UNIVERSITÀ DEL SALENTO
FACOLTÀ DI SS.MM.FF.NN.
DIPARTIMENTO DI MATEMATICA

Corso di

Statistica Applicata

A.A. 2009/2010

Lezioni del
DR. GIANFAUSTO SALVADORI
Dipartimento di Matematica
Facoltà di SS.MM.FF.NN.
Università del Salento

gianfausto.salvadori@unisalento.it

©

Indice

0 PREMESSA **1**
- 0.1 Cenni di Teoria della Misura . . . 1
- 0.2 Convergenza di v.a. e leggi . . . 3
- 0.3 Quantili . . . 5
- 0.4 Le funzioni speciali Gamma e Beta . . . 6

1 MODELLI STATISTICI **7**
- 1.1 Definizioni e proprietà . . . 7
- 1.2 Modelli statistici esponenziali . . . 9

2 STATISTICHE D'ORDINE **12**
- 2.1 Definizioni e proprietà . . . 12
- 2.2 Statistiche d'ordine estremali . . . 12
- 2.3 Leggi delle statistiche d'ordine . . . 14

3 TEORIA DEI VALORI ESTREMI **19**
- 3.1 Modelli "a blocchi" . . . 20
- 3.2 Modelli "a soglia" . . . 32

4 STIMATORI **37**
- 4.1 Stimatori . . . 37
  - 4.1.1 Media e varianza campionaria . . . 41
- 4.2 Confronto di stimatori . . . 43
- 4.3 Disuguaglianza di Fréchet-Cramér-Rao . . . 45
- 4.4 Sufficienza e completezza . . . 49

5 TECNICHE DI STIMA **54**
- 5.1 Il Metodo dei Momenti . . . 54
- 5.2 Stimatori di Massima Verosimiglianza . . . 56

6 CAMPIONI GAUSSIANI **61**
- 6.1 Legge Chi-quadro . . . 63
- 6.2 Legge t-Student . . . 64
- 6.3 Legge di Fisher-Snedecor . . . 66

7 VERIFICA DI IPOTESI **68**
- 7.1 Il Lemma di Neyman-Pearson . . . 75
- 7.2 Rapporto di verosimiglianza monotono . . . 79
- 7.3 Rapporto di verosimiglianza generalizzato . . . 85
- 7.4 Verifica di ipotesi per campioni Gaussiani . . . 89
  - 7.4.1 Test del Chi-quadro (Varianza) . . . 89
  - 7.4.2 Test t-Student (Speranza) . . . 89
  - 7.4.3 Test di Fisher-Snedecor (Confronto Varianze) . . . 89

8 STIMA PER INTERVALLI **91**
8.1 Metodo del pivot . . . 95
8.2 IC per campioni Gaussiani . . . 96

9 STATISTICA NON PARAMETRICA **98**
9.1 Stima non parametrica corretta . . . 98
9.2 I test del Chi-quadro . . . 101
9.2.1 Test del Chi-quadro per l'indipendenza . . . 104
9.2.2 Test del Chi-quadro per l'omogeneità . . . 106
9.3 I test di Kolmogorov–Smirnov . . . 107
9.3.1 Il test di adattamento di Kolmogorov–Smirnov . . . 108
9.3.2 Il test di omogeneità di Kolmogorov–Smirnov . . . 112
9.4 I test di Kendall e Spearman . . . 112
9.4.1 Il test di indipendenza di Kendall . . . 113
9.4.2 Il test di indipendenza di Spearman . . . 116

10 ANALISI DELLA VARIANZA **118**
10.1 Analisi della varianza ad una via . . . 118
10.1.1 Inferenze su combinazioni lineari . . . 119
10.1.2 Il test ANOVA ad una via . . . 122
10.1.3 Stima simultanea di contrasti . . . 125
10.2 Analisi della varianza a due vie . . . 126

11 REGRESSIONE LINEARE **130**
11.1 Regressione lineare semplice . . . 130
11.1.1 Il metodo dei Minimi Quadrati (Interpolazione) . . . 131
11.1.2 Stimatori BLUE . . . 133
11.1.3 Il modello Normale condizionale . . . 135
11.1.4 Stima e predizione nel modello Normale condizionale . . . 141
11.2 Regressione lineare multipla . . . 143

12 SIMULAZIONE **149**
12.1 Trasformazione Integrale di Probabilità . . . 149
12.2 Ulteriori schemi di simulazione univariata . . . 153
12.3 Metodo del Rigetto . . . 154
12.4 Copule e simulazione multivariata . . . 155

13 STATISTICA BAYESIANA **162**
13.1 Legge *a priori* e *a posteriori* . . . 162
13.2 Decisioni Bayesiane . . . 166
13.3 Stima Bayesiana puntuale . . . 168

A Quantili della legge Normale **176**

B Quantili della legge t-Student **177**

C Quantili della legge Chi-quadro (I) **178**

D Quantili della legge Chi-quadro (II) **179**

E Quantili della legge Fisher-Snedecor (I) **180**

F Quantili della legge Fisher-Snedecor (II) **181**

G Risultati di Algebra Lineare **182**

H Complementi sul p-value **184**

BIBLIOGRAFIA **I**

Capitolo 0

PREMESSA

Prima di intraprendere lo studio vero e proprio della Statistica Matematica si forniscono alcuni cenni riguardanti la Teoria della Misura, la Speranza Condizionale, le modalità di convergenza di v.a. e leggi, e i quantili.

0.1 Cenni di Teoria della Misura

Diamo anzitutto alcuni cenni di *Teoria della Misura* utili nel seguito del Corso. Per ulteriori approfondimenti si consultino, ad esempio, [24, 4, 34].

DEFINIZIONE 0.1 (σ-algebra)**.** Sia Ω un insieme non vuoto. Una famiglia $\mathcal{F}$ di sottoinsiemi di Ω è una *σ-algebra* (o *tribù*) se

1. $\Omega \in \mathcal{F}$;
2. se $A \in \mathcal{F}$ allora $A^{\complement} \in \mathcal{F}$;
3. se $A_1, A_2, \ldots \in \mathcal{F}$ allora $A_1 \cup A_2 \cup \cdots \in \mathcal{F}$.

Se $\mathcal{G}$ è una famiglia di sottoinsiemi di Ω, si dice *σ-algebra generata da $\mathcal{G}$*, e si scrive $\sigma(\mathcal{G})$, la più piccola σ-algebra contenente $\mathcal{G}$: essa corrisponde[1] all'intersezione di tutte le σ-algebre contenenti $\mathcal{G}$.

DEFINIZIONE 0.2. La coppia $(\Omega, \mathcal{F})$ è detta *spazio di misura* (o *misurabile*). Un insieme $A \subseteq \Omega$ è detto *misurabile* se $A \in \mathcal{F}$.

DEFINIZIONE 0.3. Dati gli spazi $(\Omega_1, \mathcal{F}_1)$ e $(\Omega_2, \mathcal{F}_2)$, una funzione $f\colon \Omega_1 \to \Omega_2$ è *misurabile* se per ogni insieme $B \in \mathcal{F}_2$ si ha $f^{-1}(B) = A$ per qualche insieme $A \in \mathcal{F}_1$.

DEFINIZIONE 0.4 (Misura)**.** Una funzione $\mu\colon \mathcal{F} \to \overline{\mathbb{R}}_+$ è una *misura* se

1. $\mu(A) \in [0, \infty]$ se $A \in \mathcal{F}$;
2. $\mu(\emptyset) = 0$;
3. se $A_1, A_2, \ldots$ è una sequenza disgiunta di elementi di $\mathcal{F}$ allora

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mu(A_n).$$

Una misura μ è

1. *finita* se $\mu(\Omega) < \infty$;
2. *infinita* se $\mu(\Omega) = \infty$;

[1] Di particolare importanza nel seguito sarà la σ-algebra dei *Boreliani* $\mathcal{B}$ generata dagli intervalli di $\mathbb{R}^d$.

3. di *probabilità* se $\mu(\Omega) = 1$;
4. σ-*finita* se $\Omega = A_1 \cup A_2 \ldots$ per una qualche sequenza, finita o numerabile, di elementi di $\mathcal{F}$ tali che $\mu(A_n) < \infty$ per ogni $n \in \mathbb{N}$.

Una misura di probabilità è finita; una misura finita è anche σ-finita, ma il viceversa può non valere (si pensi, ad esempio, alla scomposizione della retta reale in intervalli disgiunti, con la misura data dalla lunghezza degli intervalli).

DEFINIZIONE 0.5 (Spazio di probabilità)**.** La terna $(\Omega, \mathcal{F}, \mathbf{P})$, dove $\mathbf{P}$ è una misura di probabilità su $(\Omega, \mathcal{F})$, è detta *spazio di probabilità*.

ESEMPIO 0.1. ►

Gli esempi che seguono illustrano misure "naturali" su particolari spazi misurabili.

1. **La "misura del contare"**.

   Su $(\Omega, \mathcal{P}(\Omega))$ è possibile definire una "misura del contare" μ_c data da, per $A \in \mathcal{P}(\Omega)$,

   $$\mu_c(A) = \#(A)\,,$$

   cioè la *cardinalità* di A. Evidentemente μ_c è finita se, e solo se, la cardinalità di Ω è finita, e μ_c è σ-finita se, e solo se, Ω è numerabile.

2. **La misura di Lebesgue**.

   Su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ è possibile definire la *misura di Lebesgue* λ data da, per $I = (a, b)$,

   $$\lambda(I) = b - a,$$

   dove $a, b \in \mathbb{R}$, $a < b$. Evidentemente λ è σ-finita.

◄

Siano μ, ν misure su $(\Omega, \mathcal{F})$.

DEFINIZIONE 0.6. La misura ν è *assolutamente continua* rispetto alla misura μ, e si scrive $\nu \ll \mu$, se per ogni $A \in \mathcal{F}$ tale che $\mu(A) = 0$ si ha $\nu(A) = 0$.

Il Teorema che segue è fondamentale.

TEOREMA 0.1 (Radon-Nikodym)**.** *Siano μ, ν misure σ-finite su $(\Omega, \mathcal{F})$. Se $\nu \ll \mu$ allora esiste una funzione non negativa f tale che, per ogni $A \in \mathcal{F}$, $\nu(A) = \int_A f \, \mathrm{d}\mu$, ovvero $\nu = f \cdot \mu$. La funzione f è detta densità.*

NOTA 0.1.

Si osservi che il viceversa del *Teorema (0.1)* è banale: se infatti $\nu = f \cdot \mu$, $f \geq 0$, e $\mu(A) = 0$ per un insieme $A \in \mathcal{F}$, allora $\nu(A) = \int_A f \, \mathrm{d}\mu = 0$.

Sia $(\Omega, \mathcal{F}, \mathbf{P})$ uno spazio di probabilità e sia $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ lo spazio di misura dei numeri reali dotato della σ-algebra dei Boreliani $\mathcal{B}(\mathbb{R})$.

DEFINIZIONE 0.7 (Variabile aleatoria)**.** Una *variabile aleatoria* (o *casuale*) reale X è una funzione misurabile $X \colon \Omega \to \mathbb{R}$.

In altri termini X è una v.a. reale se $\{\omega \colon X(\omega) \leq r\} \in \mathcal{F}$ per ogni $r \in \mathbb{R}$. Data la v.a. reale X definita sullo spazio di probabilità $(\Omega, \mathcal{F}, \mathbf{P})$ è possibile calcolare la probabilità *indotta* da X su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, ovvero l'"immagine" della probabilità $\mathbf{P}$ definita su $(\Omega, \mathcal{F})$.

DEFINIZIONE 0.8 (Funzione di ripartizione)**.** La funzione $F_X \colon \mathbb{R} \to [0, 1]$ data da

$$F_X(x) = \mathbf{P}\{X \leq x\} = \mathbf{P}\{\omega \colon X(\omega) \leq x\} \tag{1}$$

per ogni $x \in \mathbb{R}$ è detta *funzione di ripartizione* (o *funzione di distribuzione cumulata*) di X.

Chiaramente F_X consente di calcolare la probabilità di opportuni sottoinsiemi di $\mathbb{R}$, ovvero di quelli misurabili appartenenti a $\mathcal{B}(\mathbb{R})$.

Poiché la *speranza condizionale* gioca un ruolo fondamentale nella costruzione di svariate statistiche, si forniscono alcuni cenni introduttivi su tale argomento (si veda, ad esempio, [34]). Il Teorema che segue contiene implicitamente la definizione di speranza condizionale.

TEOREMA 0.2 (Kolmogorov)**.** *Sia $X \in \mathrm{L}^1(\Omega, \mathcal{F}, \mathbf{P})$ e sia $\mathcal{G}$ una sotto-σ-algebra di $\mathcal{F}$. Allora esiste una v.a. Y tale che:*

1. *Y è $\mathcal{G}$-misurabile;*
2. *$\mathbf{E}(|Y|) < \infty$, cioè $Y \in \mathrm{L}^1(\Omega, \mathcal{G}, \mathbf{P})$;*
3. *per ogni insieme $G \in \mathcal{G}$ vale*
$$\int_G Y \, \mathrm{d}\mathbf{P} = \int_G X \, \mathrm{d}\mathbf{P}. \tag{2}$$

Inoltre, se Z è un'altra v.a. con le stesse proprietà di Y allora vale $\mathbf{P}\{Z = Y\} = 1$.

DEFINIZIONE 0.9 (Speranza condizionale)**.** Una v.a. Y con le proprietà *(1)–(3)* del *Teorema (0.2)* è una versione della *speranza condizionale* di X data $\mathcal{G}$, e si scrive
$$Y = \mathbf{E}(X \mid \mathcal{G}) \quad \text{q.c..} \tag{3}$$

0.2 Convergenza di v.a. e leggi

Nell'ambito della Statistica Matematica esistono svariate modalità di convergenza di successioni di v.a. e delle loro leggi (per maggiori dettagli si vedano, ad esempio, [24, 12, 6]).

Una prima modalità di convergenza è la cosiddetta *convergenza in legge*.

DEFINIZIONE 0.10 (Convergenza in legge)**.** Sia $\{F_n\}_{n\in\mathbb{N}}$ una successione di f.r.; se esiste una f.r. F tale che, quando $n \to \infty$,
$$F_n(x) \to F(x) \tag{4}$$
per ogni punto di continuità di F, si dice che F_n *converge in legge (o debolmente)* a F, e si scrive $F_n \xrightarrow{\mathcal{L}} F$. Sia $\{X_n\}_{n\in\mathbb{N}}$ una successione di v.a. e sia $\{F_n\}_{n\in\mathbb{N}}$ la successione delle f.r. corrispondenti; se esiste una v.a. X con f.r. F tale che $F_n \xrightarrow{\mathcal{L}} F$ quando $n \to \infty$ si dice che X_n *converge in legge (o in distribuzione)* a X, e si scrive $X_n \xrightarrow{\mathcal{L}} X$.

Si noti che la convergenza in legge non richiede che le v.a. della successione considerata siano definite sullo stesso spazio di probabilità.

OSSERVAZIONE 0.1.

La convergenza in legge non comporta necessariamente quella dei rispettivi momenti. Si considerino infatti le seguenti f.r.:
$$F(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}, \qquad F_n(x) = \begin{cases} 0, & x < 0 \\ 1 - \frac{1}{n}, & 0 \leq x < n \\ 1, & x \geq n \end{cases}, \quad \text{con } n \in \mathbb{N}.$$
Evidentemente $F_n \xrightarrow{\mathcal{L}} F$; tuttavia, il calcolo del momento di ordine $m \in \mathbb{N}$ restituisce $\mathbf{E}(X^m) = 0$, mentre $\mathbf{E}(X_n^m) = n^{m-1}$.

Il risultato che segue è importante quando si considerino v.a. reali.

PROPOSIZIONE 0.1. *Sia $X_n \xrightarrow{\mathcal{L}} X$ e sia $g\colon \mathbb{R} \to \mathbb{R}$ una funzione continua. Allora*
$$g(X_n) \xrightarrow{\mathcal{L}} g(X). \tag{5}$$

Un'ulteriore modalità di convergenza è la cosiddetta *convergenza in probabilità*.

DEFINIZIONE 0.11 (Convergenza in probabilità)**.** Sia $\{X_n\}_{n\in\mathbb{N}}$ una successione di v.a. definite sullo stesso spazio di probabilità $(\Omega, \mathcal{F}, \mathbf{P})$. Si dice che X_n converge *in probabilità* alla v.a. X se, per ogni $\epsilon > 0$,

$$\mathbf{P}\{|X_n - X| > \epsilon\} \to 0 \tag{6}$$

per $n \to \infty$, e si scrive $X_n \xrightarrow{\mathbf{P}} X$.

Si noti che la convergenza in probabilità richiede che le v.a. della successione considerata siano definite sullo stesso spazio di probabilità.

OSSERVAZIONE 0.2.

Il controesempio dell'*Osservazione (0.1)* mostra che la convergenza in probabilità non comporta necessariamente quella dei rispettivi momenti.

Il risultato che segue è importante quando si considerino v.a. reali.

PROPOSIZIONE 0.2. *Sia $X_n \xrightarrow{\mathbf{P}} X$ e sia $g\colon \mathbb{R} \to \mathbb{R}$ una funzione continua. Allora*

$$g(X_n) \xrightarrow{\mathbf{P}} g(X). \tag{7}$$

Il legame tra la convergenza in probabilità e quella in legge è sancito dal seguente Teorema.

TEOREMA 0.3. *Sia $X_n \xrightarrow{\mathbf{P}} X$; allora $X_n \xrightarrow{\mathcal{L}} X$.*

OSSERVAZIONE 0.3.

In generale il viceversa del *Teorema (0.3)* non vale. Si considerino infatti la v.a. $X \sim \mathtt{U_d}(0, 1)$ e la successione $\{X_n\}_{n\in\mathbb{N}}$, con $X_n = 1 - X$ per ogni indice n. Evidentemente X e X_n hanno la stessa legge, ma $|X_n - X| = 1$ per ogni indice n; dunque X_n non può convergere in probabilità a X.

Un'ulteriore modalità di convergenza è la cosiddetta *convergenza in media-r*.

DEFINIZIONE 0.12 (Convergenza in media-r)**.** Sia $\{X_n\}_{n\in\mathbb{N}}$ una successione di v.a. definite sullo stesso spazio di probabilità $(\Omega, \mathcal{F}, \mathbf{P})$ tali che $\mathbf{E}(|X_n|^r) < \infty$, dove $r > 0$. Si dice che X_n converge *in media-r* alla v.a. X, con $\mathbf{E}(|X|^r) < \infty$, se

$$\mathbf{E}(|X_n - X|^r) \to 0 \tag{8}$$

per $n \to \infty$, e si scrive $X_n \xrightarrow{r} X$.

Anche la convergenza in media-r richiede che le v.a. della successione considerata siano definite sullo stesso spazio di probabilità. Il risultato che segue è immediato.

PROPOSIZIONE 0.3. *Sia $r > s > 0$ e sia $X_n \xrightarrow{r} X$. Allora*

$$X_n \xrightarrow{s} X. \tag{9}$$

Il legame tra la convergenza in media-r e quella in probabilità è sancito dal seguente Teorema.

TEOREMA 0.4. *Sia $X_n \xrightarrow{r} X$, $r > 0$; allora $X_n \xrightarrow{\mathbf{P}} X$.*

OSSERVAZIONE 0.4.

In generale il viceversa del *Teorema (0.4)* non vale. Sia infatti $r > 0$ e si consideri la successione di v.a. così definite, per $n \in \mathbb{N}$:

$$\mathbf{P}\{X_n = 0\} = 1 - \frac{1}{n^r} \qquad \text{e} \qquad \mathbf{P}\{X_n = n\} = \frac{1}{n^r}.$$

Evidentemente $\mathbf{E}(|X_n|^r) = 1$ per ogni indice n, e dunque X_n non può convergere in media-r alla v.a. degenere nulla. Tuttavia si ha, per ogni $\epsilon > 0$,

$$\mathbf{P}\{|X_n| > \epsilon\} = \begin{cases} \mathbf{P}\{X_n = n\}, & \text{se } \epsilon \le n \\ 0, & \text{se } \epsilon > n \end{cases},$$

che tende a zero per $n \to \infty$, e dunque $X_n \xrightarrow{\mathbf{P}} 0$.

NOTA 0.2.

Grazie al *Teorema (0.3)*, dal *Teorema (0.4)* scende che la convergenza in media-r comporta quella in legge.

Un'ulteriore modalità di convergenza è la cosiddetta *convergenza quasi certa (q.c.)*.

DEFINIZIONE 0.13 (Convergenza quasi certa)**.** Sia $\{X_n\}_{n\in\mathbb{N}}$ una successione di v.a. definite sullo stesso spazio di probabilità $(\Omega, \mathcal{F}, \mathbf{P})$. Si dice che X_n converge *quasi certamente (q.c.)* alla v.a. X se

$$\mathbf{P}\Big\{\omega\colon X_n(\omega) \xrightarrow{n\to\infty} X(\omega)\Big\} = 1, \tag{10}$$

e si scrive $X_n \xrightarrow{\text{q.c.}} X$.

Anche la convergenza quasi certa richiede che le v.a. della successione considerata siano definite sullo stesso spazio di probabilità.

OSSERVAZIONE 0.5.

La definizione di convergenza quasi certa può essere riformulata come segue: X_n *converge quasi certamente a* X *se, e solo se, per ogni* $\epsilon > 0$*, vale:*

$$\lim_{n\to\infty} \mathbf{P}\left\{\sup_{m\geq n} |X_m - X| > \epsilon\right\} = 0. \tag{11}$$

Il legame tra la convergenza quasi certa e quella in probabilità è sancito dal seguente Teorema.

TEOREMA 0.5. *Sia* $X_n \xrightarrow{q.c.} X$*; allora* $X_n \xrightarrow{\mathbf{P}} X$.

NOTA 0.3.

Grazie al *Teorema (0.3)*, dal *Teorema (0.5)* scende che la convergenza quasi certa comporta quella in legge.

In generale il viceversa del *Teorema (0.5)* non vale. Il diagramma che segue mostra le relazioni tra le diverse modalità di convergenza illustrate.

$$\begin{array}{ccccc} X_n \xrightarrow{\text{q.c.}} X & & & & \\ \Downarrow \quad \not\Uparrow & \Rightarrow & X_n \xrightarrow{\mathbf{P}} X & \Rightarrow & X_n \xrightarrow{\mathcal{L}} X \\ X_n \xrightarrow{r} X & \not\Leftarrow & & \not\Leftarrow & \end{array}$$

0.3 Quantili

Nella pratica i *quantili* di una distribuzione forniscono informazioni spesso essenziali per l'analisi di un fenomeno e in fase di progettazione.

DEFINIZIONE 0.14 (Funzione dei quantili e quantili)**.** Sia $F\colon \mathbb{R} \to [0,1]$ una distribuzione arbitraria. La funzione $q_F\colon (0,1) \to \mathbb{R}$ data da

$$q_F(p) = \inf\{x\colon p \leq F(x)\}, \tag{12}$$

dove $p \in (0,1)$, è detta *funzione dei quantili* di F. Fissato $p \in (0,1)$ il numero reale

$$x_p = q_F(p) \tag{13}$$

è detto *quantile* (o *percentile*) *di ordine* p di F.

ESEMPIO 0.2 (Mediana)**.** ▶

Prendendo $p = \frac{1}{2}$, il quantile $x_{\frac{1}{2}} = q_F\left(\frac{1}{2}\right)$ di ordine $\frac{1}{2}$ corrisponde alla *mediana* della distribuzione F. ◀

I risultati che seguono sono intuitivi.

PROPOSIZIONE 0.4. *Sia q_F la funzione dei quantili della distribuzione F; allora q_F è una funzione non decrescente e inoltre valgono le seguenti relazioni:*

$$\{x\colon p \leq F(x)\} = [q_F(p), \infty), \tag{14a}$$

$$p \leq F(x) \Leftrightarrow q_F(p) \leq x, \tag{14b}$$

$$F\left(q_F(p)\right) \geq p, \tag{14c}$$

dove $p \in (0, 1)$.

Si noti che se X è *assolutamente continua* con f.r. F strettamente crescente, allora la disuguaglianza dell'*equazione (14c)* si trasforma nell'uguaglianza $F\left(x_p\right) = p$.

0.4 Le funzioni speciali Gamma e Beta

La *funzione Gamma di Eulero* e la *funzione Beta* giocano un ruolo fondamentale nel Calcolo delle Probabilità.

DEFINIZIONE 0.15 (Funzione Gamma di Eulero)**.** Sia $\alpha > 0$; la funzione *Gamma di Eulero* $\Gamma(\alpha)$ è definita come:

$$\Gamma(\alpha) = \int_{0^+}^{\infty} x^{\alpha-1}\, e^{-x}\, \mathrm{d}x. \tag{15}$$

I valori di $\Gamma(\alpha)$ sono calcolabili esplicitamente solo in alcuni casi particolari, spesso ricorrendo alla seguente formula ricorsiva ottenibile tramite una semplice integrazione per parti:

$$\Gamma(\alpha) = (\alpha - 1) \int_{0^+}^{\infty} x^{\alpha-2}\, e^{-x}\, dx = (\alpha - 1)\, \Gamma(\alpha - 1)\,, \tag{16}$$

dove $\alpha > 1$. In virtù della proprietà espressa dall'*equazione (16)* la funzione Gamma è anche detta *fattoriale generalizzato*. In particolare, $\Gamma(1) = 1$ e, per $n \in \mathbb{N}$:

$$\Gamma(n) = (n - 1)\, \Gamma(n - 1) = \cdots = (n - 1)! \tag{17}$$

DEFINIZIONE 0.16 (Funzione Beta)**.** Siano $a > 0$ e $b > 0$; la funzione *Beta* $\mathrm{B}(a, b)$ è definita come:

$$\mathrm{B}(a, b) = \int_{0^+}^{1^-} x^{a-1}\, (1 - x)^{b-1}\, \mathrm{d}x. \tag{18}$$

I valori di $\mathrm{B}(a, b)$ sono calcolabili esplicitamente solo in alcuni casi particolari, spesso ricorrendo alla seguente formula:

$$\mathrm{B}(a, b) = \frac{\Gamma(a)\, \Gamma(b)}{\Gamma(a + b)} \tag{19}$$

dove $a, b > 0$. Ne segue che:

$$\frac{1}{\mathrm{B}(a, b)} \int_{0^+}^{1^-} x^{a-1}\, (1 - x)^{b-1}\, dx = 1.$$

Infine, è possibile definire le versioni "incomplete" delle due funzioni speciali appena introdotte.

DEFINIZIONE 0.17 (Funzione Gamma Incompleta)**.** Siano $t > 0$ e $\alpha > 0$; la funzione *Gamma Incompleta* $\Gamma_t(\alpha)$ è definita come:

$$\Gamma_t(\alpha) = \int_{t}^{\infty} x^{\alpha-1}\, e^{-x}\, \mathrm{d}x. \tag{20}$$

DEFINIZIONE 0.18 (Funzione Beta Incompleta)**.** Siano $p \in (0, 1)$ e $a, b > 0$; la funzione *Beta Incompleta* $\mathrm{B}_p(a, b)$ è definita come:

$$\mathrm{B}_p(a, b) = \int_{0^+}^{p} x^{a-1}\, (1 - x)^{b-1}\, \mathrm{d}x. \tag{21}$$

Capitolo 1

MODELLI STATISTICI

1.1 Definizioni e proprietà

Diamo anzitutto la definizione di *modello statistico dominato* e di *verosimiglianza*.

DEFINIZIONE 1.1 (Modello statistico dominato)**.** Sia $\Theta \subseteq \mathbb{R}^d$, $d \geq 1$. Si dice *modello statistico dominato* un insieme $\mathfrak{M} = \left(\Omega, \mathcal{F}, \{\mathbf{P}_\theta\}_{\theta\in\Theta}\right)$ se si ha $\mathbf{P}_\theta \ll \mu$ per ogni $\theta \in \Theta$, dove $\mathbf{P}_\theta$ è una misura di probabilità su $(\Omega, \mathcal{F})$ e μ una misura σ-finita su $(\Omega, \mathcal{F})$.

OSSERVAZIONE 1.1.

Ovviamente $\mathbf{P}_\theta$ è una misura finita e quindi anche σ-finita. La misura μ è detta *dominante*. Grazie al *Teorema (0.1)*, implicitamente la *Definizione (1.1)* consente di affermare l'esistenza di una *densità*

$$f_\theta = \frac{\mathrm{d}\mathbf{P}_\theta}{\mathrm{d}\mu} \tag{1.1}$$

per le v.a. definite su $\mathfrak{M}$. In particolare, nel caso di v.a. discrete μ tipicamente coinciderà con la "misura del contare" μ_c, mentre nel caso di v.a. assolutamente continue μ tipicamente coinciderà con la misura di Lebesgue λ (o con una sua restrizione ad un opportuno Boreliano di $\mathbb{R}^d$).

ESEMPIO 1.1. ►

Gli esempi che seguono illustrano particolari modelli statistici dominati.

1. Su $(\{0,1\}, \mathcal{P}(\{0,1\}))$ le leggi Bernoulliane $\mathtt{Be}(\theta)$, $\theta \in (0,1)$, formano un modello statistico dominato; se μ è la "misura del contare" μ_c su $\{0,1\}$ si ha

   $$\mathbf{P}_\theta(k) = \theta^k \, (1-\theta)^{1-k} \, \mu_c(k),$$

   dove $k \in \{0,1\}$; evidentemente è $f_\theta(k) = \theta^k \, (1-\theta)^{1-k} \, \mathbb{I}_{\{0,1\}}(k)$.

2. Su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ le leggi Esponenziali $\mathtt{e}(\theta)$, $\theta > 0$, formano un modello statistico dominato; se μ è la misura di Lebesgue λ su $\mathbb{R}_+$ si ha

   $$\mathbf{P}_\theta(\mathrm{d}x) = \theta \, e^{-\theta x} \, \lambda(\mathrm{d}x),$$

   dove $x > 0$; evidentemente è $f_\theta(x) = \theta \, e^{-\theta x} \, \mathbb{I}_{\mathbb{R}_+}(x)$.

◄

DEFINIZIONE 1.2 (Verosimiglianza)**.** Assegnato un modello statistico dominato $\mathfrak{M}$ si dice *verosimiglianza* una funzione $\mathbf{L}\colon \Theta \times \Omega \to \mathbb{R}_+$ tale che, fissato $\theta \in \Theta$, la funzione $\omega \mapsto \mathbf{L}_\theta(\omega)$ sia una *versione* della densità $f_\theta = \frac{\mathrm{d}\mathbf{P}_\theta}{\mathrm{d}\mu}$, cioè $\mu\left(\{f_\theta \neq \mathbf{L}_\theta\}\right) = 0$.

DEFINIZIONE 1.3 (Modello statistico dominato regolare)**.** Un modello statistico dominato $\mathfrak{M}$ si dice *regolare* se il supporto della verosimiglianza $\mathbf{L}_\theta$ (ovvero della densità f_θ) non dipende da θ.

ESEMPIO 1.2. ►

Su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ si ha che:

1. le leggi Esponenziali $\mathtt{e}(\theta)$, $\theta > 0$, formano un modello statistico dominato regolare: il supporto della densità f_θ è $\mathbb{R}_+$, che non dipende da θ;
2. la legge Uniforme $\mathtt{U_c}(0,1)$ forma un modello statistico dominato regolare: il supporto della densità f_θ è $(0,1)$, che non dipende da θ (questa legge non dipende da alcun parametro!);
3. le leggi Uniformi $\mathtt{U_c}(0,\theta)$, $\theta > 0$, formano un modello statistico dominato non regolare: il supporto della densità f_θ è $(0,\theta)$, che dipende da θ.

◄

Diamo ora le definizioni di *campione* e di *statistica*.

DEFINIZIONE 1.4 (Campione)**.** Siano $X_1, \ldots, X_n$, $n \in \mathbb{N}$, v.a. i.i.d.[1] definite su $\mathfrak{M}$ con funzione di ripartizione comune F. L'insieme di tali v.a. è detto *campione (statistico) di ampiezza (o taglia)* n (o, ancora, n*-campione*) estratto dalla distribuzione F.

Per denotare un campione statistico $X_1, \ldots, X_n$ di ampiezza n estratto dalla distribuzione F useremo la notazione $\mathbf{X}_F^n$. Se le v.a. $X_1, \ldots, X_n$ sono indipendenti con f.r., rispettivamente, $F_1, \ldots, F_n$, allora scriveremo $\mathbf{X}_{F_i}^n$ — tuttavia in quest'ultimo caso l'appellativo di "campione" non è del tutto corretto e lecito. Si noti che $\mathbf{X}_F^n$ non è altro che un *vettore aleatorio* $\boldsymbol{X} = (X_1, \ldots, X_n)$ le cui componenti rappresentano n osservazioni distinte e i.i.d. della v.a. X avente distribuzione F.

ESEMPIO 1.3. ►

Gli esempi che seguono illustrano particolari campioni statistici.

1. Si considerino n lanci indipendenti di una moneta, con probabilità $\theta \in (0,1)$ di ottenere testa ad ogni lancio. Se le v.a. X_i, $i = 1, \ldots, n$, valgono zero o uno a seconda che esca croce o testa al lancio i-esimo, il vettore aleatorio $\boldsymbol{X} = (X_1, \ldots, X_n)$ rappresenta un campione di ampiezza n estratto dalla distribuzione (discreta) $F \sim \mathtt{Be}(\theta)$.
2. Si considerino n lampade identiche e indipendenti, la cui durata è rappresentata da una v.a. $X \sim \mathtt{e}(\theta)$, $\theta > 0$. Il vettore aleatorio $\boldsymbol{X} = (X_1, \ldots, X_n)$, che consiste delle n osservazioni della durata delle lampade, rappresenta un campione di ampiezza n estratto dalla distribuzione (continua) $F \sim \mathtt{e}(\theta)$.

◄

DEFINIZIONE 1.5 (Statistica)**.** Si dice *statistica* una v.a. $T = t(X_1, \ldots, X_n)$, $n \in \mathbb{N}$, dove $t\colon \mathbb{R}^n \to \mathbb{R}$ è una funzione misurabile che non dipende da alcun parametro incognito.

Dunque una statistica è semplicemente una funzione (misurabile) del campione in esame che non dipende dagli eventuali parametri della distribuzione di probabilità soggiacente.

ESEMPIO 1.4. ►

Si consideri il campione dell'*Esempio (1.3)(2).* Ciascuna delle seguenti v.a. è una statistica:

1. $T = \overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ (media campionaria o vita media delle lampade);
2. $T = X_1 = \pi_1(X_1, \ldots, X_n)$ (primo esito o durata della lampada "1");
3. $T = X_n = \pi_n(X_1, \ldots, X_n)$ (ultimo esito o durata della lampada "n");
4. $T = X_{(1)} = \min\{X_1, \ldots, X_n\}$ (minimo o durata della lampada bruciata per prima);
5. $T = X_{(n)} = \max\{X_1, \ldots, X_n\}$ (massimo o durata della lampada bruciata per ultima);
6. $T = c \in \mathbb{R}$ (è pur sempre una funzione misurabile del campione!).

Al contrario, la v.a. $T = \theta\left(X_{(n)} - X_{(1)}\right)$ non è una statistica. ◄

[1] L'acronimo "i.i.d." indica variabili *i*ndipendenti e *i*denticamente *d*istribuite.

1.2 Modelli statistici esponenziali

I modelli statistici *esponenziali* occupano un posto privilegiato nell'ambito della Statistica Matematica.

DEFINIZIONE 1.6 (Modello statistico esponenziale)**.** Un modello statistico dominato $\mathfrak{M}$ si dice *esponenziale* se è possibile scrivere la verosimiglianza nella forma

$$\mathbf{L}_\theta(\omega) = \frac{\mathrm{d}\mathbf{P}_\theta}{\mathrm{d}\mu}(\omega) = C(\theta)\, e^{\langle \theta, T(\omega)\rangle} = e^{\langle \theta, T(\omega)\rangle + \ln C(\theta)}, \tag{1.2}$$

dove $C(\theta) \in \mathbb{R}_+$ è una costante di normalizzazione, $T\colon \Omega \to \mathbb{R}^d$ è una v.a. e $\theta \in \Theta$, essendo Θ un aperto convesso contenuto nell'insieme $\mathfrak{D}(\mu)$ definito da

$$\mathfrak{D}(\mu) = \left\{ \theta \in \mathbb{R}^d \colon \int_{\mathbb{R}^d} e^{\langle \theta, x\rangle}\, \mathrm{d}\mu(x) < +\infty \right\}.$$

La costante di normalizzazione $C(\theta)$ si ottiene dalla condizione ovvia $\int_\Omega f_\theta\, \mathrm{d}\mu = 1$, cioè $\frac{1}{C(\theta)} = \int_\Omega e^{\langle \theta, T\rangle}\, \mathrm{d}\mu$.

L'insieme $\mathfrak{D}(\mu)$ rappresenta il *dominio proprio* della *Trasformata di Laplace* $\mathcal{L}_\mu$ della misura μ, cioè $\mathcal{L}_\mu(\theta) = \int_{\mathbb{R}^d} e^{\langle \theta, x\rangle}\, \mathrm{d}\mu(x)$, che si assume non vuoto.

OSSERVAZIONE 1.2.

La *Definizione (1.6)* implicitamente *esclude* che in un modello esponenziale il supporto della verosimiglianza $\mathbf{L}_\theta$ (e quindi della densità f_θ) possa dipendere da θ; dunque un modello esponenziale è necessariamente *regolare*.

Ponendo

$$\chi(\theta) = \ln \int_\Omega e^{\langle \theta, T\rangle}\, \mathrm{d}\mu = -\ln C(\theta) \tag{1.3}$$

si perviene alla seguente *forma canonica*.

DEFINIZIONE 1.7 (Forma canonica)**.** La verosimiglianza di un modello esponenziale può essere scritta nella *forma canonica*

$$\mathbf{L}_\theta = e^{\langle \theta, T\rangle - \chi(\theta)}. \tag{1.4}$$

In particolare, nel caso $d = 1$ di un parametro scalare, l'*equazione (1.4)* si scrive $\mathbf{L}_\theta = e^{\theta T - \chi(\theta)}$. Gli esempi che seguono mostrano alcuni tra i modelli esponenziali più frequentemente usati.

ESEMPIO 1.5 (Legge $\mathtt{Be}(\theta)$)**.** ▶

Sia $X \sim \mathtt{Be}(\theta)$, $\theta \in (0,1)$. La densità di X rispetto alla "misura del contare" μ_c su $(\{0,1\}, \mathcal{P}(\{0,1\}))$ è

$$f_\theta(k) = \theta^k\, (1-\theta)^{1-k} = e^{k \ln \frac{\theta}{1-\theta} + \ln(1-\theta)},$$

dove $k \in \{0,1\}$. Basta dunque porre $\theta' = \ln \frac{\theta}{1-\theta}$ per ricondursi alla forma canonica

$$f_{\theta'}(k) = e^{\theta' k - \ln\left(1+e^{\theta'}\right)}.$$

◀

ESEMPIO 1.6 (Legge $\mathtt{P}(\theta)$)**.** ▶

Sia $X \sim \mathtt{P}(\theta)$, $\theta > 0$. La densità di X rispetto alla "misura del contare" μ_c su $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ è

$$f_\theta(k) = e^{-\theta}\, \frac{\theta^k}{k!} = e^{-\theta + k \ln \theta - \ln k!},$$

dove $k \in \mathbb{N}$. Operando il cambio di parametro $\theta' = \ln \theta$ e prendendo come nuova misura dominante la misura μ, definita sui *singoletti*, data da

$$\mu(\{k\}) = \frac{\mu_c(\{k\})}{k!},$$

dove $k \in \mathbb{N}$, è possibile ricondursi alla forma canonica

$$f_{\theta'}(k) = e^{\theta' k - e^{\theta'}}.$$

◀

ESEMPIO 1.7 (Legge $\Gamma(\theta_1, \theta_2)$). ►
Sia $X \sim \Gamma(\theta_1, \theta_2)$, $\theta_1, \theta_2 > 0$; evidentemente $\theta = (\theta_1, \theta_2)$ e $\Theta = \mathbb{R}_+ \times \mathbb{R}_+$. La densità di X rispetto alla misura di Lebesgue λ su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ è

$$f_\theta(x) = \frac{\theta_2^{\theta_1}}{\Gamma(\theta_1)} x^{\theta_1 - 1} e^{-\theta_2 x} = e^{-\theta_2 x + (\theta_1 - 1)\ln x + \theta_1 \ln \theta_2 - \ln \Gamma(\theta_1)},$$

dove $x > 0$. Operando il cambio di parametro $\theta' = (\theta_1 - 1, \theta_2)$ e prendendo come v.a. $T = (\ln X, -X)$ è possibile ricondursi alla forma canonica. ◄

ESEMPIO 1.8 (Legge $\mathtt{N}(\theta_1, \theta_2)$). ►
Sia $X \sim \mathtt{N}(\theta_1, \theta_2)$, con $\theta_1 \in \mathbb{R}$ e $\theta_2 > 0$; evidentemente $\theta = (\theta_1, \theta_2)$ e $\Theta = \mathbb{R} \times \mathbb{R}_+$. La densità di X rispetto alla misura di Lebesgue λ su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ è

$$f_\theta(x) = \frac{1}{\sqrt{2\pi\theta_2}} e^{-\frac{(x-\theta_1)^2}{2\theta_2}} = e^{-\frac{x^2}{2\theta_2} + \frac{\theta_1}{\theta_2} x - \frac{\theta_1^2}{2\theta_2} - \frac{1}{2}\ln \theta_2 - \frac{1}{2}\ln(2\pi)},$$

dove $x \in \mathbb{R}$. Operando il cambio di parametro $\theta' = (\frac{\theta_1}{\theta_2}, -\frac{1}{2\theta_2})$ e prendendo come v.a. $T = (X, X^2)$ è possibile ricondursi alla forma canonica. ◄

ESEMPIO 1.9. ►
Sia $X \sim \mathtt{U_c}(0, \theta)$, $\theta > 0$. Le v.a. con questa legge non sono riconducibili ad un modello esponenziale: infatti la densità di X rispetto alla misura di Lebesgue λ su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ è

$$f_\theta(x) = \frac{1}{\theta} \mathbb{I}_{(0,\theta)}(x),$$

dove $x \in \mathbb{R}$. Evidentemente il modello non è *regolare* e non è possibile ricondursi alla forma canonica in quanto il supporto di f_θ dipende da θ, ciò che è implicitamente escluso dalla *Definizione (1.6).* ◄

OSSERVAZIONE 1.3 (Regola "s.p.q.r.").
Per quanto visto finora ci si convince facilmente che una famiglia $\{\mu_\theta\}_{\theta \in \Theta}$, $\Theta \subseteq \mathbb{R}$, su $(\mathbb{Z}, \mathcal{P}(\mathbb{Z}))$ o $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, è di classe esponenziale se la densità f_θ di μ_θ rispetto alla misura dominante (rispettivamente, la "misura del contare" μ_c o la misura di Lebesgue λ) si può scrivere nella forma

$$f_\theta(x) = s(x)\, e^{p(\theta)q(x) + r(\theta)}, \tag{1.5}$$

dove s, p, q, r sono opportune funzioni.

ESEMPIO 1.10. ►
Si riconsideri l'*Esempio (1.6).* Alla luce dell'*Osservazione (1.3)* è chiaro che basta porre $s(k) = \frac{1}{k!}$, $p(\theta) = \ln \theta$, $q(k) = k$ e $r(\theta) = -\theta$ per ricondursi alla forma canonica data dall'*equazione (1.5).* ◄

ESEMPIO 1.11 (Legge $\mathtt{Bi}(n, \theta)$). ►
Sia $X \sim \mathtt{Bi}(n, \theta)$, dove n è *noto* e $\theta \in (0, 1)$ (e perciò $\Theta = (0, 1)$). La densità di X rispetto alla "misura del contare" μ_c su $(\{0, \ldots, n\}, \mathcal{P}(\{0, \ldots, n\}))$ è

$$f_\theta(k) = \binom{n}{k} \theta^k (1 - \theta)^{n-k} = \binom{n}{k} e^{k \ln \frac{\theta}{1-\theta} + n \ln(1-\theta)},$$

dove $k \in \{0, \ldots, n\}$. Basta dunque porre $s(k) = \binom{n}{k}$, $p(\theta) = \ln \frac{\theta}{1-\theta}$, $q(k) = k$ e $r(\theta) = n \ln(1 - \theta)$ per ricondursi alla forma canonica data dall'*equazione (1.5).* ◄

I risultati del Teorema che segue sono di uso frequente.

TEOREMA 1.1. *In un modello esponenziale, con la verosimiglianza data nella forma canonica, valgono le seguenti relazioni:*

1. *per ogni $\theta \in \Theta$ si ha $\mathbf{E}_\theta(T_i) = \mathrm{D}_i^{(1)} \chi(\theta)$, con $i = 1, \ldots, d$;*
2. *per ogni $\theta \in \Theta$ si ha $\mathbf{V}_\theta(T_i) = \mathrm{D}_i^{(2)} \chi(\theta)$, con $i = 1, \ldots, d$;*
3. *per ogni $\theta \in \Theta$ si ha $\mathbf{C}_\theta(T_i, T_j) = \mathrm{D}_i^{(1)} \mathrm{D}_j^{(1)} \chi(\theta)$, con $i, j = 1, \ldots, d$.*

Qui $\mathrm{D}_i^{(r)}$ indica la derivata parziale r-esima[2] rispetto alla i-esima componente di θ, cioè $\mathrm{D}_i^{(r)} = \frac{\partial^r}{\partial \theta_i^r}$.

[2] Per semplicità nel seguito porremo $\mathrm{D}_i = \mathrm{D}_i^{(r)}$ quando $r = 1$.

DIM. Dimostreremo solo il risultato *(1)*. Poiché in un modello esponenziale il supporto della verosimiglianza (densità) non dipende da θ si ha

$$\begin{aligned} \mathrm{D}_i \int_\Omega e^{\langle \theta, T \rangle} \, \mathrm{d}\mu &= \int_\Omega \mathrm{D}_i \, e^{\langle \theta, T \rangle} \, \mathrm{d}\mu \\ &= \int_\Omega T_i \, e^{\langle \theta, T \rangle} \, \mathrm{d}\mu. \end{aligned}$$

Perciò la derivata parziale $\mathrm{D}_i \, \chi(\theta)$ vale

$$\begin{aligned} \mathrm{D}_i \left(\ln \int_\Omega e^{\langle \theta, T \rangle} \, \mathrm{d}\mu \right) &= \frac{\int_\Omega T_i \, e^{\langle \theta, T \rangle} \, \mathrm{d}\mu}{\int_\Omega e^{\langle \theta, T \rangle} \, \mathrm{d}\mu} \\ &= \int_\Omega T_i \, e^{\langle \theta, T \rangle - \chi(\theta)} \, \mathrm{d}\mu \\ &= \int_\Omega T_i \, \mathbf{L}_\theta \, \mathrm{d}\mu \\ &= \int_\Omega T_i \, \mathrm{d}\mathbf{P}_\theta = \mathbf{E}_\theta(T_i) \, , \end{aligned}$$

da cui la tesi. Gli altri risultati si dimostrano analogamente. □

Capitolo 2

STATISTICHE D'ORDINE

2.1 Definizioni e proprietà

Un insieme di *statistiche* (cioè di funzioni misurabili del campione) di grande importanza nell'ambito della Statistica Matematica è rappresentato dalle cosiddette *statistiche d'ordine*; per ulteriori dettagli si consultino [9, 7].

DEFINIZIONE 2.1 (Statistiche d'ordine)**.** Sia $\mathbf{X}_F^n$ un campione. Allora le v.a. $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$, ovvero le stesse v.a. $X_1, \ldots, X_n$ disposte in ordine *crescente*, rappresentano le *statistiche d'ordine* corrispondenti al campione.

OSSERVAZIONE 2.1.

Evidentemente, a differenza del campione $\mathbf{X}_F^n$ formato da v.a. i.i.d., le statistiche d'ordine *non sono indipendenti*: infatti, se ad esempio $X_{(1)} > x$, allora necessariamente (q.c.) $X_{(i)} > x$ per ogni indice i. Inoltre, come vedremo meglio in seguito, in generale le statistiche d'ordine *non sono identicamente distribuite*.

ESEMPIO 2.1. ►

Si supponga di eseguire un esperimento consistente nella misurazione della durata (aleatoria) della vita di n lampade identiche e indipendenti. Se X_i, $i = 1, \ldots, n$, denota la durata della vita della lampada i-esima, al termine dell'esperimento avremo che $X_{(1)}$ corrisponderà alla durata della vita della lampada, tra le n in prova, che è bruciata per *prima* (cioè alla durata *minima*), $X_{(2)}$ alla durata della vita della lampada che è bruciata per *seconda*, …, $X_{(n)}$ alla durata della vita della lampada che è bruciata per *ultima* (cioè alla durata *massima*); ovviamente è $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$. ◄

ESEMPIO 2.2 (Livelli fluviali)**.** ►

Il livello dell'acqua in un fiume è una variabile di estremo interesse per gli ingegneri idraulici, in quanto può fornire informazioni essenziali sia per l'irrigazione sia per la costruzione di opere di contenimento delle eventuali piene (argini, casse di espansione, ecc…). Ora, siano X_i, $i = 1, \ldots, n$, i livelli d'acqua misurati in giorni successivi in un certo fiume: è chiaro che le statistiche d'ordine $X_{(1)}, \ldots, X_{(k)}$ (ovvero quelle "minimali") riflettono il comportamento critico del sistema nei periodi di siccità, mentre le statistiche d'ordine $X_{(n-k)}, \ldots, X_{(n)}$ (ovvero quelle "massimali") consentono lo studio di fenomeni altrettanto disastrosi quali alluvioni ed esondazioni. ◄

2.2 Statistiche d'ordine estremali

Spesso le statistiche d'ordine *estremali*, cioè $X_{(1)}$ e $X_{(n)}$, sono quelle di maggior interesse nelle applicazioni e il calcolo delle leggi del *minimo* e del *massimo* risulta essere immediato. Evidentemente si ha

$$X_{(1)} = \min\{X_1, \ldots, X_n\}, \tag{2.1a}$$

$$X_{(n)} = \max\{X_1, \ldots, X_n\}. \tag{2.1b}$$

Le *equazioni (2.1)* costituiscono il punto di partenza per il calcolo delle leggi di $X_{(1)}$ e di $X_{(n)}$.

OSSERVAZIONE 2.2 (Sistemi in serie e in parallelo)**.**
Si indichino con $X_1, \ldots, X_n$, $n \in \mathbb{N}$, la durata della vita degli elementi di un sistema a n componenti (formato, ad esempio, dalle lampade dell'*Esempio (2.1)*). Se gli elementi sono disposti *in serie* è sufficiente che se ne rompa uno affinché il sistema si guasti (e dunque la durata della vita (aleatoria) del circuito è pari a $X_{(1)}$). Se gli elementi sono disposti *in parallelo* è necessario che tutti si rompano affinché il sistema si guasti (e dunque la durata della vita (aleatoria) del circuito è pari a $X_{(n)}$).

TEOREMA 2.1 (Legge del minimo)**.** *Siano date le v.a. $X_1, \ldots, X_n$, $n \in \mathbb{N}$. Allora $X_{(1)}$ ha f.r. $F_{(1)}$ data da*

$$F_{(1)}(x) = \mathbf{P}\{X_{(1)} \leq x\} = 1 - \mathbf{P}\{X_1 > x, \ldots, X_n > x\}\,. \tag{2.2}$$

DIM. Evidentemente

$$\{X_{(1)} > x\} \equiv \{X_1 > x\} \cap \cdots \cap \{X_n > x\}\,,$$

da cui la tesi. □

COROLLARIO 2.1. *Sia $\mathbf{X}_{F_i}^n$ un campione. Allora $X_{(1)}$ ha f.r. $F_{(1)}$ data da*

$$F_{(1)}(x) = 1 - \prod_{i=1}^{n} \mathbf{P}\{X_i > x\} = 1 - \prod_{i=1}^{n} (1 - F_i(x))\,. \tag{2.3}$$

COROLLARIO 2.2. *Sia $\mathbf{X}_F^n$ un campione. Allora $X_{(1)}$ ha f.r. $F_{(1)}$ data da*

$$F_{(1)}(x) = 1 - \prod_{i=1}^{n} \mathbf{P}\{X_i > x\} = 1 - (1 - F(x))^n\,. \tag{2.4}$$

COROLLARIO 2.3. *Sia $\mathbf{X}_F^n$ un campione di v.a. assolutamente continue con densità comune f. Allora $X_{(1)}$ ha densità $f_{(1)}$ data da*

$$f_{(1)}(x) = n\ (1 - F(x))^{n-1}\ f(x). \tag{2.5}$$

TEOREMA 2.2 (Legge del massimo)**.** *Siano date le v.a. $X_1, \ldots, X_n$, $n \in \mathbb{N}$. Allora $X_{(n)}$ ha f.r. $F_{(n)}$ data da*

$$F_{(n)}(x) = \mathbf{P}\{X_{(n)} \leq x\} = \mathbf{P}\{X_1 \leq x, \ldots, X_n \leq x\}\,. \tag{2.6}$$

DIM. Evidentemente

$$\{X_{(n)} \leq x\} \equiv \{X_1 \leq x\} \cap \cdots \cap \{X_n \leq x\}\,,$$

da cui la tesi. □

COROLLARIO 2.4. *Sia $\mathbf{X}_{F_i}^n$ un campione. Allora $X_{(n)}$ ha f.r. $F_{(n)}$ data da*

$$F_{(n)}(x) = \prod_{i=1}^{n} \mathbf{P}\{X_i \leq x\} = \prod_{i=1}^{n} F_i(x). \tag{2.7}$$

COROLLARIO 2.5. *Sia $\mathbf{X}_F^n$ un campione. Allora $X_{(n)}$ ha f.r. $F_{(n)}$ data da*

$$F_{(n)}(x) = \prod_{i=1}^{n} \mathbf{P}\{X_i \leq x\} = (F(x))^n\,. \tag{2.8}$$

COROLLARIO 2.6. *Sia $\mathbf{X}_F^n$ un campione di v.a. assolutamente continue con densità comune f. Allora $X_{(n)}$ ha densità $f_{(n)}$ data da*

$$f_{(n)}(x) = n\ (F(x))^{n-1}\ f(x). \tag{2.9}$$

ESEMPIO 2.3. ►

Si consideri un campione $\mathbf{X}^n_{\mathsf{g}(p_i)}$, con $p_i \in (0,1)$ per $i = 1, \ldots, n$. Il calcolo della legge di $X_{(1)}$ può essere effettuato nel seguente modo. La generica v.a. X_i, $i = 1, \ldots, n$, ha legge

$$\mathbf{P}\{X_i = k\} = p_i\, q_i^{k-1}, \quad k = 1, 2, \ldots$$

dove $q_i = 1 - p_i$, e f.r. F_i data da

$$F_i(x) = \mathbf{P}\{X_i \leq x\} = 1 - q_i^{\lfloor x \rfloor}, \quad x > 0,$$

e nulla altrove. Usando il *Corollario (2.1)* avremo che, per $x > 0$,

$$F_{(1)}(x) = 1 - \prod_{i=1}^{n} (1 - F_i(x)) = 1 - \prod_{i=1}^{n} \left(1 - 1 + q_i^{\lfloor x \rfloor}\right) = 1 - q^{\lfloor x \rfloor},$$

dove $q = q_1 \cdots q_n$; dunque $X_{(1)} \sim \mathsf{g}(p)$, essendo $p = 1 - q$. ◄

OSSERVAZIONE 2.3.

L'*Esempio (2.3)* mostra che il *minimo* di un campione di v.a. indipendenti con legge *geometrica* è esso stesso una v.a. con legge *geometrica*. Inoltre, essendo $q < q_i$ per ogni $i = 1, \ldots, n$, ne segue che

$$\mathbf{P}\{X_{(1)} > k\} = q^k < q_i^k = \mathbf{P}\{X_i > k\}, \quad k = 1, 2, \ldots$$

Ricordando l'esempio del circuito in serie dell'*Osservazione (2.2)* se ne deduce che la probabilità che il sistema funzioni per un tempo superiore a k è minore della probabilità che il singolo elemento i-esimo funzioni per un tempo superiore a k stesso.

ESEMPIO 2.4. ►

Si consideri un campione $\mathbf{X}^n_{\mathsf{e}(\lambda_i)}$, con $\lambda_i > 0$ per $i = 1, \ldots, n$. Il calcolo della legge di $X_{(1)}$ può essere effettuato nel seguente modo. La generica v.a. X_i, $i = 1, \ldots, n$, ha f.r. F_i data da

$$F_i(x) = \mathbf{P}\{X_i \leq x\} = 1 - e^{-\lambda_i x}, \quad x > 0,$$

e nulla altrove. Usando il *Corollario (2.1)* avremo che, per $x > 0$,

$$F_{(1)}(x) = 1 - \prod_{i=1}^{n} (1 - F_i(x)) = 1 - \prod_{i=1}^{n} \left(1 - 1 + e^{-\lambda_i x}\right) = 1 - e^{-\lambda x},$$

dove $\lambda = \lambda_1 + \cdots + \lambda_n$; dunque $X_{(1)} \sim \mathsf{e}(\lambda)$. ◄

OSSERVAZIONE 2.4.

L'*Esempio (2.4)* mostra che il *minimo* di un campione di v.a. indipendenti con legge *Esponenziale* è esso stesso una v.a. con legge *Esponenziale*. Inoltre, essendo $\lambda > \lambda_i$ per ogni $i = 1, \ldots, n$, ne segue che

$$\mathbf{P}\{X_{(1)} > x\} = e^{-\lambda x} < e^{-\lambda_i x} = \mathbf{P}\{X_i > x\}, \quad x > 0.$$

Ricordando l'esempio del circuito in serie dell'*Osservazione (2.2)* se ne deduce che la probabilità che il sistema funzioni per un tempo superiore a x è minore della probabilità che il singolo elemento i-esimo funzioni per un tempo superiore a x stesso.

2.3 Leggi delle statistiche d'ordine

Calcoliamo ora la legge della generica statistica d'ordine.

TEOREMA 2.3 (Legge della singola statistica d'ordine)**.** *Sia $\mathbf{X}^n_F$ un campione. Allora $X_{(i)}$, $i = 1, \ldots, n$, ha f.r. $F_{(i)}$ data da*

$$F_{(i)}(x) = \mathbf{P}\{X_{(i)} \leq x\} = \sum_{j=i}^{n} \binom{n}{j} \left(F(x)\right)^j \left(1 - F(x)\right)^{n-j}. \tag{2.10}$$

DIM. Evidentemente

$$\left\{X_{(i)} \leq x\right\} \equiv \left\{X_j \leq x \text{ per almeno } i \text{ delle } n \text{ v.a. } X_j\right\},$$

da cui la tesi.

In alternativa, per x fissato e $j = 1, \ldots, n$, si definiscano le v.a. $Y_j = \mathbb{I}_{(-\infty,x]}(X_j)$; dunque $Y_j \sim \texttt{Be}(F(x))$, cioè $Y_j = 0$ se $X_j > x$ (che avviene con probabilità $1 - F(x)$) e $Y_j = 1$ se $X_j \leq x$ (che avviene con probabilità $F(x)$). Ne segue che $Y = \sum_{j=1}^{n} Y_j \sim \texttt{Bi}(n, F(x))$ e inoltre $Y = \#(X_j \leq x)$. Dalla equivalenza degli eventi

$$\left\{X_{(i)} \leq x\right\} \equiv \{Y \geq i\}$$

si ha la tesi. □

OSSERVAZIONE 2.5.

Ponendo $p = F(x)$ nell'*equazione (2.10)* si ottiene

$$F_{(i)}(x) = \sum_{j=i}^{n} \binom{n}{j} p^j \ (1-p)^{n-j}.$$

Usando la funzione *Beta Incompleta* introdotta nella *Sezione 0.4* se ne deduce che

$$F_{(i)}(x) = \frac{\mathrm{B}_p(i, n-i+1)}{\mathrm{B}(i, n-i+1)}, \tag{2.10′}$$

che si dimostra essere equivalente all'*equazione (2.10)* tramite una tediosa integrazione per parti.

COROLLARIO 2.7. *Sia $\mathbf{X}_F^n$ un campione di v.a. assolutamente continue con densità comune f. Allora $X_{(i)}$, $i = 1, \ldots, n$, ha densità $f_{(i)}$ data da*

$$f_{(i)}(x) = \frac{1}{\mathrm{B}(i, n-i+1)} \ (F(x))^{i-1} \ (1-F(x))^{n-i} \ f(x). \tag{2.11}$$

DIM. Usando il risultato dell'*Osservazione (2.5)* si ha che

$$f_{(i)}(x) = \frac{1}{\mathrm{B}(i, n-i+1)} \frac{\mathrm{d}}{\mathrm{d}x} \int_{0^+}^{F(x)} t^{i-1} (1-t)^{n-i+1-1} \ \mathrm{d}t,$$

da cui la tesi. □

OSSERVAZIONE 2.6.

Euristicamente il risultato del *Corollario (2.7)* può essere ottenuto ragionando come segue. L'evento

$$\left\{x < X_{(i)} \leq x + \mathrm{d}x\right\}$$

si realizza quando $i-1$ delle n v.a. X_i sono minori di x, esattamente una è compresa tra x e $x+\mathrm{d}x$, e le rimanenti $n-i$ sono maggiori di $x + \mathrm{d}x$. Usando il fatto che il campione è composto di v.a. i.i.d. ne segue che vi sono esattamente

$$\frac{n!}{(i-1)!\,1!\,(n-i)!} = \binom{n}{i-1} \binom{n-(i-1)}{1} \binom{n-i}{n-i}$$

modi per realizzare tale evento, e ciascuna di tali modalità ha probabilità

$$(F(x))^{i-1} \left(F(x+\mathrm{d}x) - F(x)\right) (1 - F(x+\mathrm{d}x))^{n-i}.$$

Dividendo per $\mathrm{d}x$ e prendendo il limite per $\mathrm{d}x \to 0$ si ottiene il risultato desiderato. Il *Corollario (2.3)* e il *Corollario (2.6)* sono casi particolari del *Corollario (2.7)*.

ESEMPIO 2.5 (Sistemi i-su-n). ►

Si consideri un sistema formato da n componenti identici e indipendenti, ciascuno avente una durata pari a X, e si supponga che tale sistema vada fuori uso non appena i di questi componenti si rompono, con $1 \leq i \leq n$ (si pensi, ad esempio, al numero di scarichi ostruiti di una diga). Sistemi di questo tipo si chiamano *i-su-n*. Ora, è chiaro che la durata del sistema è pari a $X_{(i)}$, ed è dunque importante saperne calcolare la legge. ◄

Una proprietà interessante delle statistiche d'ordine riguarda il loro legame con i quantili (si veda la *Sezione 0.3*). Per semplicità assumeremo che X sia una v.a. *assolutamente continua* con f.r. F strettamente crescente; si osservi che in questo caso la disuguaglianza dell'*equazione (14c)* si trasforma nell'uguaglianza $F(x_p) = p$. Fissato $p \in (0,1)$ l'evento $\{X_{(i)} \leq x_p\}$ si può scrivere come

$$\{X_{(i)} \leq x_p\} = \{X_{(i)} \leq x_p, X_{(j)} \geq x_p\} \cup \{X_{(i)} \leq x_p, X_{(j)} < x_p\},$$

dove $i < j$. Poiché l'unione è disgiunta e (q.c.) $X_{(j)} < x_p$ implica $X_{(i)} \leq x_p$ si ha

$$\mathbf{P}\{X_{(i)} \leq x_p\} = \mathbf{P}\{X_{(i)} \leq x_p \leq X_{(j)}\} + \mathbf{P}\{X_{(j)} < x_p\},$$

da cui

$$\mathbf{P}\{X_{(i)} \leq x_p \leq X_{(j)}\} = \mathbf{P}\{X_{(i)} \leq x_p\} - \mathbf{P}\{X_{(j)} < x_p\}.$$

Usando il *Teorema (2.3)* è possibile dimostrare il seguente importante risultato.

PROPOSIZIONE 2.1. *L'*intervallo aleatorio $(X_{(i)}, X_{(j)})$, $i < j$, *contiene il quantile* x_p, $p \in (0,1)$, *con probabilità data da*

$$\mathbf{P}\{X_{(i)} \leq x_p \leq X_{(j)}\} = F_{(i)}(x_p) - F_{(j)}(x_p) = \sum_{k=i}^{j} \binom{n}{k} p^k \, (1-p)^{n-k}, \tag{2.12}$$

dove si è assunto che il campione sia di ampiezza n.

NOTA 2.1.

Si osservi che la probabilità calcolata nella *Proposizione (2.1) non* dipende in alcun modo dalla distribuzione F di X; dunque l'intervallo aleatorio $(X_{(i)}, X_{(j)})$ rappresenta una "stima (debole)" di x_p *indipendente* dalla legge considerata (ovvero una stima "non parametrica").

Il calcolo della *legge congiunta* di due (o più) statistiche d'ordine procede sulla falsariga di quanto appena visto nel caso di una singola statistica d'ordine.

TEOREMA 2.4 (Legge congiunta di coppie di statistiche d'ordine)**.** *Sia* $\mathbf{X}_F^n$ *un campione di v.a. assolutamente continue con densità comune* f. *Allora le statistiche d'ordine* $X_{(i)}$ *e* $X_{(j)}$, $1 \leq i < j \leq n$, *hanno densità congiunta* $f_{(i,j)}$ *data da*

$$\begin{aligned} f_{(i,j)}(x,y) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \, & (F(x))^{i-1} \, f(x) \cdot \\ & \cdot (F(y) - F(x))^{j-i-1} \, (1 - F(y))^{n-j} \, f(y) \end{aligned} \tag{2.13}$$

per $x < y$, *e nulla altrove.*

Grazie al *Teorema (2.4)* è possibile calcolare la *densità condizionale* della statistica d'ordine $X_{(i)}$ dato l'evento $\{X_{(j)} = y\}$, con $i < j$. Tale calcolo è di importanza fondamentale, ad esempio in fase di simulazione.

TEOREMA 2.5 (Legge condizionale di coppie di statistiche d'ordine)**.** *Sia* $\mathbf{X}_F^n$ *un campione di v.a. assolutamente continue con densità comune* f. *Allora la densità condizionale della statistica d'ordine* $X_{(i)}$ *dato l'evento* $\{X_{(j)} = y\}$, $1 \leq i < j \leq n$, *vale*

$$\begin{aligned} f_{(i|j)}(x \mid y) &= \frac{f_{(i,j)}(x,y)}{f_{(j)}(y)} \\ &= \frac{(j-1)!}{(i-1)!(j-i-1)!} \, (F(x))^{i-1} \, f(x) \, (F(y) - F(x))^{j-i-1} \, (F(y))^{1-j} \end{aligned} \tag{2.14}$$

per $x < y$, *e nulla altrove.*

NOTA 2.2.

L'*equazione (2.14)* mostra che la distribuzione condizionale di $X_{(i)}$ dato l'evento $\{X_{(j)} = y\}$ è la stessa di quella della i-esima statistica d'ordine, in un campione di ampiezza $j-1$, estratta dalla legge F, troncata a destra in y.

Il calcolo della *legge congiunta* di tutte le n statistiche d'ordine si basa sull'osservazione che le $n!$ permutazioni delle n v.a. X_i, $i = 1, \ldots, n$, sono egualmente probabili.

TEOREMA 2.6 (Legge congiunta di tutte le statistiche d'ordine)**.** *Sia $\mathbf{X}_F^n$ un campione di v.a. assolutamente continue con densità comune f. Allora le statistiche d'ordine $\left\{X_{(i)}\right\}$, $i = 1, \ldots, n$, hanno densità congiunta $f_{(1,\ldots,n)}$ data da*

$$f_{(1,\ldots,n)}(x_1, \ldots, x_n) = n! \prod_{i=1}^{n} f(x_i) \tag{2.15}$$

per $x_1 < \cdots < x_n$, e nulla altrove.

NOTA 2.3.

Il risultato del *Teorema (2.6)* può essere utilizzato per ottenere la legge congiunta di un qualsiasi insieme di statistiche d'ordine: infatti è sufficiente integrare $f_{(1,\ldots,n)}$ rispetto alle variabili che si desidera eliminare, ottenendo così la legge voluta quale legge marginale.

Grazie al *Teorema (2.4)* è possibile calcolare la legge della *differenza* di due statistiche d'ordine qualsiasi.

PROPOSIZIONE 2.2 (Legge della differenza di statistiche d'ordine)**.** *Sia $\mathbf{X}_F^n$ un campione di v.a. assolutamente continue con densità comune f. Allora la legge della differenza $V = X_{(j)} - X_{(i)}$, $1 \leq i < j \leq n$, vale*

$$\begin{aligned} f_V(v) = {} & \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \cdot \\ & \cdot \int_{\mathbb{R}} (F(u))^{i-1} \, f(u) \, (F(u+v) - F(u))^{j-i-1} \, (1 - F(u+v))^{n-j} \, f(u+v) \, \mathrm{d}u \end{aligned} \tag{2.16}$$

per $v > 0$, e nulla altrove.

DIM. Useremo il *metodo della variabile ausiliaria*. Si consideri la trasformazione invertibile

$$\begin{cases} U = X_{(i)} \\ V = X_{(j)} - X_{(i)} \end{cases} \iff \begin{cases} X_{(i)} = U \\ X_{(j)} = U + V \end{cases},$$

avente determinante Jacobiano $J = \left|\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right| = 1$. Ne segue che la densità congiunta $g(u, v)$ del vettore (U, V) è data da

$$\begin{aligned} g(u,v) = {} & \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \cdot \\ & \cdot (F(u))^{i-1} \, f(u) \, (F(u+v) - F(u))^{j-i-1} \, (1 - F(u+v))^{n-j} \, f(u+v) \end{aligned}$$

per $v > 0$; dunque basta integrare rispetto a u per ottenere il risultato enunciato. □

ESEMPIO 2.6 (Legge del Range)**.** ►

Nelle applicazioni è spesso di interesse la variabile

$$R = X_{(n)} - X_{(1)} \tag{2.17}$$

detta *Range* (o *campo di variazione*); in pratica essa fornisce una stima approssimativa dell'ampiezza dell'intervallo dei valori assumibili dal fenomeno osservato (si veda a tale proposito l'*Esempio (4.6)*). Grazie alla *Proposizione (2.2)* è immediato calcolare la distribuzione di R. Infatti, ponendo $i = 1$ e $j = n$ nell'*equazione (2.16)*, si ottiene

$$f_R(r) = n(n-1) \int_{\mathbb{R}} f(u) \, (F(u+r) - F(u))^{n-2} \, f(u+r) \, \mathrm{d}u,$$

$$F_R(r) = \int_{-\infty}^{r} f_R(t) \, \mathrm{d}t = n \int_{\mathbb{R}} f(u) \, (F(u+r) - F(u))^{n-1} \, \mathrm{d}u$$

per $r > 0$. ◄

ESEMPIO 2.7 (Plotting positions). ►
Nella pratica capita spesso di dover confrontare la funzione di ripartizione "teorica" F di un certo campione (ovvero quella presunta) con le osservazioni disponibili $\{X_i\}$. Più in dettaglio, dopo aver ordinato il campione in senso crescente (cioè dopo aver calcolato le statistiche d'ordine $\{X_{(i)}\}$), il problema consiste nel valutare le differenze tra i valori "teorici" $F(x_{(i)})$ e quelli "campionari" $C(x_{(i)})$, dove C è una opportuna funzione dell'indice i e dell'ampiezza del campione n che identifica le cosiddette *plotting positions*.

In Letteratura (si vedano, ad esempio, [17, 26]) esistono numerose possibilità per quanto concerne la scelta di C, ciascuna con i propri pregi e difetti. In particolare, le cosiddette *plotting positions standard* sono tra le più usate nella pratica; esse sono definite tramite la sequenza di ordinate

$$y_i = \frac{i}{n+1}, \tag{2.18}$$

con $i = 1, \ldots, n$; si noti che $0 < y_i < 1$. La ragione per cui si sceglie di usare tale sequenza è semplice: usando il *Corollario (2.7)*, il calcolo di $\mathbf{E}\big(F\left(X_{(i)}\right)\big)$ mostra che

$$\begin{aligned}
\mathbf{E}\big(F\left(X_{(i)}\right)\big) &= \int_{\mathbb{R}} F(x)\, f_{(i)}(x)\, \mathrm{d}x \\
&= \frac{1}{\mathrm{B}(i, n-i+1)} \int_{\mathbb{R}} (F(x))^i\, (1-F(x))^{n-i}\, f(x)\, \mathrm{d}x \quad \bigstar\ t = F(x) \\
&= \frac{1}{\mathrm{B}(i, n-i+1)} \int_0^1 t^i\, (1-t)^{n-i}\, \mathrm{d}t \\
&= \frac{\mathrm{B}(i+1, n-i+1)}{\mathrm{B}(i, n-i+1)} \\
&= \frac{\Gamma(i+1)\,\Gamma(n-i+1)}{\Gamma(n+2)}\, \frac{\Gamma(n+1)}{\Gamma(i)\,\Gamma(n-i+1)} \\
&= \frac{i!\,(n-i)!}{(n+1)!}\, \frac{n!}{(i-1)!\,(n-i)!} \\
&= \frac{i}{n+1}.
\end{aligned}$$

Dunque, *in media*, $F\left(X_{(i)}\right) = \frac{i}{n+1}$, ciò che giustifica la scelta empirica delle coordinate $\left\{(x_{(i)}, y_i)\right\}$ per confrontare il comportamento statisico previsto dalla funzione di ripartizione "teorica" F con quello effettivamente mostrato dal campione $\{X_i\}$. Si osservi come le plotting positions non dipendano da F, ossia forniscano una sorta di stima "non parametrica" della funzione di ripartizione (non nota) del campione; in particolare, esse saranno usate nelle figure del *Capitolo 3* e del *Capitolo 8*. La *Tabella (2.1)* mostra alcune delle plotting positions più utilizzate nelle applicazioni. ◄

Plotting position	$\widehat{\mathbf{F}}_\mathbf{i}$	**Plotting position**	$\widehat{\mathbf{F}}_\mathbf{i}$
Adamowski	$\frac{i-0.26}{n+0.5}$	**Gringorten**	$\frac{i-0.44}{n+0.12}$
Blom	$\frac{i-3/8}{n+1/4}$	**Hazen**	$\frac{i-1/2}{n}$
California	$\frac{i-1}{n}$	**Hosking (APL)**	$\frac{i-0.35}{n}$
Chegodayev	$\frac{i-0.3}{n+0.4}$	**Tukey**	$\frac{i-1/3}{n+1/3}$
Cunanne	$\frac{i-1/5}{n+2/5}$	**Weibull**	$\frac{i}{n+1}$

Tabella 2.1: espressioni di alcune delle plotting positions ($\widehat{F}_i$) più utilizzate nelle applicazioni.

Capitolo 3

TEORIA DEI VALORI ESTREMI

Quanto visto nel *Capitolo 2* circa le statistiche d'ordine estremali potrebbe indurre a pensare che, nelle applicazioni, (quasi) tutti i problemi concernenti i valori estremi di un fenomeno siano facilmente risolvibili una volta date la distribuzione F di X e l'ampiezza n del campione disponibile. Sfortunatamente queste condizioni non sempre si realizzano, per almeno tre ragioni:

1. la distribuzione F di X può non essere nota;
2. l'ampiezza n del campione può non essere nota;
3. l'ampiezza n del campione può divergere.

Il caso *(1)* è frequente nelle applicazioni, in quanto quasi mai è possibile conoscere la vera distribuzione "parente" F. Solitamente si aggira il problema decidendo a priori che X abbia una certa legge; tuttavia è interessante domandarsi quali metodi siano applicabili nel caso in cui F sia incognita, e quale sia la loro credibilità ed efficacia. Il caso *(2)* può, a prima vista, apparire paradossale: come può non essere nota l'ampiezza n del campione una volta raccolti i dati? Tuttavia è possibile che lo strumento di misura utilizzato campioni il fenomeno solo al di sopra (o al di sotto) di una certa soglia — si parla, in tal caso, di *dati censurati* — e dunque che una parte delle rilevazioni vada persa; perciò è interessante chiedersi quanto possa influire la mancata conoscenza di n, e cosa possa essere fatto per affrontare il problema. Il caso *(3)* nasce dal fatto che, data la distribuzione F,

$$\lim_{n\to\infty} \mathbf{P}\big\{X_{(n)} \leq x\big\} = \lim_{n\to\infty} F^n(x) = \begin{cases} 1, & \text{se } F(x) = 1 \\ 0, & \text{se } F(x) < 1 \end{cases}, \tag{3.1a}$$

$$\lim_{n\to\infty} \mathbf{P}\big\{X_{(1)} \leq x\big\} = \lim_{n\to\infty} 1 - (1 - F(x))^n = \begin{cases} 0, & \text{se } F(x) = 0 \\ 1, & \text{se } F(x) > 0 \end{cases}. \tag{3.1b}$$

Dunque, apparentemente, i limiti delle distribuzioni del massimo e del minimo (nel caso i.i.d.) sono degeneri! Ciò conduce ad una serie di domande interessanti, quali ad esempio: (*a*) esiste una distribuzione limite non degenere per i massimi e i minimi? e se la risposta è positiva, (*b*) quante distribuzioni limite esistono? (*c*) quali sono? (*d*) possono due diverse distribuzioni "parenti" avere la stessa distribuzione limite? e infine (*e*) come è possibile calcolare la distribuzione limite associata ad una data distribuzione "parente"?

In questo Capitolo cercheremo di rispondere ad alcune delle domande appena poste (si vedano anche [7, 11, 3, 17, 18, 8, 26] e i lavori ivi citati). Al fine di semplificare la trattazione assumeremo che le distribuzioni in gioco siano (almeno) continue e che i campioni disponibili siano formati da v.a. i.i.d.; in particolare, la convergenza in legge (o *convergenza debole* — si veda la *Definizione (0.10)*) potrà essere sostituita da quella puntuale (o, addirittura, da quella uniforme).

OSSERVAZIONE 3.1.

Nel seguito ci limiteremo allo studio dei massimi di una distribuzione, in quanto i risultati ottenuti per questi ultimi sono di applicazione immediata (tramite facili modifiche) ai corrispondenti minimi: infatti, operando la semplice sostituzione

$$Y_i = -X_i, \quad i = 1, \ldots, n, \tag{3.2}$$

è evidente che tra le variabili $X_{(1)} = \min\{X_1, \ldots, X_n\}$ e $Y_{(n)} = \max\{Y_1, \ldots, Y_n\}$ sussiste la relazione $X_{(1)} = -Y_{(n)}$. Dunque sarà sufficiente studiare solamente i massimi, ovvero $-\max\{Y_1, \ldots, Y_n\}$.

Premettiamo una definizione importante.

DEFINIZIONE 3.1 (Limite inferiore e superiore di una distribuzione)**.** Il *limite inferiore* α_F della f.r. F è dato da

$$\alpha_F = \inf\{x\colon F(x) > 0\}. \tag{3.3a}$$

Il *limite superiore* ω_F della f.r. F è dato da

$$\omega_F = \sup\{x\colon F(x) < 1\}. \tag{3.3b}$$

In termini pratici α_F e ω_F rappresentano, rispettivamente, i valori minimi e massimi assumibili dalla v.a. X associata alla f.r. F; evidentemente può essere $\alpha_F = -\infty$ e/o $\omega_F = +\infty$ nel caso di distribuzioni non limitate.

3.1 Modelli "a blocchi"

Il titolo di questa Sezione deriva essenzialmente dalla procedura con cui, nella pratica, solitamente si studia il comportamento estremo di un fenomeno, come vedremo meglio più avanti nella *Nota (3.4)*. Le *equazioni (3.1)* mostrano che la ricerca delle possibili distribuzioni limite per i massimi (e, parimenti, per i minimi) debba essere abbastanza accurata onde evitare problemi di degenerazione. A tale fine è possibile studiare se esistono successioni di costanti $\{a_n\}, \{b_n > 0\}$, $n \in \mathbb{N}$, tali che

$$G(x) = \lim_{n\to\infty} F^n(a_n + b_n x), \quad x \in \mathbb{R}, \tag{3.4}$$

sia effettivamente una distribuzione non degenere. In altri termini, usando un procedimento classico della Statistica Matematica (si pensi, ad esempio, al *Teorema del Limite Centrale*), si cerca di vedere se, rinormalizzando opportunamente le variabili in gioco, sia possibile ottenere qualche forma di convergenza.

DEFINIZIONE 3.2 (Dominio di attrazione per i massimi di una distribuzione)**.** Si dice che la f.r. F appartiene al (o giace nel) *dominio di attrazione per i massimi* della f.r. G se soddisfà all'*equazione (3.4)* per opportune successioni $\{a_n\}$ e $\{b_n > 0\}$.

Una definizione analoga può essere data per i minimi. Lo studio delle distribuzioni limite può dunque ridursi ai seguenti punti: (*a*) stabilire le condizioni sotto cui valga l'*equazione (3.4)*; (*b*) fornire regole per la costruzione delle successioni $\{a_n\}$ e $\{b_n\}$; (*c*) trovare quali f.r. possano giocare il ruolo del limite G nell'*equazione (3.4)*. Il Teorema che segue è fondamentale; per semplicità adotteremo la notazione $M_n = \max\{X_1, \ldots, X_n\}$.

TEOREMA 3.1 (Distribuzioni del valore estremo per i massimi)**.** *Se esistono successioni* $\{a_n\}$ *e* $\{b_n > 0\}$ *tali che, per* $z \in \mathbb{R}$*,*

$$\lim_{n\to\infty} \mathbf{P}\left\{\frac{M_n - a_n}{b_n} \le z\right\} = G(z), \tag{3.5}$$

dove G *è una f.r. non degenere, allora* G *appartiene ad una delle seguenti* distribuzioni del valore estremo per i massimi*:*

I. ***Tipo I (Gumbel)***

$$G(z) = \exp\left\{-\exp\left(-\frac{z-a}{b}\right)\right\}, \quad -\infty < z < \infty; \tag{3.6}$$

II. ***Tipo II (Fréchet)***

$$G(z) = \begin{cases} 0, & z \le a \\ \exp\left\{-\left(\frac{z-a}{b}\right)^{-\gamma}\right\}, & z > a \end{cases}; \tag{3.7}$$

III. ***Tipo III (Weibull (inversa))***

$$G(z) = \begin{cases} \exp\left\{-\left[\left(-\frac{z-a}{b}\right)^{\gamma}\right]\right\}, & z < a \\ 1, & z \geq a \end{cases}. \tag{3.8}$$

Qui $a \in \mathbb{R}$ *è un parametro di* posizione*,* $b > 0$ *è un parametro di* scala *e* $\gamma > 0$ *è un parametro di* forma*.*

Si osservi che la formulazione data nel *Teorema (3.1)* è solo una tra le molte possibili, così come diverse sono in Letteratura le convenzioni circa i parametri a, b, γ.

NOTA 3.1.

Il *Teorema (3.1)* fornisce un analogo del *Teorema del Limite Centrale* nell'ambito della Teoria dei Valori Estremi: in buona sostanza esso afferma che le tre distribuzioni date sono i soli limiti possibili per la legge di M_n (una volta riscalato grazie ad opportune trasformazioni affini) indipendentemente dalla distribuzione "parente" F. Tuttavia non tutte le distribuzioni F soddisfanno alle ipotesi del *Teorema (3.1)*, come ad esempio la legge

$$F(x) = 1 - \frac{1}{\ln x}, \quad x > e.$$

Nelle *Figure (3.1)–(3.3)* sono mostrati i grafici delle tre distribuzioni limite per i massimi in *forma canonica* (ovvero con $a = 0$ e $b = 1$) e delle rispettive funzioni dei quantili, per diversi valori del parametro di forma.

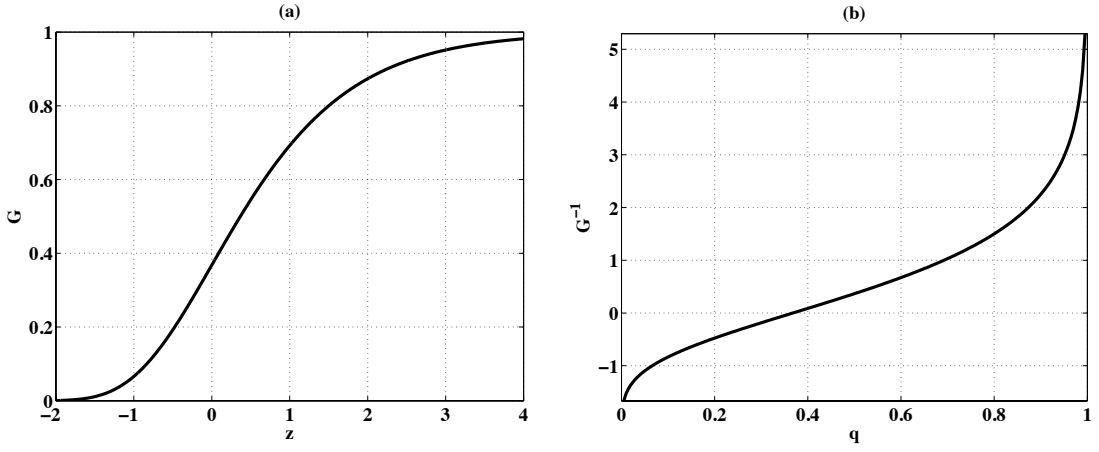

Figura 3.1: (*a*) Funzione di ripartizione della legge Gumbel; (*b*) funzione dei quantili della legge Gumbel.

OSSERVAZIONE 3.2.

L'analisi del comportamento di G in un opportuno intorno di ω_G mostra come le tre distribuzioni limite date nel *Teorema (3.1)* abbiano comportamenti asintotici alquanto differenti, corrispondenti ad andamenti diversi della coda destra della distribuzione "parente" F. Infatti, considerando senza perdita di generalità la *forma canonica* delle varie distribuzioni limite G (ovvero con $a = 0$ e $b = 1$), si ottiene quanto segue.

I. Per la distribuzione di Gumbel $\omega_G = +\infty$ e la densità corrispondente, ovvero

$$g(z) \propto \exp\left\{-\exp\left(-z\right)\right\} \exp\left(-z\right), \tag{3.9}$$

ha una caduta "esponenziale" per $z \gg 1$.

II. Per la distribuzione di Fréchet $\omega_G = +\infty$ e la densità corrispondente, ovvero

$$g(z) \propto \exp\left\{-z^{-\gamma}\right\} z^{-\gamma-1}, \tag{3.10}$$

ha una caduta "algebrica" per $z \gg 1$ (e si parla, in tal caso, di distribuzioni a "coda pesante"); si noti, in particolare, che non esistono i momenti di ordine $m \geq \gamma > 0$.

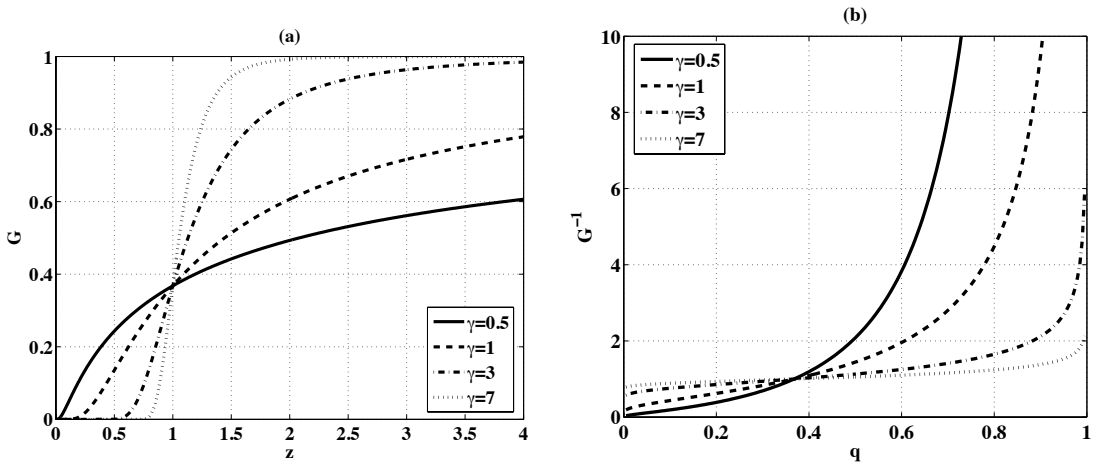

Figura 3.2: (*a*) Funzione di ripartizione della legge Fréchet per diversi valori del parametro di forma; (*b*) funzione dei quantili della legge Fréchet.

III. Per la distribuzione di Weibull $\omega_G < +\infty$ e la densità corrispondente è nulla per $z > \omega_G$.

Tali differenze si riflettono in comportamenti profondamente diversi dei valori estremi associati, come mostrato negli Esempi presentati più avanti.

Il *Teorema (3.1)* può essere riformulato in modo compatto introducendo la distribuzione *Generalized Extreme Value (GEV)*, anche detta *forma di von Mises*, di grande utilità nelle applicazioni.

TEOREMA 3.2 (Distribuzione Generalized Extreme Value per i massimi)**.** *Nelle ipotesi del Teorema (3.1), la distribuzione G è un membro della famiglia GEV per i massimi definita sull'insieme $\left\{z\colon 1+\gamma'\,\frac{z-a'}{b'}>0\right\}$:*

$$G(z) = \exp\left\{-\left[1+\gamma'\,\frac{z-a'}{b'}\right]^{-1/\gamma'}\right\}. \tag{3.11}$$

Qui $a' \in \mathbb{R}$ è un parametro di posizione*, $b' > 0$ è un parametro di* scala *e $\gamma' \in \mathbb{R}$ è un parametro di* forma*. Il caso limite $\gamma' = 0$ corrisponde (si riduce) alla famiglia di Gumbel data dall'*equazione *(3.6).*

NOTA 3.2.

Si osservi che per $\gamma' < 0$ si ottiene la famiglia di Weibull, mentre per $\gamma' > 0$ si ottiene la famiglia di Fréchet.

Per completezza, diamo di seguito l'analogo del *Teorema (3.2)* per i minimi.

TEOREMA 3.3 (Distribuzione Generalized Extreme Value per i minimi)**.** *Sia $\widetilde{M}_n = \min\{X_1, \ldots, X_n\}$. Se esistono successioni $\{c_n\}$ e $\{d_n > 0\}$ tali che, per $z \in \mathbb{R}$,*

$$\lim_{n\to\infty} \mathbf{P}\left\{\frac{\widetilde{M}_n - c_n}{d_n} \le z\right\} = \widetilde{G}(z), \tag{3.12}$$

dove $\widetilde{G}$ è una f.r. non degenere, allora $\widetilde{G}$ è un membro della famiglia GEV per i minimi definita sull'insieme $\left\{z\colon 1-\tilde{\gamma}\,\frac{z-\tilde{c}}{\tilde{d}}>0\right\}$:

$$\widetilde{G}(z) = 1 - \exp\left\{-\left[1-\tilde{\gamma}\,\frac{z-\tilde{c}}{\tilde{d}}\right]^{-1/\tilde{\gamma}}\right\}. \tag{3.13}$$

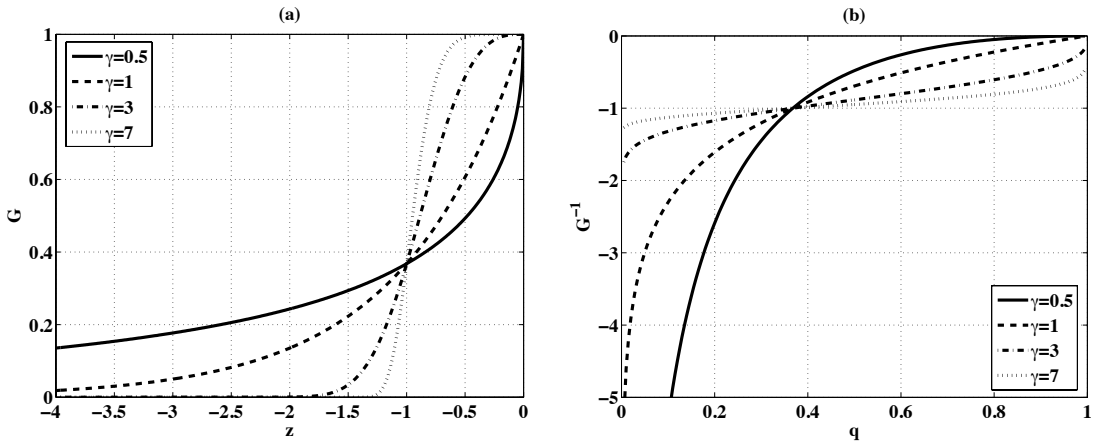

Figura 3.3: (*a*) Funzione di ripartizione della legge Weibull per diversi valori del parametro di forma; (*b*) funzione dei quantili della legge Weibull.

Qui $\tilde{c} \in \mathbb{R}$ *è un parametro di* posizione*,* $\tilde{d} > 0$ *è un parametro di* scala *e* $\tilde{\gamma} \in \mathbb{R}$ *è un parametro di* forma*. Il caso limite* $\tilde{\gamma} = 0$ *corrisponde (si riduce) alla famiglia di Gumbel data da*

$$\widetilde{G}(z) = 1 - \exp\left\{-\exp\left(\frac{z - \tilde{c}}{\tilde{d}}\right)\right\}. \tag{3.14}$$

Nella *Tabella (3.1)* si riporta il dominio di attrazione, rispettivamente, per il massimo ed il minimo di alcune tra le distribuzioni più utilizzate nelle applicazioni; per la loro identificazione si vedano alcuni dei Teoremi presentati più avanti.

Distribuzione	**Massimo**	**Minimo**	**Distribuzione**	**Massimo**	**Minimo**
Normale	Gumbel	Gumbel	**Rayleigh**	Gumbel	Weibull
Esponenziale	Gumbel	Weibull	**LogNormale**	Gumbel	Gumbel
Gamma	Gumbel	Weibull	**Uniforme**	Weibull	Weibull
Cauchy	Fréchet	Fréchet	**Pareto**	Fréchet	Weibull
Gumbel (Max)	Gumbel	Gumbel	**Gumbel (Min)**	Gumbel	Gumbel
Fréchet (Max)	Fréchet	Gumbel	**Fréchet (Min)**	Gumbel	Fréchet
Weibull (Max)	Weibull	Gumbel	**Weibull (Min)**	Gumbel	Weibull

Tabella 3.1: dominio di attrazione per il massimo ed il minimo di alcune distribuzioni.

La dimostrazione del *Teorema (3.2)* (o del *Teorema (3.1)*) è assai tecnica e va oltre gli scopi di questo Corso; tuttavia è possibile dare una giustificazione informale dei risultati mostrati, come illustrato di seguito. A tale fine premettiamo la seguente nozione.

DEFINIZIONE 3.3 (Distribuzione max-stabile)**.** Si dice che la distribuzione G è *max-stabile* se, per ogni $n \in \mathbb{N}$, esistono costanti a_n e $b_n > 0$ tali che, per ogni $z \in \mathbb{R}$,

$$G^n(a_n + b_n z) = G(z). \tag{3.15}$$

NOTA 3.3.

Poiché G^n rappresenta la distribuzione della v.a. M_n, dove il campione corrispondente è formato da v.a. i.i.d. aventi f.r. comune G, è evidente che la proprietà di max-stabilità è soddisfatta da tutte quelle distribuzioni per le quali il calcolo della legge del massimo fornisce, a meno di una trasformazione affine, la stessa distribuzione "parente" G.

Il legame tra la famiglia GEV e le distribuzioni max-stabili è sancito dal Teorema che segue.

TEOREMA 3.4. *Una distribuzione è max-stabile se, e solo se, appartiene alla famiglia GEV.*

Non è difficile mostrare che i membri della famiglia GEV godono della proprietà di max-stabilità: infatti, è sufficiente trovare costanti a_n e $b_n > 0$ che rendano vera l'*equazione (3.15)*, ovvero (usando, senza perdita di generalità, la forma canonica della distribuzione GEV)

$$G^n(a_n + b_n z) = \exp\left\{-n\,\left[1 + \gamma'\,(a_n + b_n z)\right]^{-1/\gamma'}\right\} = \exp\left\{-\left[1 + \gamma'\,z\right]^{-1/\gamma'}\right\} = G(z),$$

da cui la semplice relazione

$$n^{-\gamma'}\,\left(1 + \gamma'\,(a_n + b_n z)\right) = 1 + \gamma'\,z.$$

Invece, la prova dell'implicazione nell'altro senso richiede tecniche di Analisi Funzionale e va oltre gli scopi di questo Corso.

Il *Teorema (3.4)* entra il gioco direttamente nella dimostrazione del *Teorema (3.2)* (o del *Teorema (3.1)*): l'idea è quella di considerare la v.a. M_{nk}, ovvero il massimo di un campione di ampiezza nk, con $n \gg 1$ e $k \in \mathbb{N}$, e di "interpretarla" in modi differenti. Evidentemente M_{nk} può essere considerata (*i*) sia come il massimo di nk variabili (*ii*) sia come il massimo di k massimi ciascuno calcolato su un campione di ampiezza n. Ora, si supponga che esista e sia non degenere la distribuzione (limite) G di $\frac{M_n - a_n}{b_n}$; dunque si ha, per $n \gg 1$,

$$\mathbf{P}\left\{\frac{M_n - a_n}{b_n} \leq z\right\} \approx G(z).$$

Analogamente, essendo $nk \geq n \gg 1$, vale

$$\mathbf{P}\left\{\frac{M_{nk} - a_{nk}}{b_{nk}} \leq z\right\} \approx G(z).$$

Tuttavia, poiché M_{nk} è il massimo di k variabili aventi la stessa legge di M_n, si ha

$$\mathbf{P}\left\{\frac{M_{nk} - a_{nk}}{b_{nk}} \leq z\right\} \approx \left(\mathbf{P}\left\{\frac{M_n - a_n}{b_n} \leq z\right\}\right)^k.$$

Ne segue che

$$\mathbf{P}\{M_{nk} \leq z\} \approx G\left(\frac{z - a_{nk}}{b_{nk}}\right) \quad \text{e} \quad \mathbf{P}\{M_{nk} \leq z\} \approx G^k\left(\frac{z - a_n}{b_n}\right),$$

ovvero G soddisfà al *postulato di stabilità* dato nella *Definizione (3.3)*. Dunque G e G^k sono identiche, a meno di una trasformazione affine dei rispettivi argomenti; ne segue che G è max-stabile e quindi, in base al *Teorema (3.4)*, è un membro della famiglia GEV.

NOTA 3.4.

La dimostrazione euristica data sopra mette in evidenza un punto assai interessante, già accennato all'inizio di questa Sezione: nelle applicazioni capita spesso che i dati raccolti siano a scala, diciamo, giornaliera (si pensi, ad esempio, alle temperature massime (e/o minime) misurate nelle 24 ore, alla quantità di pioggia caduta nelle 24 ore, ecc...), e che l'interesse primario risieda nel calcolo dei massimi annuali. Da un punto di vista pratico l'intero campione è pensato come se fosse suddiviso in k "blocchi" consecutivi indipendenti (dove k è il numero di anni di campionamento disponibili), ciascuno contenente $n = 365$ osservazioni indipendenti; quindi, per ogni anno si calcola l'osservazione massima, che fornisce un elemento del

sotto-campione (di ampiezza k) dei massimi annuali; infine, su quest'ultimo si adatta una delle distribuzioni dei massimi trattate in precedenza.
Quello appena descritto costituisce il procedimento standard di raccolta e di analisi dei massimi (annuali) di un certo fenomeno. Sebbene tale pratica possa, a prima vista, apparire ragionevole, è bene rendersi conto che esistono delle controindicazioni anche abbastanza serie. Infatti, se lo studio dei massimi annuali è svolto al fine di indagare il comportamento estremo del fenomeno di interesse, si corre il rischio scartare dati rilevanti: ad esempio, se solamente uno degli anni considerati evidenzia una fenomenologia estremale (si pensi a valori ripetuti di temperature elevate, o a periodi prolungati di piogge intense), mentre nel resto del campione il fenomeno non mostra eccessive fluttuazioni, è evidente che tener conto solo del singolo massimo raccolto nell'anno di maggiore "instabilità" conduce ad una perdita effettiva di informazione circa la dinamica più estremale del fenomeno stesso. Vedremo più avanti, nella *Sezione 3.2*, come sia possibile "isolare" proprio le componenti di maggiore interesse per lo studio dei valori estremi.

Il Teorema che segue fornisce indicazioni utili per l'identificazione della distribuzione limite.

TEOREMA 3.5 (Dominio di attrazione per i massimi di una distribuzione)**.** *La distribuzione F giace nel dominio di attrazione per i massimi della famiglia*

I. ***Tipo I (Gumbel)*** *se, e solo se,*

$$\lim_{n\to\infty} n\left[1 - F\left(x_{1-1/n} + x(x_{1-1/(ne)} - x_{1-1/n})\right)\right] = e^{-x}; \tag{3.16}$$

II. ***Tipo II (Fréchet)*** *se, e solo se, $\omega_F = +\infty$ e*

$$\lim_{t\to\infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-\gamma}, \quad x, \gamma > 0; \tag{3.17}$$

III. ***Tipo III (Weibull)*** *se, e solo se, $\omega_F < +\infty$ e*

$$\lim_{t\to\infty} \frac{1 - F\left(\omega_F - 1/(tx)\right)}{1 - F\left(\omega_F - 1/t\right)} = x^{-\gamma}, \quad x, \gamma > 0. \tag{3.18}$$

Qui x_q è il quantile di ordine q di F (si veda la Sezione 0.3).

Il Teorema che segue fornisce un criterio generale per il calcolo delle costanti di normalizzazione a_n, b_n.

TEOREMA 3.6. *Le costanti di normalizzazione a_n, b_n dell'equazione (3.4) possono essere scelte come segue.*

I. ***Tipo I (Gumbel):***

$$a_n = F^{-1}\left(1 - \frac{1}{n}\right), \quad b_n = F^{-1}\left(1 - \frac{1}{ne}\right) - a_n. \tag{3.19}$$

II. ***Tipo II (Fréchet):***

$$a_n = 0, \quad b_n = F^{-1}\left(1 - \frac{1}{n}\right). \tag{3.20}$$

III. ***Tipo III (Weibull):***

$$a_n = \omega_F, \quad b_n = \omega_F - F^{-1}\left(1 - \frac{1}{n}\right). \tag{3.21}$$

Evidentemente F^{-1} rappresenta la funzione dei quantili della distribuzione F (si veda la *Sezione 0.3*). La Proposizione che segue mostra che la scelta delle successioni $\{a_n\}, \{b_n\}$ può non essere unica.

PROPOSIZIONE 3.1. *Se $\{a_n\}, \{b_n\}$ sono successioni che soddisfanno all'equazione (3.4) e $\{a'_n\}, \{b'_n\}$ sono successioni tali che*

$$\lim_{n\to\infty} \frac{a_n - a'_n}{b_n} = 0 \quad e \quad \lim_{n\to\infty} \frac{b'_n}{b_n} = 1, \tag{3.22}$$

allora anche $\{a'_n\}, \{b'_n\}$ sono successioni normalizzanti ammissibili per l'equazione (3.4).

L'esempio che segue illustra l'uso dei risultati appena esposti.

ESEMPIO 3.1. ►

Nel *Teorema (3.6)* si supponga che la distribuzione "parente" F sia quella di Gumbel, Fréchet e Weibull nei casi, rispettivamente *(I)*, *(II)* e *(III)*. Grazie al *Teorema (3.5)* è facile calcolare quali siano le corrispondenti distribuzioni limite per i massimi. Quindi, usando le costanti a_n, b_n prescritte dal *Teorema (3.6)*, e considerando la forma canonica delle varie distribuzioni limite G, si ottiene quanto segue (si veda anche la *Tabella (3.1)*).

I. **Tipo I (Gumbel):** se $z \in \mathbb{R}$ e $t \in (0,1)$ si ha

$$F(z) = \exp\{-\exp(-z)\} \quad \text{e} \quad F^{-1}(t) = -\ln(-\ln(t)).$$

I quantili dell'*equazione (3.16)* valgono, rispettivamente, $z_{1-1/n} = -\ln(-\ln(1-1/n))$ e $z_{1-1/(ne)} = -\ln(-\ln(1-1/(ne)))$. Per $n \gg 1$ si ha $z_{1-1/n} \approx \ln(n)$ e $z_{1-1/(ne)} \approx \ln(n)+1$; ne segue che il limite dell'*equazione (3.16)* è e^{-z}. Dunque il dominio di attrazione è quello della legge Gumbel. Le costanti a_n, b_n prescritte dal *Teorema (3.6)* valgono, rispettivamente,

$$a_n = -\ln\left(-\ln\left(1-\frac{1}{n}\right)\right) \quad \text{e} \quad b_n = -\ln\left(-\ln\left(1-\frac{1}{ne}\right)\right) + \ln\left(-\ln\left(1-\frac{1}{n}\right)\right).$$

Effettuate le dovute sostituzioni e svolti alcuni tediosi passaggi si ottiene

$$\lim_{n\to\infty} F^n(a_n + b_n z) = \exp\{-\exp(-z)\}.$$

II. **Tipo II (Fréchet):** se $z > 0$ e $t \in (0,1)$ si ha

$$F(z) = \exp\left\{-z^{-\gamma}\right\} \quad \text{e} \quad F^{-1}(t) = (-\ln(t))^{-1/\gamma}.$$

Evidentemente $\omega_F = +\infty$ e il limite dell'*equazione (3.17)* è $x^{-\gamma}$; dunque il dominio di attrazione è quello della legge Fréchet. Le costanti a_n, b_n prescritte dal *Teorema (3.6)* valgono, rispettivamente,

$$a_n = 0 \quad \text{e} \quad b_n = \left(-\ln\left(1-\frac{1}{n}\right)\right)^{-1/\gamma}.$$

Effettuate le dovute sostituzioni e svolti alcuni tediosi passaggi si ottiene

$$\lim_{n\to\infty} F^n(a_n + b_n z) = \exp\left\{-z^{-\gamma}\right\}.$$

III. **Tipo III (Weibull):** se $z < 0$ e $t \in (0,1)$ si ha

$$F(z) = \exp\{-[(-z)^{\gamma}]\} \quad \text{e} \quad F^{-1}(t) = -(-\ln(t))^{1/\gamma}.$$

Evidentemente $\omega_F = 0 < +\infty$ e il limite dell'*equazione (3.18)* è $x^{-\gamma}$; dunque il dominio di attrazione è quello della legge Weibull. Le costanti a_n, b_n prescritte dal *Teorema (3.6)* valgono, rispettivamente,

$$a_n = 0 \quad \text{e} \quad b_n = 0 + \left(-\ln\left(1-\frac{1}{n}\right)\right)^{1/\gamma}.$$

Effettuate le dovute sostituzioni e svolti alcuni tediosi passaggi si ottiene

$$\lim_{n\to\infty} F^n(a_n + b_n z) = \exp\{-[(-z)^{\gamma}]\}.$$

I risultati appena ottenuti non sorprendono: partendo da un distribuzione "parente" GEV (e quindi *max-stabile*) è legittimo aspettarsi di cadere nel dominio di attrazione della stessa legge di partenza. ◄

OSSERVAZIONE 3.3.

Alla luce del *Teorema (3.4)* e della libertà di scelta consentita dalla *Proposizione (3.1)*, è possibile calcolare le successioni $\{a_n\}, \{b_n\}$ semplicemente eguagliando le espressioni delle funzioni $G(z)$ e $G^n(a_n + b_n z)$, dopo aver scritto G in forma canonica. Così facendo si ottiene quanto segue.

I. **Tipo I (Gumbel):** $a_n = \ln n$, $b_n = 1$; infatti

$$\exp\{-n\exp[-(a_n + b_n z)]\} = \exp\{-\exp(-z)\}$$

da cui

$$ne^{-a_n}e^{-b_n z} = e^{-z}, \quad \text{ovvero} \quad \begin{cases} ne^{-a_n} = 1 \\ e^{-b_n z} = e^{-z} \end{cases},$$

onde la soluzione proposta.

II. **Tipo II (Fréchet):** $a_n = 0$, $b_n = n^{1/\gamma}$; infatti

$$\exp\left\{-n\left(a_n + b_n z\right)^{-\gamma}\right\} = \exp\left\{-z^{-\gamma}\right\}$$

da cui

$$n(a_n + b_n z)^{-\gamma} = z^{-\gamma}, \quad \text{ovvero} \quad \begin{cases} n^{-1/\gamma}a_n = 0 \\ n^{-1/\gamma}b_n z = z \end{cases},$$

onde la soluzione proposta.

III. **Tipo III (Weibull):** $a_n = 0$, $b_n = n^{-1/\gamma}$; infatti

$$\exp\{-n\left(-a_n - b_n z\right)^{\gamma}\} = \exp\{-(-z)^{\gamma}\}$$

da cui

$$n(-a_n - b_n z)^{\gamma} = (-z)^{\gamma}, \quad \text{ovvero} \quad \begin{cases} n^{1/\gamma}a_n = 0 \\ n^{1/\gamma}b_n z = z \end{cases},$$

onde la soluzione proposta.

Come già accennato, le successioni proposte non sono le uniche possibili, e al loro variare si modifica la velocità di convergenza della funzione $G^n(a_n + b_n z)$; tuttavia non cambia il parametro di forma γ. Si noti come, nel limite $n \to \infty$, le costanti trovate ora siano le stesse di quelle dell'*Esempio (3.1)* e soddisfanno all'*equazione (3.22)*.

Il Teorema che segue fornisce un criterio utile per l'individuazione del dominio di attrazione per i massimi.

TEOREMA 3.7. *Una condizione necessaria e sufficiente affinché la distribuzione continua F appartenga al dominio di attrazione per i massimi di G è che*

$$\lim_{\epsilon \to 0} \frac{F^{-1}(1-\epsilon) - F^{-1}(1-2\epsilon)}{F^{-1}(1-2\epsilon) - F^{-1}(1-4\epsilon)} = 2^c. \tag{3.23}$$

In particolare, G appartiene alla famiglia

I. ***Tipo I (Gumbel)** se $c = 0$;*

II. ***Tipo II (Fréchet)** se $c > 0$;*

III. ***Tipo III (Weibull)** se $c < 0$.*

NOTA 3.5.

Si osservi come nel *Teorema (3.7)* il limite mostrato nell'*equazione (3.23)* corrisponda ad uno studio della distribuzione F in un intorno arbitrariamente piccolo di ω_F.

Volendo riassumere alcuni dei risultati enunciati è possibile affermare che, nelle ipotesi del *Teorema (3.1)*:

- solo tre famiglie di distribuzioni (ovvero *Gumbel*, *Fréchet* e *Weibull*) possono intervenire quali distribuzioni limite per i massimi (minimi) di successioni di osservazioni indipendenti e identicamente distribuite;
- esistono regole per determinare se una data distribuzione F cade nel dominio di attrazione di una opportuna distribuzione limite;

- esistono regole per determinare le costanti di normalizzazione a_n, b_n;
- se $\omega_F = +\infty$ allora F non può giacere nel dominio di attrazione della famiglia Weibull;
- se $\omega_F < +\infty$ allora F non può giacere nel dominio di attrazione della famiglia Fréchet.

NOTA 3.6.

Il fatto che una distribuzione "parente" F abbia la coda (destra e/o sinistra) limitata non è sufficiente a garantire che essa giaccia nel dominio di attrazione della famiglia Weibull; un semplice controesempio è fornito dalla legge LogNormale, come mostrato in *Tabella (3.1)*.

ESEMPIO 3.2. ►

Si consideri la distribuzione Esponenziale (standard) avente f.r.

$$F(x) = 1 - e^{-x}$$

per $x > 0$ e nulla altrove; la funzione dei quantili corrispondente vale $F^{-1}(t) = -\ln(1-t)$, $t \in (0,1)$. Grazie al *Teorema (3.7)* si ha

$$\lim_{\epsilon \to 0} \frac{-\ln(1-(1-\epsilon)) + \ln(1-(1-2\epsilon))}{-\ln(1-(1-2\epsilon)) + \ln(1-(1-4\epsilon))} = \lim_{\epsilon \to 0} \frac{-\ln(\epsilon) + \ln 2 + \ln(\epsilon)}{-\ln 2 - \ln(\epsilon) + \ln 4 + \ln(\epsilon)} = 1 = 2^0.$$

Dunque $c = 0$ e il dominio di attrazione per i massimi è quello della famiglia Gumbel; d'altro canto, il limite dell'*equazione (3.16)* risulta essere proprio e^{-x}. La *Figura (3.4)* mostra l'analisi "a blocchi" dei massimi di un campione di ampiezza $n \times k = 3000$ dato da $k = 30$ simulazioni indipendenti (separate dalle linee verticali tratteggiate) estratte della distribuzione Esponenziale (standard), ciascuna di ampiezza $n = 100$.

I parametri a_n, b_n sono dati dall'*equazione (3.19)*, con $n = 100$, e valgono, rispettivamente, $a_n = \ln n$ e $b_n = 1$. Si noti come l'asse delle ordinate della funzione di sopravvivenza sia logaritmico e come la coda destra mostri un andamento (asintotico) lineare; ciò implica che il comportamento limite della distribuzione sia di tipo "esponenziale", come anticipato nell'*Osservazione (3.2)*: infatti dalla relazione $\ln y \approx \alpha + \beta z$, con $z \gg 1$, si deduce che $y \propto e^{\beta z}$. ◄

ESEMPIO 3.3. ►

Si consideri la distribuzione di Cauchy (standard) avente f.r.

$$F(x) = 1/2 + \arctan(x)/\pi$$

per $x \in \mathbb{R}$; la funzione dei quantili corrispondente vale $F^{-1}(t) = \tan\left((t-1/2)\,\pi\right)$, $t \in (0,1)$. Grazie al *Teorema (3.7)* si ha

$$\lim_{\epsilon \to 0} \frac{\tan\left((1-\epsilon-1/2)\,\pi\right) - \tan\left((1-2\epsilon-1/2)\,\pi\right)}{\tan\left((1-2\epsilon-1/2)\,\pi\right) - \tan\left((1-4\epsilon-1/2)\,\pi\right)} = \lim_{\epsilon \to 0} \frac{\tan\left((1/2-\epsilon)\,\pi\right) - \tan\left((1/2-2\epsilon)\,\pi\right)}{\tan\left((1/2-2\epsilon)\,\pi\right) - \tan\left((1/2-4\epsilon)\,\pi\right)} = 2^1.$$

Dunque $c = 1 > 0$ e il dominio di attrazione per i massimi è quello della famiglia Fréchet. Il parametro γ dell'*equazione (3.17)* è facile da calcolare: infatti, essendo $\arctan(y) \approx (\pi/2 - 1/y)$ per $y \gg 1$,

$$\lim_{t \to \infty} \frac{1/2 - \arctan(tx)/\pi}{1/2 - \arctan(t)/\pi} = \lim_{t \to \infty} \frac{\pi - 2\arctan(tx)}{\pi - 2\arctan(t)} = x^{-1},$$

e dunque $\gamma = 1$. La *Figura (3.5)* mostra l'analisi "a blocchi" dei massimi di un campione di ampiezza $n \times k = 3000$ dato da $k = 30$ simulazioni indipendenti (separate dalle linee verticali tratteggiate) estratte della distribuzione di Cauchy (standard), ciascuna di ampiezza $n = 100$.

I parametri a_n, b_n sono dati dall'*equazione (3.20)*, con $n = 100$, e valgono, rispettivamente, $a_n = 0$ e $b_n = \tan\left(\pi\,(1/2 - 1/n)\right)$. Si noti come gli assi della funzione di sopravvivenza siano logaritmici e come la coda destra mostri un andamento (asintotico) lineare; ciò implica che il comportamento limite della distribuzione sia di tipo "algebrico", come anticipato nell'*Osservazione (3.2)*: infatti dalla relazione $\ln y \approx \alpha + \beta \ln z$, con $z \gg 1$, si deduce che $y \propto z^{\beta}$.

A differenza dei valori simulati dell'*Esempio (3.2)*, che visivamente non presentavano fluttuazioni "troppo" marcate, qui siamo invece in presenza di autentici "outlier", come appare evidente dalla *Figura (3.5)-(a)*; tuttavia,

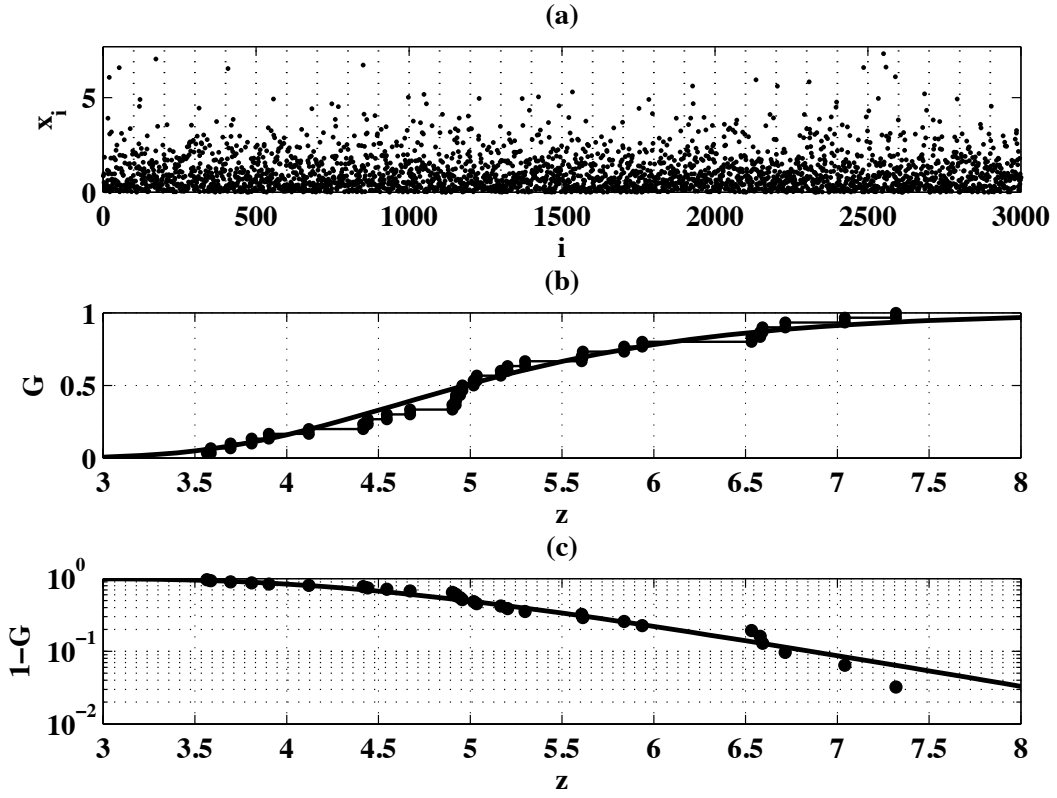

Figura 3.4: analisi dei massimi "a blocchi" del campione dell'*Esempio (3.2).* (*a*) I 3000 valori simulati estratti della distribuzione Esponenziale (standard). (*b*) Confronto tra la funzione di ripartizione empirica dei 30 massimi simulati (*cerchi*) e la legge Gumbel dell'*equazione (3.6)* con $a = a_n$ e $b = b_n$ (*linea*). (*c*) Confronto tra la funzione di sopravvivenza empirica dei 30 massimi simulati (*cerchi*) — usando le *plotting positions* dell'*Esempio (2.7)* — e quella della legge Gumbel (*linea*).

ciò non deve soprendere: infatti la distribuzione di Cauchy non ha momenti (interi) di alcun ordine, proprio a causa della presenza di valori estremi (sia positivi sia negativi) eccezionali. Come già accennato nell'*Osservazione (3.2)* (si veda, in particolare, l'*equazione (3.10)*), tale comportamento è tipico delle distribuzioni appartenenti alla famiglia Fréchet, per le quali non esistono i momenti di ordine superiore ad una data funzione del parametro di forma. Nella pratica è proprio per questo motivo che le leggi Fréchet risultano adatte a descrivere fenomenologie estremali e catastrofiche, tipiche di numerosi sistemi quali, ad esempio, quelli geofisici e finanziari. ◄

ESEMPIO 3.4. ►

Si consideri la distribuzione Uniforme avente f.r.

$$F(x) = x\,\mathbb{I}_{(0,1]}(x) + \mathbb{I}_{(1,\infty)}(x);$$

la funzione dei quantili corrispondente vale $F^{-1}(t) = t$, $t \in (0,1)$. Grazie al *Teorema (3.7)* si ha

$$\lim_{\epsilon \to 0} \frac{1 - \epsilon - 1 + 2\epsilon}{1 - 2\epsilon - 1 + 4\epsilon} = 2^{-1}.$$

Dunque $c = -1 < 0$ e il dominio di attrazione per i massimi è quello della famiglia Weibull. Il parametro γ dell'*equazione (3.18)* è facile da calcolare: infatti, essendo $\omega_F = 1$,

$$\lim_{t \to \infty} \frac{1 - (1 - 1/tx)}{1 - (1 - 1/t)} = \lim_{t \to \infty} \frac{1/tx}{1/t} = x^{-1},$$

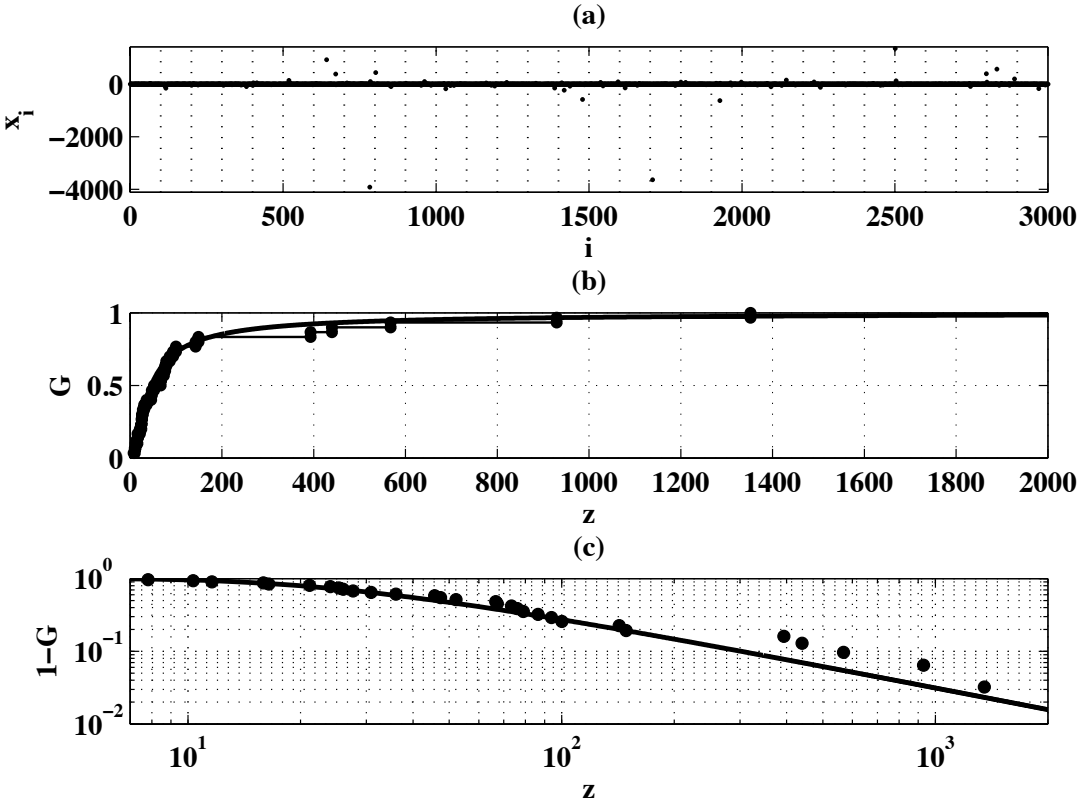

Figura 3.5: analisi dei massimi "a blocchi" del campione dell'*Esempio (3.3).* (*a*) I 3000 valori simulati estratti della distribuzione di Cauchy (standard). (*b*) Confronto tra la funzione di ripartizione empirica dei 30 massimi simulati (*cerchi*) e la legge Fréchet dell'*equazione (3.7)* con $a = a_n$ e $b = b_n$ (*linea*). (*c*) Confronto tra la funzione di sopravvivenza empirica dei 30 massimi simulati (*cerchi*) — usando le *plotting positions* dell'*Esempio (2.7)* — e quella della legge Fréchet (*linea*).

e dunque $\gamma = 1$. La *Figura (3.6)* mostra l'analisi "a blocchi" dei massimi di un campione di ampiezza $n \times k = 3000$ dato da $k = 30$ simulazioni indipendenti (separate dalle linee verticali tratteggiate) estratte della distribuzione Uniforme, ciascuna di ampiezza $n = 100$.

I parametri a_n, b_n sono dati dall'*equazione (3.21)*, con $n = 100$, e valgono, rispettivamente, $a_n = \omega_F = 1$ e $b_n = \omega_F - (1 - 1/n) = 1/n$. Si noti come $\omega_G = 1 < +\infty$, come anticipato nell'*Osservazione (3.2).* ———◀

Il comportamento asintotico delle distribuzioni dei massimi è di grande importanza nelle applicazioni, in quanto esso determina le probabilità degli eventi più estremi e catastrofici (e lo stesso vale per i minimi). Al fine di studiare il comportamento limite di una distribuzione si introduce la seguente nozione.

DEFINIZIONE 3.4 (Equivalenza delle code)**.** Due distribuzioni F e H sono dette *coda-destra equivalenti (right-tail equivalent)* se, e solo se,

$$\omega_F = \omega_H \quad \text{e} \quad \lim_{x \to \omega_F^-} \frac{1 - F(x)}{1 - H(x)} = 1. \tag{3.24a}$$

Analogamente, F e H sono dette *coda-sinistra equivalenti (left-tail equivalent)* se, e solo se,

$$\alpha_F = \alpha_H \quad \text{e} \quad \lim_{x \to \alpha_F^+} \frac{F(x)}{H(x)} = 1. \tag{3.24b}$$

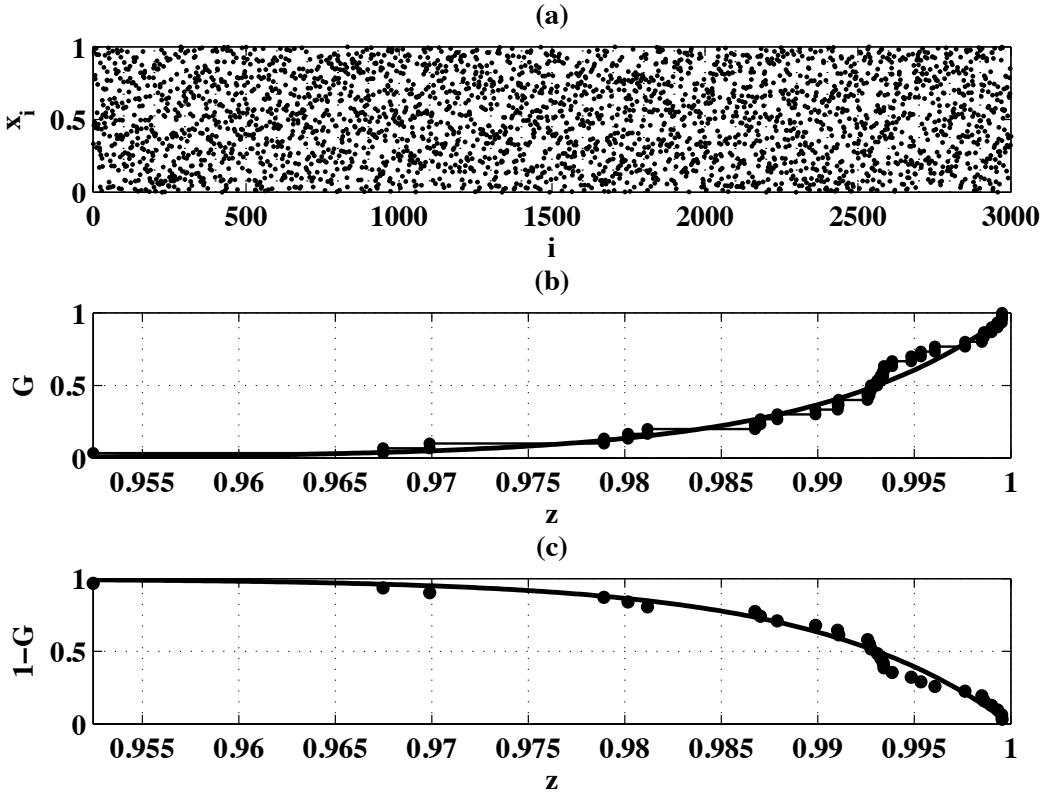

Figura 3.6: analisi dei massimi "a blocchi" del campione dell'*Esempio (3.4).* (*a*) I 3000 valori simulati estratti della distribuzione Uniforme. (*b*) Confronto tra la funzione di ripartizione empirica dei 30 massimi simulati (*cerchi*) e la legge Weibull dell'*equazione (3.8)* con $a = a_n$ e $b = b_n$ (*linea*). (*c*) Confronto tra la funzione di sopravvivenza empirica dei 30 massimi simulati (*cerchi*) — usando le *plotting positions* dell'*Esempio (2.7)* — e quella della legge Weibull (*linea*).

Il risultato che segue chiarisce la connessione tra l'equivalenza delle code destre e il dominio di attrazione per i massimi.

PROPOSIZIONE 3.2. *Se F e H soddisfanno alle condizioni del Teorema (3.1) e sono coda-destra equivalenti, e inoltre*

$$\lim_{n\to\infty} F^n(a_n + b_n\, x) = G(x) \quad \textit{per ogni } x \in \mathbb{R}, \tag{3.25}$$

allora

$$\lim_{n\to\infty} H^n(a_n + b_n\, x) = G(x) \quad \textit{per ogni } x \in \mathbb{R}. \tag{3.26}$$

OSSERVAZIONE 3.4.

La *Proposizione (3.2)* afferma che: (*i*) se due distribuzioni sono coda-destra equivalenti e una delle due appartiene a qualche dominio di attrazione per i massimi, allora anche l'altra giace nello stesso dominio di attrazione; (*ii*) le successioni delle costanti di normalizzazione sono le stesse per le due distribuzioni. Le implicazioni pratiche della *Proposizione (3.2)* sono notevoli: infatti, se la distribuzione F può essere (asintoticamente) "rimpiazzata" da un'altra distribuzione ad essa coda-equivalente, allora per lo studio del comportamento statistico dei valori estremi può essere sufficiente l'utilizzo della sola famiglia GEV. Dunque, per $n \gg 1$, $F^n(a_n + b_n x)$ potrebbe essere sostituita da $G(x)$; analogamente, per $x \gg 1$, $F(x)$ potrebbe essere sostituita da $G^{1/n}((x - a_n)/b_n)$.

Chiudiamo questa Sezione con un commento riguardante alcuni aspetti pratici dello studio dei massimi "a blocchi". Una volta raccolto il campione (diciamo di ampiezza totale $n \times k$) è necessario calcolare i parametri della legge GEV che "meglio" si adatta (secondo criteri opportuni che studieremo nel *Capitolo 4*) alle osservazioni disponibili dei massimi. La *vexata quaestio* è rappresentato dalla scelta (*i*) della dimensione n di ciascun blocco e (*ii*) dell'ampiezza k del campione dei massimi. In alcune situazioni (ad esempio nello studio dei massimi annuali di pioggia) sia n sia k sono già dati a priori; in altri contesti, invece, possono essere arbitrariamente scelti dallo statistico incaricato delle analisi. In quest'ultimo caso si tratta di trovare un bilanciamento adeguato tra due esigenze conflittuali: da un lato n deve essere quanto più grande possibile, così da soddisfare alle condizioni di "asintoticità" prescritte dai diversi Teoremi enunciati (ovvero $n \gg 1$); dall'altro anche k deve essere sufficientemente grande così da fornire una base statistica sufficiente (come vedremo meglio nel *Capitolo 4*). Purtroppo non sempre è possibile attuare questo compromesso in maniera efficace.

3.2 Modelli "a soglia"

Come già accennato nella *Nota (3.4)*, lo studio "a blocchi" dei massimi può condurre alla perdita di informazioni importanti circa la dinamica più estremale del fenomeno in esame. Un modo di procedere alternativo è quello cosiddetto "POT", acronimo di "Peaks-Over-Threshold" (ovvero Picchi al di Sopra di una Soglia — in Letteratura è talora usato anche l'acronimo "PDS", ovvero "Partial-Duration-Series" [17]). Tale vocabolo indica la procedura per l'analisi dei massimi (minimi) che descriveremo brevemente nel seguito.

Si consideri una successione $X_1, X_2, \ldots$ di v.a. i.i.d. aventi f.r. comune F. Al fine di studiare il comportamento estremo del fenomeno descritto da tali variabili è naturale considerare quali eventi "critici" quelli del tipo $E = \{X > u\}$, dove u rappresenta una soglia prefissata arbitrariamente alta (o bassa, se l'evento di interesse è del tipo $E = \{X < u\}$ come nel caso dei minimi). In altri termini, la filosofia di fondo è quella di analizzare i cosiddetti *eccessi* al di sopra una soglia prestabilita. Infatti, un modo naturale di descrivere efficacemente la dinamica estremale del sistema in esame è dato dalla probabilità condizionale

$$1 - H_u(x) = \mathbf{P}\{X > u + x \mid X > u\} = \frac{1 - F(u+x)}{1 - F(u)}, \quad x > 0. \tag{3.27}$$

Se la distribuzione "parente" F fosse nota, tale sarebbe anche la probabilità dell'*equazione (3.27)*; tuttavia, poiché nelle applicazioni questa possibilità è quasi sempre negata, è necessario cercare di approssimare la distribuzione condizionale indicata tramite un'espressione che sia valida in generale, indipendentemente da F (si noti come questo approccio sia simile a quello insito nell'uso dell'approssimazione GEV). Il Teorema che segue è fondamentale nello studio dei massimi POT.

TEOREMA 3.8 (Distribuzione Generalized Pareto per i massimi)**.** *Si supponga che la distribuzione "parente" F soddisfi alle ipotesi del Teorema (3.2). Allora, per $u \gg 1$,*

$$H_u(x) = \mathbf{P}\{X \leq u + x \mid X > u\} \approx 1 - \left(1 + \gamma \frac{x}{b}\right)^{-1/\gamma} \tag{3.28}$$

sull'insieme $\{x\colon x > 0 \ \ e \ \ 1 + \gamma x/b > 0\}$, dove i parametri $b = b' + \gamma(u - a')$ e $\gamma = \gamma'$ sono funzioni di quelli del Teorema (3.2). Il caso limite $\gamma = 0$ corrisponde (si riduce) alla distribuzione Esponenziale

$$H_u(x) \approx 1 - e^{-x/b}, \tag{3.29}$$

che non dipende esplicitamente da u.

DEFINIZIONE 3.5 (Famiglia Generalized Pareto)**.** La famiglia di leggi definita dall'*equazione (3.28)* (e dall'*equazione (3.29)* nel caso particolare $\gamma = 0$) è detta *Generalized Pareto* (*GP*).

OSSERVAZIONE 3.5.

In buona sostanza il *Teorema (3.8)* afferma che se i massimi "a blocchi" obbediscono (approssimativamente) ad una legge GEV G allora gli eccessi al di sopra di una certa soglia u sufficientemente alta seguono (approssimativamente) una distribuzione GP H. Inoltre, i parametri di quest'ultima si possono ricavare direttamente da quelli della corrispondente legge GEV (in effetti la distribuzione GP H si può ottenere dalla legge GEV G tramite la formula $H = 1 + \ln G$); in particolare, il parametro di forma γ è lo stesso. Quest'ultima osservazione, combinata con lo studio della formula data dall'*equazione (3.28)*, permette di affermare che anche nel modello POT possono presentarsi casi in cui i momenti di ordine m elevato non esistono (ovvero quando $m \geq 1/\gamma > 0$).

A scopo illustrativo rianalizziamo ora i risultati dei tre *Esempi (3.2)–(3.4)*, dandone una lettura ed una interpretazione alternativa attraverso l'approccio POT.

ESEMPIO 3.5. ►

Si consideri il campione estratto dalla distribuzione Esponenziale (standard) dell'*Esempio (3.2)*. Per quanto già visto, la legge GEV dei massimi "a blocchi" è quella legata alla famiglia Gumbel; ne segue che la distribuzione GP per lo studio POT dei massimi è quella data dall'*equazione (3.29)*. In effetti il calcolo diretto fornisce

$$1 - H_u(x) = \frac{1 - F(u+x)}{1 - F(u)} = \frac{e^{-(u+x)}}{e^{-u}} = e^{-x}$$

per ogni $x > 0$, che corrisponde alla legge GP con $\gamma = 0$ e $b = 1$. In particolare, si osservi come questo sia un risultato *esatto*, valido per tutte le soglie $u > 0$.

La *Figura (3.7)* mostra l'analisi POT degli stessi dati dell'*Esempio (3.2)*; la soglia $u > 0$ è stata scelta così da ottenere 30 eccessi, che sono poi mostrati utilizzando la variabile $X - u$.

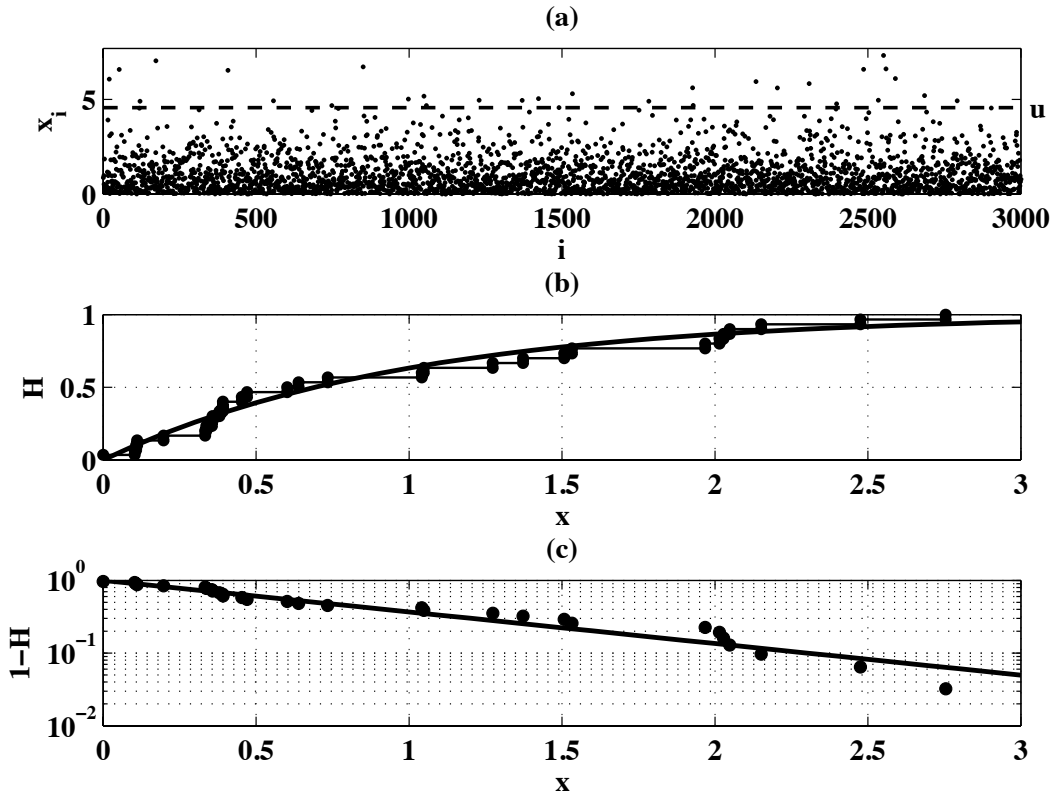

Figura 3.7: analisi dei massimi POT del campione dell'*Esempio (3.2)*. (*a*) I 3000 valori simulati estratti della distribuzione Esponenziale (standard); la *linea tratteggiata* indica la soglia $u \approx 4.57$. (*b*) Confronto tra la funzione di ripartizione empirica dei 30 massimi POT (*cerchi*) e la legge GP dell'*equazione (3.29)* (*linea*). (*c*) Confronto tra la funzione di sopravvivenza empirica dei 30 massimi POT (*cerchi*) — usando le *plotting positions* dell'*Esempio (2.7)* — e quella della legge mostrata in (*b*).

Qui $n = 100$ e i parametri GP valgono, rispettivamente, $\gamma = 0$ e $b = b_n + \gamma(u - a_n) = 1$. Si noti come l'asse delle ordinate della funzione di sopravvivenza sia logaritmico e come la coda destra mostri un andamento (asintotico) lineare; ciò implica che il comportamento limite della distribuzione sia di tipo "esponenziale".

L'analisi della *Figura (3.7)-(a)* mette in evidenza come, ad esempio, nel settore $2501 \leq i \leq 2600$ siano presenti almeno quattro "eccessi": tutti questi valori concorrono a fornire informazioni circa la dinamica estremale del fenomeno in esame, e sono riportati nei grafici (*b*) e (*c*). Al contrario, solo il massimo tra questi può essere utilizzato nel caso dello studio "a blocchi" dell'*Esempio (3.2).* ◀

ESEMPIO 3.6. ▶

Si consideri il campione estratto dalla distribuzione Cauchy (standard) dell'*Esempio (3.3).* Per quanto già visto, la legge GEV dei massimi "a blocchi" è quella legata alla famiglia Fréchet; ne segue che la distribuzione GP per lo studio POT dei massimi è quella data dall'*equazione (3.28)* con $\gamma > 0$. In effetti il calcolo diretto fornisce (essendo $\arctan(t) \approx \left(\frac{\pi}{2} - \frac{1}{t}\right)$ per $t \gg 1$)

$$1 - H_u(x) = \frac{1 - F(u+x)}{1 - F(u)} = \frac{\frac{1}{2} - \frac{\arctan(u+x)}{\pi}}{\frac{1}{2} - \frac{\arctan(u)}{\pi}} \approx \frac{\frac{1}{u+x}}{\frac{1}{u}} = \left(1 + \frac{x}{u}\right)^{-1}$$

per ogni $x > 0$, che corrisponde alla legge GP con $\gamma = 1$ e $b = u$.

La *Figura (3.8)* mostra l'analisi POT degli stessi dati dell'*Esempio (3.3)*; la soglia $u > 0$ è stata scelta così da ottenere 30 eccessi, che sono poi mostrati utilizzando la variabile $X - u$.

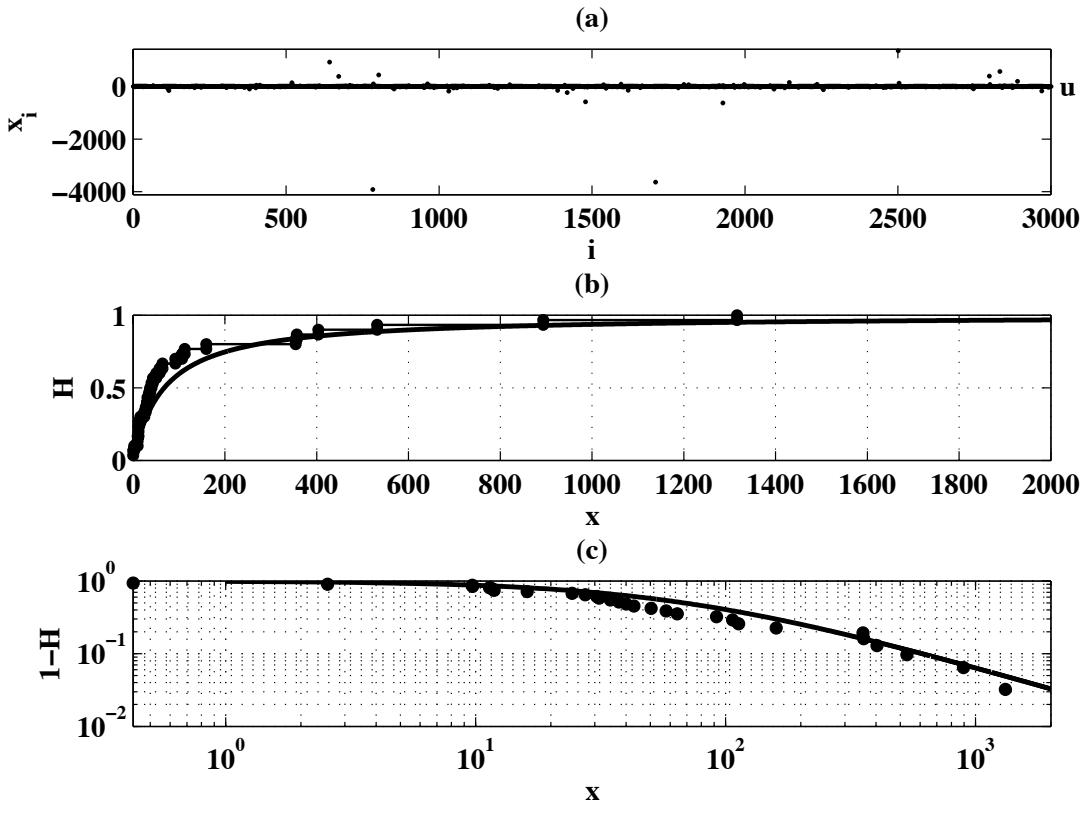

Figura 3.8: analisi dei massimi POT del campione dell'*Esempio (3.3).* (*a*) I 3000 valori simulati estratti della distribuzione di Cauchy (standard); la *linea tratteggiata* indica la soglia $u \approx 36.04$. (*b*) Confronto tra la funzione di ripartizione empirica dei 30 massimi POT (*cerchi*) e la legge GP dell'*equazione (3.28)* (*linea*). (*c*) Confronto tra la funzione di sopravvivenza empirica dei 30 massimi POT (*cerchi*) — usando le *plotting positions* dell'*Esempio (2.7)* — e quella della legge mostrata in (*b*).

Qui $n = 100$ e i parametri GP valgono, rispettivamente, $\gamma = 1$ e $b = b_n + \gamma(u - a_n) \approx 57.3914$. Si noti come gli assi della funzione di sopravvivenza siano logaritmici e come la coda destra mostri un andamento (asintotico)

lineare; ciò implica che il comportamento limite della distribuzione sia di tipo "algebrico". Valgono qui le stesse considerazioni espresse nel caso GEV per quanto concerne il comportamento estremale delle osservazioni.

Anche la *Figura (3.8)-(a)* mostra come, ad esempio, nei settori $601 \leq i \leq 700$ e $2801 \leq i \leq 2900$ siano presenti più di un "eccesso". Nello studio "a soglia" tutti questi valori entrano a far parte dell'analisi della dinamica estremale del fenomeno in esame, mentre solo il massimo tra questi può essere utilizzato nel caso dello studio "a blocchi" dell'*Esempio (3.3).* ◀

ESEMPIO 3.7. ▶

Si consideri il campione estratto dalla distribuzione Uniforme dell'*Esempio (3.4).* Per quanto già visto, la legge GEV dei massimi "a blocchi" è quella legata alla famiglia Weibull; ne segue che la distribuzione GP per lo studio POT dei massimi è quella data dall'*equazione (3.28)* con $\gamma < 0$. In effetti il calcolo diretto fornisce

$$1 - H_u(x) = \frac{1 - F(u+x)}{1 - F(u)} = \frac{1 - (u+x)}{1 - u} = 1 - \frac{x}{1-u}$$

per $0 < x < 1 - u$, che corrisponde alla legge GP con $\gamma = -1$ e $b = 1 - u$.

La *Figura (3.9)* mostra l'analisi POT degli stessi dati dell'*Esempio (3.4)*; la soglia $u > 0$ è stata scelta così da ottenere 30 eccessi, che sono poi mostrati utilizzando la variabile $X - u$.

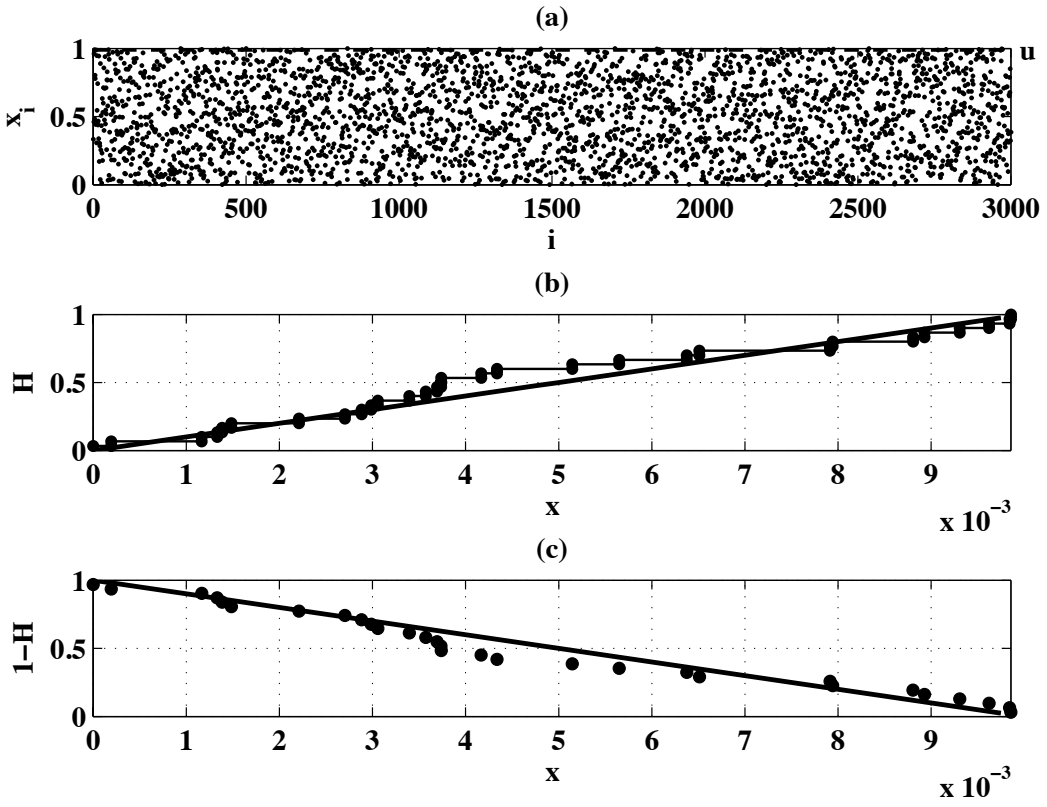

Figura 3.9: analisi dei massimi POT del campione dell'*Esempio (3.4).* (*a*) I 3000 valori simulati estratti della distribuzione Uniforme; la *linea tratteggiata* indica la soglia $u \approx 0.99$. (*b*) Confronto tra la funzione di ripartizione empirica dei 30 massimi POT (*cerchi*) e la legge GP dell'*equazione (3.28)* (*linea*). (*c*) Confronto tra la funzione di sopravvivenza empirica dei 30 massimi POT (*cerchi*) — usando le *plotting positions* dell'*Esempio (2.7)* — e quella della legge mostrata in (*b*).

Qui $n = 100$ e i parametri GP valgono, rispettivamente, $\gamma = -1$ e $b = b_n + \gamma(u - a_n) \approx 0.0203$. Si noti come $\omega_H = b < +\infty$. ◀

Una giustificazione informale del risultato del *Teorema (3.8)* è la seguente. Grazie al *Teorema (3.2)* sappiamo che, per $n \gg 1$,

$$F^n(z) \approx \exp\left\{ -\left[1 + \gamma' \frac{z - a'}{b'}\right]^{-1/\gamma'} \right\}$$

e dunque

$$n \ln F(z) \approx -\left[1 + \gamma' \frac{z - a'}{b'}\right]^{-1/\gamma'} .$$

Tuttavia, essendo $\ln F(z) = \ln(1 - (1 - F(z)))$, per $z \gg 1$ si ha $\ln F(z) \approx -(1 - F(z))$. Considerando la soglia $u \gg 1$ e sostituendo si ottiene

$$1 - F(u) \approx \frac{1}{n} \left[1 + \gamma' \frac{u - a'}{b'}\right]^{-1/\gamma'} ,$$
$$1 - F(u + x) \approx \frac{1}{n} \left[1 + \gamma' \frac{u + x - a'}{b'}\right]^{-1/\gamma'} .$$

Dunque, volendo calcolare $1 - H_u(x)$, si ottiene

$$\begin{aligned} \mathbf{P}\{X > u + x \mid X > u\} &\approx \frac{n^{-1} \left[1 + \gamma' \frac{u+x-a'}{b'}\right]^{-1/\gamma'}}{n^{-1} \left[1 + \gamma' \frac{u-a'}{b'}\right]^{-1/\gamma'}} \\ &= \left(1 + \gamma \frac{x}{b}\right)^{-1/\gamma} , \end{aligned}$$

dove $b = b' + \gamma(u - a')$ e $\gamma = \gamma'$, come richiesto.

Chiudiamo questa Sezione con un commento riguardante alcuni aspetti pratici dello studio dei massimi POT, così come si è fatto nel caso dell'approccio "a blocchi". Qui la *vexata quaestio* riguarda la scelta della soglia u, che deve essere fissata in maniera adeguata onde trovare un bilanciamento ottimale tra due esigenze conflittuali: da un lato u deve essere quanto più grande possibile, così da soddisfare alle condizioni di "asintoticità" prescritte dal *Teorema (3.8)*; dall'altro u non deve essere "troppo" grande, così da fornire un campione di eccessi sufficiente alle ulteriori indagini statistiche. Purtroppo non sempre è possibile attuare questo compromesso in maniera efficace.

Capitolo 4

STIMATORI

Uno dei problemi fondamentali della Statistica Matematica consiste nell'inferire "informazioni" sul parametro $\theta \in \Theta$ (o su una sua funzione misurabile $\xi(\theta)$) avendo osservato un risultato $\omega \in \Omega$, essendo cioè in presenza di un campione $\mathbf{X}_F^n$ definito su un modello statistico dominato $\mathfrak{M}$ soggiacente. Nel seguito si introducono alcuni concetti di importanza fondamentale per il Corso.

4.1 Stimatori

Al fine di raccogliere "informazioni" sul parametro di interesse si usano gli *stimatori*.

DEFINIZIONE 4.1 (Stimatore)**.** Si dice *stimatore* del parametro θ (o di una sua funzione misurabile $\xi(\theta)$) ogni statistica $\omega \mapsto T(\omega)$ che si pensi di usare al posto di θ (o di $\xi(\theta)$).

Seppur vaga, la *Definizione (4.1)* ha una portata notevole, come si apprezzerà meglio nel seguito. Se la funzione ξ è l'identità, la stima di θ è banalmente ricondotta alla stima di $\xi(\theta)$.

NOTA 4.1.

Si osservi, in particolare, che *uno stimatore T di $\xi(\theta)$ non può dipendere da θ stesso*.

ESEMPIO 4.1. ►

Nel caso dell'*Esempio (1.3)(1)* può essere di interesse la stima di θ, cioè della probabilità di ottenere testa in un lancio; nel caso dell'*Esempio (1.3)(2)* può essere di interesse la stima della vita media delle lampade, cioè della funzione $\xi(\theta) = 1/\theta$. Si noti che in entrambi i casi possiamo pensare di usare quale stimatore la *media campionaria*, cioè la statistica $T = \overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ (si veda la *Sezione 4.1.1*). ◄

Gli *stimatori lineari* giocano un ruolo importante nell'ambito della Statistica Matematica.

DEFINIZIONE 4.2 (Stimatore lineare)**.** Dato un campione $\mathbf{X}_F^n$, uno stimatore T si dice *lineare* se è della forma

$$T = \sum_{i=1}^{n} c_i\, X_i, \tag{4.1}$$

dove $c_i \in \mathbb{R}$, $i = 1, \ldots, n$, ovvero una combinazione lineare delle osservazioni.

Assegnata una funzione $\xi\colon \Theta \to \mathbb{R}^d$ spesso ne esistono più stimatori; si pone dunque il problema di scegliere tra gli stimatori possibili, introducendo criteri che consentano sia di imporre che gli stimatori soddisfacciano a condizioni opportune sia il confronto tra stimatori. Per ragioni che si comprenderanno tra breve, è usuale considerare la *classe degli stimatori di quadrato sommabile*, indicata solitamente con $\mathfrak{D}^2$. Inoltre, una richiesta assai frequente è che gli stimatori siano *corretti*.

DEFINIZIONE 4.3 (Correttezza)**.** In un modello statistico dominato $\mathfrak{M}$ uno stimatore T della funzione $\xi\colon \Theta \to \mathbb{R}$ si dice *corretto* (o *non distorto*) se per ogni $\theta \in \Theta$ si ha

$$\mathbf{E}_\theta(T) = \xi(\theta). \tag{4.2}$$

In maniera del tutto analoga si definisce la correttezza per uno stimatore $T\colon \Omega \to \mathbb{R}^d$ della funzione vettoriale $\xi\colon \Theta \to \mathbb{R}^d$ quando lo spazio dei parametri è $\Theta \subseteq \mathbb{R}^d$, $d > 1$: in questo caso l'*equazione (4.2)* si supporrà valida per ciascuna componente. La *classe degli stimatori corretti di quadrato sommabile* si indicherà con $\mathfrak{D}_c^2$.

OSSERVAZIONE 4.1.

Siano T_1, T_2 due *differenti*[1] stimatori non distorti di $\xi(\theta)$. Allora esiste un'infinità di stimatori corretti di $\xi(\theta)$ dati dalla loro combinazione lineare convessa, ovvero

$$T_c = c\,T_1 + (1-c)\,T_2,$$

dove $c \in (0,1)$; infatti si ha: $\mathbf{E}_\theta(T_c) = c\,\mathbf{E}_\theta(T_1) + (1-c)\,\mathbf{E}_\theta(T_2) = \xi(\theta)$.

Inoltre, se T è uno stimatore corretto di θ, in generale $\xi(T)$ *non* è uno stimatore corretto di $\xi(\theta)$, tranne nel caso in cui ξ sia una funzione *lineare* di θ, ovvero $\xi(\theta) = a\theta + b$ per opportune costanti $a, b \in \mathbb{R}$: infatti

$$\mathbf{E}_\theta(\xi(T)) = \mathbf{E}_\theta(a\,T + b) = a\,\mathbf{E}_\theta(T) + b = a\,\theta + b = \xi(\theta).$$

Dunque $\xi(T)$ è uno stimatore corretto di $\xi(\theta)$.

ESEMPIO 4.2. ►

Rifacendoci all'*Esempio (4.1)* si provano facilmente i seguenti risultati.

1. Nel caso dell'*Esempio (1.3)(1)* il parametro θ coincide con la speranza della legge in questione; considerando la statistica $T = \overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ si ha

   $$\mathbf{E}_\theta(T) = \frac{1}{n}\sum_{i=1}^n \mathbf{E}_\theta(X_i) = \frac{n\theta}{n} = \theta,$$

   e dunque $T = \overline{X}_n$ è uno stimatore corretto di $\xi(\theta) = \theta$.

2. Nel caso dell'*Esempio (1.3)(2)* la funzione $\xi(\theta) = 1/\theta$ coincide con la speranza della legge in questione; considerando la statistica $T = \overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ si ha

   $$\mathbf{E}_\theta(T) = \frac{1}{n}\sum_{i=1}^n \mathbf{E}_\theta(X_i) = \frac{n}{n\theta} = \frac{1}{\theta},$$

   e dunque $T = \overline{X}_n$ è uno stimatore corretto di $\xi(\theta) = 1/\theta$.

◄

La correttezza è per uno stimatore una proprietà desiderabile, ma non irrinunciabile: infatti uno stimatore T può essere corretto e tuttavia fornire stime inattendibili.

ESEMPIO 4.3. ►

Si consideri un campione $\mathbf{X}^1_{\mathrm{P}(\theta)}$, $\theta > 0$. La v.a. $T = (-2)^X$ è uno stimatore corretto della funzione (positiva) $\xi(\theta) = e^{-3\theta}$, dove X è l'unico elemento del campione: infatti, per ogni $\theta > 0$,

$$\mathbf{E}_\theta(T) = \sum_{i=0}^{\infty}(-2)^i\,e^{-\theta}\,\frac{\theta^i}{i!} = e^{-\theta}\sum_{i=0}^{\infty}\frac{(-2\theta)^i}{i!} = e^{-\theta}\,e^{-2\theta} = e^{-3\theta}.$$

Si noti che $T > 0$ se X è pari e $T < 0$ se X è dispari; tuttavia è sempre $\xi(\theta) > 0$. ◄

Addirittura vi sono modelli statistici per i quali non esistono stimatori corretti del parametro.

ESEMPIO 4.4. ►

Si consideri un campione $\mathbf{X}^1_{\mathrm{e}(\theta)}$, $\theta > 0$. Si supponga che esista uno stimatore corretto di θ della forma $T = t(X)$,

[1] Qui per "differenti" si intende che esiste almeno un valore $\tilde\theta \in \Theta$ tale che $\mathbf{P}_{\tilde\theta}\{T_1 \neq T_2\} > 0$.

dove X è l'unico elemento del campione e la funzione t è non negativa (una richiesta non restrittiva). Se un tale stimatore esistesse si avrebbe, per ogni $\theta > 0$,

$$\theta = \mathbf{E}_\theta(T) = \int_0^\infty t(x)\, \theta\, e^{-\theta x}\, \mathrm{d}x,$$

ovvero $\int_0^\infty t(x)\, e^{-\theta x}\, \mathrm{d}x = 1$. Se $0 < \theta_1 < \theta_2$ sono due diversi valori del parametro sarebbe dunque

$$\int_0^\infty t(x)\, \left(e^{-\theta_1 x} - e^{-\theta_2 x}\right)\, \mathrm{d}x = 0,$$

ciò che implicherebbe $t(x) = 0$ q.o., cioè $T = 0$, ma in tal caso T non potrebbe essere corretto. ◀

ESEMPIO 4.5. ▶

Si consideri un campione $\mathbf{X}^1_{\mathrm{P}(\theta)}$, $\theta > 0$. Si supponga che esista uno stimatore corretto della funzione $\xi(\theta) = 1/\theta$ della forma $T = t(X)$, dove X è l'unico elemento del campione. Se un tale stimatore esistesse si avrebbe, per ogni $\theta > 0$,

$$\frac{1}{\theta} = \mathbf{E}_\theta(T) = \sum_{i=0}^\infty t(i)\, e^{-\theta}\, \frac{\theta^i}{i!},$$

ovvero

$$\sum_{i=0}^\infty t(i)\, \frac{\theta^{i+1}}{i!} = e^\theta = \sum_{i=0}^\infty \frac{\theta^i}{i!},$$

il che è assurdo, in quanto due serie di potenze sono uguali se, e solo se, sono uguali i coefficienti delle potenze con lo stesso esponente. Dunque T non può essere corretto. ◀

L'Esempio che segue chiarisce alcuni punti importanti.

ESEMPIO 4.6. ▶

Siano $a, b \in \mathbb{R}$, $a < b$, e si consideri un campione $\mathbf{X}^n_{\mathrm{U}_\mathrm{c}(a,b)}$. Mostriamo ora che

1. la v.a. $X_{(1)}$ è uno stimatore distorto del limite inferiore della distribuzione α_F (cioè a);
2. la v.a. $X_{(n)}$ è uno stimatore distorto del limite superiore della distribuzione ω_F (cioè b).

Infatti, la legge di X è, per $x \in \mathbb{R}$,

$$f(x) = \frac{1}{b-a}\, \mathbb{I}_{(a,b)}(x),$$
$$F(x) = \frac{x-a}{b-a}\, \mathbb{I}_{[a,b)}(x) + \mathbb{I}_{[b,\infty)}(x),$$

sicché, usando il *Corollario (2.3)* e il *Corollario (2.6)*, otteniamo, per $x \in \mathbb{R}$,

$$f_{(1)}(x) = \frac{n}{(b-a)^n}\, (b-x)^{n-1}\, \mathbb{I}_{(a,b)}(x),$$
$$f_{(n)}(x) = \frac{n}{(b-a)^n}\, (x-a)^{n-1}\, \mathbb{I}_{(a,b)}(x),$$

Infine, dal calcolo della speranza di $X_{(1)}$ e di $X_{(n)}$, cioè

$$\mathbf{E}\big(X_{(1)}\big) = a + \frac{b-a}{n+1} \quad \left(\xrightarrow{n\to\infty}\ a\right),$$
$$\mathbf{E}\big(X_{(n)}\big) = b - \frac{b-a}{n+1} \quad \left(\xrightarrow{n\to\infty}\ b\right),$$

si ottiene il risultato indicato. Dunque, *in media*, $X_{(1)}$ *sovrastima* a mentre $X_{(n)}$ *sottostima* b: ciò è una conseguenza del fatto che, q.c., $a \le X_{(1)} < X_{(n)} \le b$. Si osservi tuttavia che, asintoticamente (cioè per $n \to \infty$), entrambi gli stimatori sono corretti.

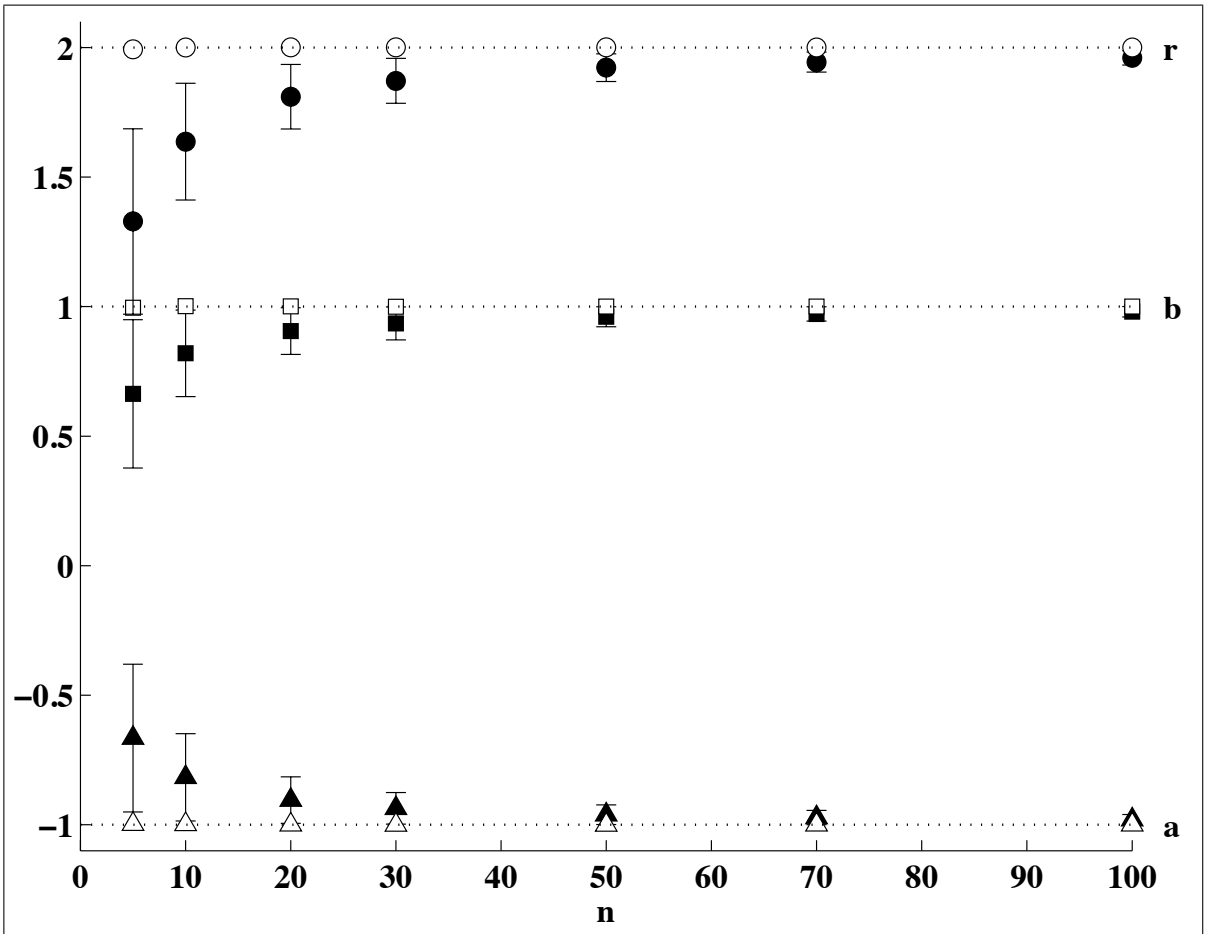

Figura 4.1: stima dei parametri $a = -1$, $b = 1$ e $r = b - a$ della legge $\mathtt{U_c}(-1, 1)$ usando gli stimatori dell'*Esempio (4.6)* in funzione dell'ampiezza n del campione. Sono inoltre mostrate le barre d'errore relative alla stima dei parametri a, b e r, aventi lunghezza pari ad una deviazione standard empirica.

In *Figura (4.1)* sono mostrati i valori medi campionari $\mathbf{E}\big(X_{(1)}\big)$ (*triangoli pieni*) e $\mathbf{E}\big(X_{(n)}\big)$ (*quadrati pieni*) — calcolati su 10000 simulazioni indipendenti della legge $\mathtt{U_c}(a, b)$, con $a = -1$ e $b = +1$ — al crescere dell'ampiezza n del campione: come previsto tutti gli stimatori considerati sono distorti, ma la distorsione diminuisce con l'aumentare di n e tende a zero per $n \to \infty$.

Infine, si consideri la v.a. $R = X_{(n)} - X_{(1)}$ (si veda l'*Esempio (2.6)*), che risulta essere uno stimatore distorto del *campo di variazione* $r = b - a$ della v.a. X:

$$\mathbf{E}(R) = \mathbf{E}\big(X_{(n)}\big) - \mathbf{E}\big(X_{(1)}\big) = b - a - 2\,\frac{b-a}{n+1} = r\,\frac{n-1}{n+1} = r\,\left(1 - \frac{2}{n+1}\right) \quad \left(\xrightarrow{n\to\infty} r\right).$$

Dunque R sottostima r e la distorsione tende a zero al crescere di n. In *Figura (4.1)* sono mostrati i valori medi campionari $\mathbf{E}(R)$ (*cerchi pieni*).

Si osservi tuttavia che lo stimatore

$$R' = \frac{n+1}{n-1}\,R$$

risulta essere corretto, cioè $\mathbf{E}(R') = r$. In *Figura (4.1)* sono mostrati i valori medi campionari $\mathbf{E}(R')$ (*cerchi vuoti*), che consentono di apprezzare la sostanziale correttezza dello stimatore R'.

Ora, essendo riusciti a stimare il parametro r senza distorsione, è immediato tentare di stimare correttamente sia a sia b. Introducendo i nuovi stimatori

$$X'_{(1)} = X_{(1)} - \frac{R}{n-1} \qquad \text{e} \qquad X'_{(n)} = X_{(n)} + \frac{R}{n-1},$$

è facile dimostrare che

$$\mathbf{E}\Big(X'_{(1)}\Big) = a + \frac{r}{n+1} - \frac{r}{n+1} = a,$$
$$\mathbf{E}\Big(X'_{(n)}\Big) = b - \frac{r}{n+1} + \frac{r}{n+1} = b,$$

ovvero che $X'_{(1)}$ e $X'_{(n)}$ sono stimatori corretti. In *Figura (4.1)* sono mostrati i valori medi campionari $\mathbf{E}\Big(X'_{(1)}\Big)$ (*triangoli vuoti*) e $\mathbf{E}\Big(X'_{(n)}\Big)$ (*quadrati vuoti*), che consentono di apprezzare la sostanziale correttezza dei nuovi stimatori introdotti. ◀

4.1.1 Media e varianza campionaria

Si consideri un campione $\mathbf{X}_{\mu_\theta}^n$ e si supponga che il parametro $\theta \in \Theta \subseteq \mathbb{R}$ abbia il significato di *speranza* di μ_θ (che si suppone esista finita) cioè, per ogni $\theta \in \Theta$, $\mathbf{E}_\theta(X) = \int_{\mathbb{R}} x \, \mathrm{d}\mu_\theta(x) = \theta$. Uno stimatore assai usato di θ è la *media campionaria* (o *empirica*) $\overline{X}_n$.

DEFINIZIONE 4.4 (Media campionaria)**.** La *media campionaria* $\overline{X}_n$ è definita mediante

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i. \tag{4.3}$$

NOTA 4.2.

La media campionaria $\overline{X}_n$ è una statistica e quindi una v.a. (e non un semplice valore numerico).

Una proprietà fondamentale della media campionaria è la seguente.

PROPOSIZIONE 4.1. *La media campionaria è uno stimatore corretto della speranza di μ_θ, cioè per ogni $\theta \in \Theta$*

$$\mathbf{E}_\theta\big(\overline{X}_n\big) = \mathbf{E}_\theta(X) = \theta. \tag{4.4}$$

DIM. Nelle ipotesi enunciate in precedenza si ha, per ogni $\theta \in \Theta$,

$$\mathbf{E}_\theta\big(\overline{X}_n\big) = \mathbf{E}_\theta\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{E}_\theta(X_i) = \frac{n\theta}{n} = \theta,$$

cioè la tesi. □

Evidentemente la media campionaria $\overline{X}_n$ è uno *stimatore lineare corretto* della speranza di μ_θ, corrispondente alla scelta $c_i = \frac{1}{n}$ per ogni indice i nell'*equazione (4.1)*. Se $\theta_2^2 = \mathbf{V}_\theta(X)$ è la varianza di μ_θ (che si suppone esista finita), un semplice calcolo porta al seguente risultato.

PROPOSIZIONE 4.2. *La varianza della media campionaria $\overline{X}_n$ è pari a*

$$\mathbf{V}_\theta\big(\overline{X}_n\big) = \frac{\mathbf{V}_\theta(X)}{n} = \frac{\theta_2^2}{n}. \tag{4.5}$$

DIM. Nelle ipotesi enunciate in precedenza si ha, per ogni $\theta \in \Theta$,

$$\mathbf{V}_\theta\big(\overline{X}_n\big) = \mathbf{V}_\theta\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} \mathbf{V}_\theta(X_i) = \frac{n\theta_2^2}{n^2} = \frac{\theta_2^2}{n},$$

cioè la tesi. □

Se, nelle ipotesi enunciate in precedenza, si vuole stimare la varianza θ_2^2 di μ_θ (che si suppone esista finita) si ricorre spesso alla *varianza campionaria* (o *empirica*) $\overline{S}_n^2$.

DEFINIZIONE 4.5 (Varianza campionaria)**.** La *varianza campionaria* $\overline{S}_n^2$ è definita mediante

$$\overline{S}_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}_n\right)^2. \tag{4.6}$$

NOTA 4.3.

La varianza campionaria $\overline{S}_n^2$ è una statistica e quindi una v.a. (e non un semplice valore numerico).

Una proprietà fondamentale della varianza campionaria è la seguente.

PROPOSIZIONE 4.3. *La varianza campionaria è uno stimatore corretto della varianza di μ_θ, cioè per ogni $\theta \in \Theta$*

$$\mathbf{E}_\theta\left(\overline{S}_n^2\right) = \mathbf{V}_\theta(X) = \theta_2^2. \tag{4.7}$$

DIM. Nelle ipotesi enunciate in precedenza si ha, per ogni $\theta \in \Theta$,

$$\mathbf{E}_\theta\left(\overline{S}_n^2\right) = \mathbf{E}_\theta\left(\frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}_n\right)^2\right) = \frac{1}{n-1} \sum_{i=1}^{n} \mathbf{E}_\theta\left(\left(X_i - \overline{X}_n\right)^2\right).$$

Scrivendo $X_i - \overline{X}_n = X_i - \mathbf{E}_\theta(X) + \mathbf{E}_\theta(X) - \overline{X}_n$ e svolgendo i conti (tediosi) si ottiene la tesi. □

Se si suppone che esista finito il *momento centrale* del quarto ordine m_4' di μ_θ, cioè

$$m_4' = \mathbf{E}_\theta\left(\left(X - \mathbf{E}_\theta(X)\right)^4\right) < \infty,$$

un tedioso calcolo (non svolto) porta al seguente risultato.

PROPOSIZIONE 4.4. *La varianza della varianza campionaria $\overline{S}_n^2$ è pari a*

$$\mathbf{V}_\theta\left(\overline{S}_n^2\right) = \frac{1}{n}\left(m_4' - \frac{n-3}{n-1}\,\theta_2^4\right). \tag{4.8}$$

Accanto alla varianza campionaria si considera anche la cosiddetta *varianza campionaria non corretta* $\overline{\sigma}_n^2$.

DEFINIZIONE 4.6 (Varianza campionaria non corretta)**.** La *varianza campionaria non corretta* $\overline{\sigma}_n^2$ è definita mediante

$$\overline{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \overline{X}_n\right)^2. \tag{4.9}$$

OSSERVAZIONE 4.2.

Essendo $\overline{\sigma}_n^2 = \frac{n-1}{n}\,\overline{S}_n^2$ si ha, per ogni $\theta \in \Theta$

$$\mathbf{E}_\theta\left(\overline{\sigma}_n^2\right) = \frac{n-1}{n}\,\mathbf{E}_\theta\left(\overline{S}_n^2\right) = \frac{n-1}{n}\,\theta_2^2,$$

e dunque $\overline{\sigma}_n^2$ è uno stimatore distorto della varianza di μ_θ (in generale $\overline{\sigma}_n^2$ *sottostima* θ_2^2). Tuttavia, asintoticamente (cioè per $n \to \infty$), anche $\overline{\sigma}_n^2$ è uno stimatore corretto.

4.2 Confronto di stimatori

Un criterio assai diffuso per confrontare stimatori si basa sull'uso dello *scarto quadratico medio*, o *rischio (quadratico)*. Nel seguito supporremo che gli stimatori in gioco appartengano a $\mathfrak{D}^2$, ovvero che siano (almeno) di quadrato sommabile.

DEFINIZIONE 4.7 (Rischio (quadratico))**.** Dato lo stimatore $T \in \mathfrak{D}^2$ della funzione $\theta \mapsto \xi(\theta)$, il *rischio (quadratico)* $\mathbf{R}$ è definito come, per ogni $\theta \in \Theta$,

$$\mathbf{R}_\theta(T) = \mathbf{E}_\theta\Big((T - \xi(\theta))^2\Big). \tag{4.10}$$

Intuitivamente $\mathbf{R}$ rappresenta la speranza di una *funzione di perdita* quadratica.

OSSERVAZIONE 4.3 (Distorsione)**.**
Il rischio $\mathbf{R}_\theta(T)$ si scrive anche nella forma, per ogni $\theta \in \Theta$,

$$\mathbf{R}_\theta(T) = \mathbf{E}_\theta\Big((T - \mathbf{E}_\theta(T) + \mathbf{E}_\theta(T) - \xi(\theta))^2\Big) = \mathbf{V}_\theta(T) + (\mathbf{E}_\theta(T) - \xi(\theta))^2. \tag{4.11}$$

La differenza

$$d_\theta(T) = \mathbf{E}_\theta(T) - \xi(\theta) \tag{4.12}$$

rappresenta la *distorsione* dello stimatore T: essa è nulla per uno stimatore *corretto*, sicché in tal caso il rischio di T coincide con la varianza di T.

Volendo procedere al confronto di stimatori, è possibile introdurre nella classe $\mathfrak{D}^2$ un *ordinamento parziale* come segue.

DEFINIZIONE 4.8 (Ordinamento)**.** Nella classe $\mathfrak{D}^2$ degli stimatori di quadrato sommabile si dice che

1. U è *preferibile* a V se, e solo se, per ogni $\theta \in \Theta$, $\mathbf{R}_\theta(U) \leq \mathbf{R}_\theta(V)$;
2. U è *migliore* di V se, e solo se, U è preferibile a V e inoltre esiste almeno un elemento $\tilde{\theta} \in \Theta$ per il quale sia $\mathbf{R}_{\tilde{\theta}}(U) < \mathbf{R}_{\tilde{\theta}}(V)$;
3. U è *ammissibile* se, e solo se, non esistono stimatori migliori di U;
4. U è *ottimale* se, e solo se, è preferibile ad ogni altro stimatore.

Evidentemente l'ordinamento è *parziale* in quanto non sempre è possibile confrontare tra loro gli stimatori: ad esempio potrebbe accadere che per alcuni valori del parametro θ sia $\mathbf{R}_\theta(U) \leq \mathbf{R}_\theta(V)$, mentre per altri valga la disuguaglianza opposta.

Il risultato che segue mette in luce una proprietà importante della media campionaria.

PROPOSIZIONE 4.5. *Sia $\mathbf{X}^n_{\mu_\theta}$, $\theta \in \Theta$, un campione di v.a. in $\mathfrak{D}^2$. La media campionaria $\overline{X}_n$ è* ottimale *nella classe degli stimatori lineari corretti di $\mathbf{E}_\theta(X)$.*

ESEMPIO 4.7. ▸
Si consideri il campione $\mathbf{X}^n_{\mathbb{N}\left(\theta_1,\theta_2^2\right)}$ e si voglia stimare la media θ_1. Indicando per semplicità di notazione $\theta = \left(\theta_1, \theta_2^2\right)$, per quanto visto nella *Sezione 4.1.1* la media campionaria $\overline{X}_n$ è uno stimatore corretto di θ_1, di rischio

$$\mathbf{R}_\theta(\overline{X}_n) = \frac{\theta_2^2}{n}.$$

Considerando ora lo stimatore $\overline{X}_k$, $k < n$, costruito utilizzando un sotto-campione di ampiezza k, si ha che anch'esso è corretto, di rischio

$$\mathbf{R}_\theta(\overline{X}_k) = \frac{\theta_2^2}{k} > \frac{\theta_2^2}{n} = \mathbf{R}_\theta(\overline{X}_n).$$

Dunque $\overline{X}_n$ è migliore di $\overline{X}_k$ essendo, per ogni $\theta \in \Theta$, $\mathbf{R}_\theta(\overline{X}_n) < \mathbf{R}_\theta(\overline{X}_k)$. Si scelga ora lo stimatore $T_n = c\overline{X}_n$, $c \in (0, 1)$. Evidentemente T_n non è corretto (essendo $c \neq 1$) e ha rischio

$$\mathbf{R}_\theta(T_n) = c^2\frac{\theta_2^2}{n} + (1 - c)^2\theta_1^2$$

che dipende da θ_1; in particolare si ha $\mathbf{R}_\theta(T_n) \leq \mathbf{R}_\theta(\overline{X}_n)$ se, e solo se,

$$\theta_1^2 \leq \frac{1+c}{1-c}\,\frac{\theta_2^2}{n}.$$

Dunque in generale non si può dire quale dei due stimatori sia preferibile, anzi, essi non sono nemmeno confrontabili. Tuttavia si noti che $\lim_{n\to\infty} \mathbf{R}_\theta(\overline{X}_n) = 0$ mentre $\lim_{n\to\infty} \mathbf{R}_\theta(T_n) = (1-c)^2\theta_1^2 > 0$: dunque $\overline{X}_n$ è "preferibile" a T_n per grandi campioni.

In *Figura (4.2)* si mostra, fissato $\theta_2^2 = 1$, la regione in cui $\mathbf{R}_\theta(T_n) \leq \mathbf{R}_\theta(\overline{X}_n)$, al variare di $c \in (0,1)$ e per due diversi valori di n (rispettivamente, $n = 4$ in (*a*) e $n = 40$ in (*b*)): si noti che, qualunque sia il valore di c, la regione in cui $\mathbf{R}_\theta(T_n) \leq \mathbf{R}_\theta(\overline{X}_n)$ si restringe al crescere di n.

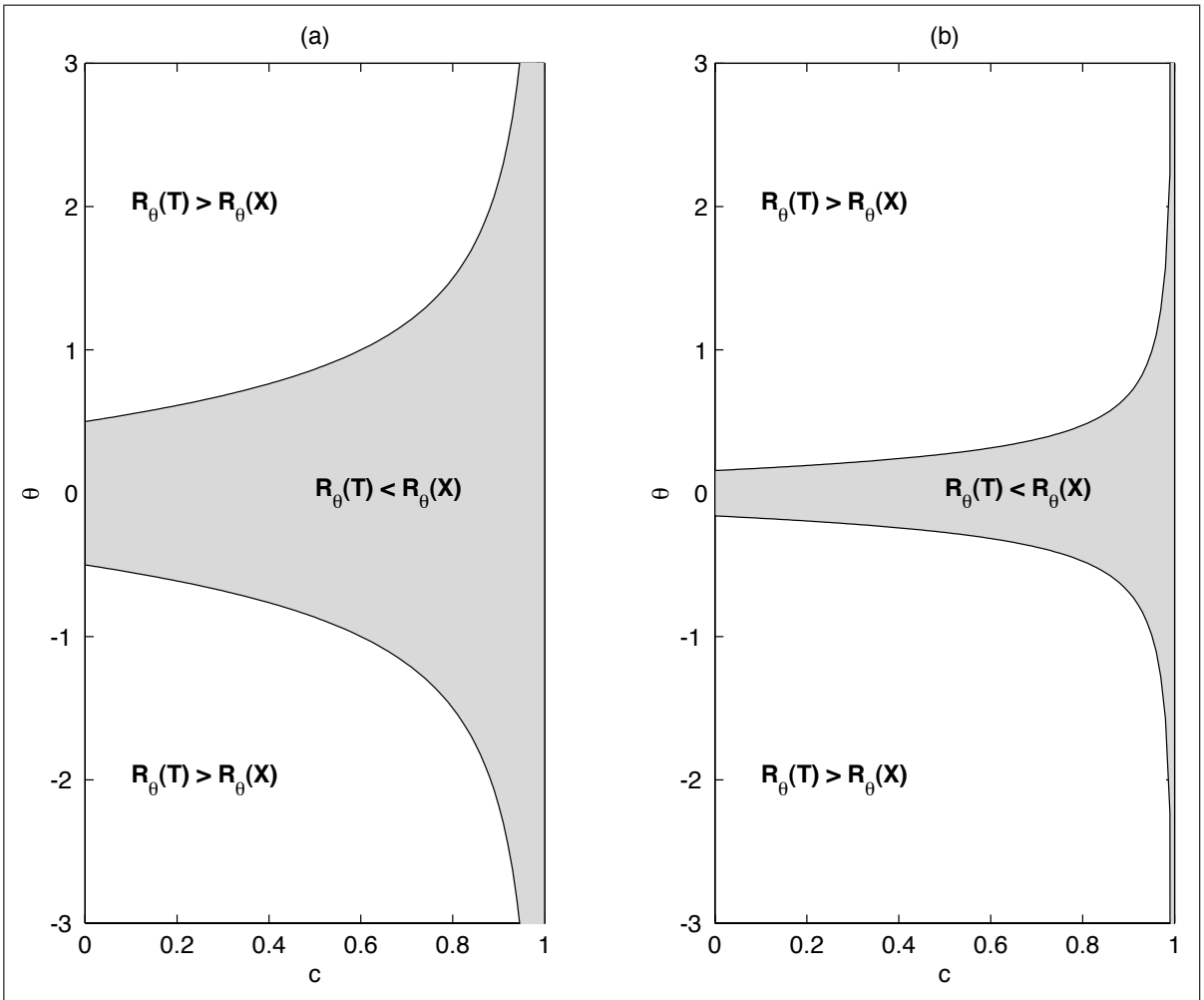

Figura 4.2: confronto delle regioni ove $\mathbf{R}_\theta(T_n) \leq \mathbf{R}_\theta(\overline{X}_n)$ per gli stimatori dell'*Esempio (4.7)*.

Volendo ora stimare la varianza θ_2^2, per quanto visto nella *Sezione 4.1.1* la varianza campionaria $\overline{S}_n^2$ è uno stimatore corretto di θ_2^2, di rischio

$$\mathbf{R}_\theta(\overline{S}_n^2) = \frac{2}{n-1}\,\theta_2^4,$$

essendo $m_4' = 3\theta_2^4$. Usando la varianza campionaria non corretta $\overline{\sigma}_n^2$ si ha invece

$$\mathbf{R}_\theta(\overline{\sigma}_n^2) = \frac{2n-1}{n^2}\,\theta_2^4.$$

Dunque, benché distorto, $\overline{\sigma}_n^2$ risulta essere uno stimatore migliore di $\overline{S}_n^2$ essendo, per ogni $\theta \in \Theta$, $\mathbf{R}_\theta(\overline{\sigma}_n^2) < \mathbf{R}_\theta(\overline{S}_n^2)$. ◀

ESEMPIO 4.8. ▶
Si consideri il campione $\mathbf{X}_{\mathrm{P}(\theta)}^n$, $\theta > 0$, e si voglia stimare il parametro θ, che coincide sia con la media sia con la varianza della legge in questione. Per quanto visto nella *Sezione 4.1.1* sia la media campionaria $\overline{X}_n$ sia la varianza campionaria $\overline{S}_n^2$ sono stimatori corretti di θ; volendone calcolare le rispettive funzioni di rischio si ha

$$\mathbf{R}_\theta(\overline{X}_n) = \frac{\theta}{n} \qquad \text{e} \qquad \mathbf{R}_\theta(\overline{S}_n^2) = \frac{\theta}{n}\left(1 + \frac{2n}{n-1}\,\theta\right),$$

essendo in questo caso $m_4' = \theta + 3\theta^2$. Dunque vale, per ogni $\theta > 0$, $\mathbf{R}_\theta(\overline{X}_n) < \mathbf{R}_\theta(\overline{S}_n^2)$, sicché $\overline{X}_n$ è migliore di $\overline{S}_n^2$. ◀

ESEMPIO 4.9. ▶
La legge Esponenziale è usata spesso nelle applicazioni quale modello della durata dei componenti di un sistema. Si consideri il campione $\mathbf{X}_{\mathrm{e}(\theta)}^n$, $\theta > 0$, e si voglia stimare la media della distribuzione $\xi(\theta) = \frac{1}{\theta}$. Per quanto visto nella *Sezione 4.1.1* la media campionaria $\overline{X}_n$ è uno stimatore corretto, con funzione di rischio pari a $\mathbf{R}_\theta(\overline{X}_n) = \frac{1}{n\theta^2}$. Considerando invece lo stimatore $T = nX_{(1)}$ si ricava facilmente (usando il risultato dell'*Esempio (2.4)*) che T è corretto, con funzione di rischio pari a $\mathbf{R}_\theta(T) = \frac{1}{\theta^2}$. Dunque $\overline{X}_n$ è preferibile a T. ◀

NOTA 4.4.
Nell'*Esempio (4.9)* per il calcolo di $\overline{X}_n$ è necessario attendere che tutte le durate dei componenti del sistema siano osservate, mentre per il calcolo di T è sufficiente osservare il valore di $X_{(1)}$, cioè la durata del componente che si guasta per primo (con ovvi vantaggi dal punto di vista pratico).

Il Teorema che segue chiarisce un punto importante.

TEOREMA 4.1. *Uno stimatore ammissibile $T \in \mathfrak{D}_c^2$ se esiste è unico.*

4.3 Disuguaglianza di Fréchet-Cramér-Rao

Per quanto visto finora uno stimatore (corretto) è tanto migliore quanto più piccolo è il suo rischio. Viene dunque spontaneo chiedersi quanto tale rischio possa essere ridotto, se ne esista una limitazione inferiore e se questa possa essere raggiunta. In questa Sezione cercheremo di rispondere a tali domande.

Nel seguito supporremo che $\mathfrak{M} = \big(\Omega, \mathcal{F}, \{\mathbf{P}_\theta\}_{\theta\in\Theta}\big)$ sia un modello statistico dominato *regolare*, dove Θ è un aperto di $\mathbb{R}$ (per semplicità); supporremo inoltre che, per ogni $\omega \in \Omega$, la funzione $\theta \mapsto \mathbf{L}_\theta(\omega)$ sia strettamente positiva e differenziabile (e dunque lo stesso vale per la funzione $\theta \mapsto f_\theta(\omega)$).

LEMMA 4.1. *Si supponga che la v.a. $\frac{\partial}{\partial\theta}\ln \mathbf{L}_\theta$ sia di quadrato sommabile. Allora, per ogni $\theta \in \Theta$,*

$$\mathbf{E}_\theta\left(\frac{\partial}{\partial\theta}\ln \mathbf{L}_\theta\right) = 0. \tag{4.13}$$

DIM. Evidentemente si ha, per ogni $\theta \in \Theta$,

$$\begin{aligned}
\mathbf{E}_\theta\left(\frac{\partial}{\partial\theta}\ln \mathbf{L}_\theta\right) &= \int_\Omega \frac{\partial}{\partial\theta}\ln \mathbf{L}_\theta \;\mathrm{d}\mathbf{P}_\theta \\
&= \int_\Omega \frac{\frac{\partial}{\partial\theta}\mathbf{L}_\theta}{\mathbf{L}_\theta}\,\mathbf{L}_\theta \;\mathrm{d}\mu \\
&= \int_\Omega \frac{\partial}{\partial\theta}\mathbf{L}_\theta \;\mathrm{d}\mu \\
&= \frac{\partial}{\partial\theta}\int_\Omega \mathbf{L}_\theta \;\mathrm{d}\mu \\
&= \frac{\partial}{\partial\theta}\int_\Omega \mathrm{d}\mathbf{P}_\theta = 0,
\end{aligned}$$

ove μ è la misura dominante e si sono usate le ipotesi che $\mathbf{L}_\theta$ sia strettamente positiva e che il modello sia regolare (cioè il supporto di $\mathbf{L}_\theta$ non dipende da θ). □

OSSERVAZIONE 4.4.

In base al *Lemma (4.1)* si ha che

$$\mathbf{E}_\theta\left(\frac{\partial}{\partial\theta}\ln\mathbf{L}_\theta\right)=\int_\Omega\left(\frac{\partial}{\partial\theta}\ln\mathbf{L}_\theta\right)\mathbf{L}_\theta\,\mathrm{d}\mu=0.$$

Derivando quest'ultima relazione rispetto a θ — nell'ipotesi che ciò sia legittimo — si ottiene

$$\begin{aligned}0&=\int_\Omega\left(\frac{\partial^2}{\partial\theta^2}\ln\mathbf{L}_\theta\right)\mathbf{L}_\theta\;\mathrm{d}\mu+\int_\Omega\left(\frac{\partial}{\partial\theta}\ln\mathbf{L}_\theta\right)\frac{\partial\mathbf{L}_\theta}{\partial\theta}\,\mathrm{d}\mu\\&=\mathbf{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\ln\mathbf{L}_\theta\right)+\int_\Omega\left(\frac{\partial}{\partial\theta}\ln\mathbf{L}_\theta\right)\left(\frac{1}{\mathbf{L}_\theta}\frac{\partial\mathbf{L}_\theta}{\partial\theta}\right)\mathbf{L}_\theta\;\mathrm{d}\mu\\&=\mathbf{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\ln\mathbf{L}_\theta\right)+\mathbf{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\ln\mathbf{L}_\theta\right)^2\right),\end{aligned}$$

da cui si ricava

$$\mathbf{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\ln\mathbf{L}_\theta\right)^2\right)=-\mathbf{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\ln\mathbf{L}_\theta\right).$$

L'*Informazione di Fisher* è fondamentale nell'ambito della Statistica Matematica.

DEFINIZIONE 4.9 (Informazione di Fisher)**.** Il numero reale positivo

$$\mathcal{I}(\theta)=\mathbf{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\ln\mathbf{L}_\theta\right)^2\right)\tag{4.14}$$

è detto *Informazione di Fisher*.

Grazie all'*Osservazione (4.4)*, sotto opportune condizioni si ha $\mathcal{I}(\theta)=-\mathbf{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\ln\mathbf{L}_\theta\right)$.

OSSERVAZIONE 4.5.

Considerando un campione $\mathbf{X}_{\mu_\theta}^n$ è possibile mostrare che l'Informazione di Fisher del campione $\mathcal{I}_\mathbf{X}(\theta)$ è *lineare* in n, ovvero, per ogni $\theta\in\Theta$,

$$\mathcal{I}_\mathbf{X}(\theta)=n\,\mathcal{I}(\theta).\tag{4.15}$$

Infatti, poiché la verosimiglianza del campione vale $\mathbf{L}_\mathbf{X}=\prod_{i=1}^n\mathbf{L}_\theta\circ X_i$, osservando che

$$\mathcal{I}_\mathbf{X}(\theta)=\mathbf{E}_\theta\left(\left(\sum_{i=1}^n\frac{\partial}{\partial\theta}\ln\mathbf{L}_\theta\circ X_i\right)^2\right),$$

e sviluppando opportunamente il quadrato tra parentesi si giunge al risultato voluto. Si noti che l'"informazione" $\mathcal{I}_\mathbf{X}(\theta)$ fornita da un campione di ampiezza $n>1$ è maggiore di $\mathcal{I}(\theta)$.

Siamo ora in grado di dimostrare un risultato fondamentale.

TEOREMA 4.2 (Disuguaglianza di Fréchet-Cramér-Rao (FCR))**.** *Sia $T\in\mathfrak{D}_c^2$ uno stimatore corretto di quadrato sommabile di una funzione $\xi\colon\Theta\to\mathbb{R}$. Se tanto $\frac{\partial}{\partial\theta}\ln\mathbf{L}_\theta$ quanto $\frac{\partial}{\partial\theta}\mathbf{L}_\theta$ sono di quadrato sommabile allora, per ogni $\theta\in\Theta$,*

$$\mathbf{R}_\theta(T)=\mathbf{V}_\theta(T)\geq\frac{\left(\frac{\partial}{\partial\theta}\xi(\theta)\right)^2}{\mathcal{I}(\theta)}.\tag{4.16}$$

DIM. Per ogni $\theta \in \Theta$ si ha

$$\begin{aligned}\frac{\partial}{\partial\theta}\mathbf{E}_\theta(T) &= \frac{\partial}{\partial\theta}\int_\Omega T\ \mathrm{d}\mathbf{P}_\theta \\ &= \frac{\partial}{\partial\theta}\int_\Omega T\,\mathbf{L}_\theta\ \mathrm{d}\mu \\ &= \int_\Omega T\,\frac{\partial}{\partial\theta}\ln\mathbf{L}_\theta\ \mathrm{d}\mathbf{P}_\theta \\ &= \mathbf{E}_\theta\left(T\,\frac{\partial}{\partial\theta}\ln\mathbf{L}_\theta\right) \\ &= \mathbf{E}_\theta\left(\frac{\partial}{\partial\theta}\ln\mathbf{L}_\theta\cdot(T-\mathbf{E}_\theta(T))\right),\end{aligned}$$

essendo la v.a. $\frac{\partial}{\partial\theta}\ln\mathbf{L}_\theta$ centrata. Pertanto, usando la *disuguaglianza di Schwartz*[2] si ha, per ogni $\theta\in\Theta$,

$$\frac{\partial}{\partial\theta}\mathbf{E}_\theta(T) \le \sqrt{\mathbf{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\ln\mathbf{L}_\theta\right)^2\right)}\sqrt{\mathbf{V}_\theta(T)},$$

cioè la tesi. □

Se ξ è la funzione identità, cioè T è uno stimatore corretto del parametro θ, allora la disuguaglianza FCR si può scrivere nella forma

$$\mathbf{R}_\theta(T) = \mathbf{V}_\theta(T) \ge \frac{1}{\mathcal{I}(\theta)}, \tag{4.17}$$

dove $\frac{1}{\mathcal{I}(\theta)}$ rappresenta il *limite inferiore* della disuguaglianza FCR.

DEFINIZIONE 4.10 (Stimatore efficiente)**.** Uno stimatore corretto di quadrato sommabile $T\in\mathfrak{D}_c^2$ di una funzione $\xi\colon\Theta\to\mathbb{R}$ si dice *efficiente* (o *efficace*) se il suo rischio coincide con la limitazione inferiore stabilita dalla disuguaglianza FCR.

OSSERVAZIONE 4.6.

Si consideri un modello esponenziale $\mathfrak{M}$, con Θ un aperto di $\mathbb{R}$. Poiché $\ln\mathbf{L}_\theta = \theta T - \chi(\theta)$ e, in base al *Teorema (1.1)*, T è uno stimatore corretto di $\chi'(\theta)$, l'Informazione di Fisher vale

$$\mathcal{I}(\theta) = \mathbf{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\ln\mathbf{L}_\theta\right)^2\right) = \mathbf{E}_\theta\Big((T-\chi'(\theta))^2\Big) = \mathbf{E}_\theta\Big((T-\mathbf{E}_\theta(T))^2\Big) = \mathbf{V}_\theta(T) = \chi''(\theta).$$

Dunque il rischio di T vale

$$\mathbf{R}_\theta(T) = \mathbf{V}_\theta(T) = \chi''(\theta) = \frac{(\chi''(\theta))^2}{\chi''(\theta)} = \frac{1}{\mathcal{I}(\theta)}\left(\frac{\partial}{\partial\theta}\chi'(\theta)\right)^2 = \frac{1}{\mathcal{I}(\theta)}\left(\frac{\partial}{\partial\theta}\xi(\theta)\right)^2,$$

cioè il *limite inferiore* della disuguaglianza FCR. Ne segue che T è uno stimatore efficiente di $\chi'(\theta)$.

Il Teorema che segue è di importanza fondamentale.

TEOREMA 4.3. *Se Θ è un aperto di $\mathbb{R}$ e se, dato il campione $\mathbf{X}_{\mu_\theta}^n$, la statistica $T = t(X_1,\dots,X_n)$ è uno stimatore corretto ed efficiente di una funzione $\xi\colon\Theta\to\mathbb{R}$, allora esistono una funzione misurabile $g\colon\mathbb{R}\to\mathbb{R}$ e una funzione $h\colon\Theta\to\mathbb{R}$ tali che per la densità f_θ della legge μ_θ vale*

$$\frac{\partial}{\partial\theta}\ln f_\theta(x) = \frac{h(\theta)}{n}\,(g(x)-\xi(\theta)), \tag{4.18}$$

[2] Si considerino le v.a. $X,Y\in\mathrm{L}^2(\Omega)$ e sia $g(t)=\mathbf{E}\Big((X+tY)^2\Big)\ge 0$, $t\in\mathbb{R}$. Se per ogni $t\in\mathbb{R}$ si ha $g(t)>0$ allora il discriminante dell'equazione quadratica corrispondente è negativo e dunque $(\mathbf{E}(XY))^2 < \mathbf{E}\big(X^2\big)\,\mathbf{E}\big(Y^2\big)$, cioè la *disuguaglianza di Schwartz*. Se invece si ha $g(t)=0$ per un qualche $t=t'\in\mathbb{R}$ allora $\mathbf{E}\Big((X+t'Y)^2\Big)=0$ e dunque, usando la *disuguaglianza di Markov* (ovvero $\mathbf{P}\{h(X)>x\}\le\mathbf{E}(h(X))/x$, dove $h>0$), vale $\mathbf{P}\{X=aY+b\}=1$ per opportuni valori $a,b\in\mathbb{R}$.

e inoltre

$$T = \frac{1}{n} \sum_{i=1}^{n} g \circ X_i. \tag{4.19}$$

OSSERVAZIONE 4.7.

Dall'*equazione (4.18)* scende immediatamente, per integrazione, che

$$f_\theta(x) = s(x)\, e^{p(\theta)q(x)+r(\theta)},$$

dove

$$s(x) = 1, \quad p(\theta) = \frac{1}{n} \int h, \quad q(x) = g(x), \quad r(\theta) = -\frac{1}{n} \int h\,\xi.$$

In altre parole, ricordando l'*Osservazione (1.3)*, il modello è *esponenziale*.

Gli esempi che seguono illustrano il calcolo di stimatori efficaci per i parametri di alcune delle principali leggi nell'ambito della Statistica Matematica; i risultati mostrati si basano sul *Teorema (1.1)*, sull'*Osservazione (4.6)* e sull'*Osservazione (4.7)*.

ESEMPIO 4.10 (Legge $\mathtt{Be}(\boldsymbol{\theta})$). ►

Sia $X \sim \mathtt{Be}(\theta)$, $\theta \in (0,1)$. Ricordando l'*Esempio (1.5)* il modello è esponenziale ed è possibile scrivere la densità di X nella forma canonica

$$f_{\theta'}(x) = e^{\theta' x - \ln\left(1+e^{\theta'}\right)},$$

dove $x \in \{0,1\}$ e $\theta' = \ln \frac{\theta}{1-\theta}$. La densità del campione vale dunque

$$f_{\theta'}(x_1, \ldots, x_n) = e^{\theta' \sum_{i=1}^n x_i - n \ln\left(1+e^{\theta'}\right)},$$

dove $x_i \in \{0,1\}$, $i = 1, \ldots, n$. Ne segue che la statistica $T = \sum_{i=1}^n X_i$ è uno stimatore corretto ed efficiente di

$$\frac{\partial}{\partial \theta'} \chi(\theta') = \frac{\partial}{\partial \theta'} \left(n \ln \left(1 + e^{\theta'} \right) \right) = n \, \frac{e^{\theta'}}{1+e^{\theta'}} = n\theta,$$

e dunque la media campionaria $\overline{X}_n = \frac{1}{n}\, T$ è uno stimatore corretto ed efficiente di θ. ◄

ESEMPIO 4.11 (Legge $\boldsymbol{\Gamma(\theta_1, \theta_2)}$). ►

Sia $X \sim \Gamma(\theta_1, \theta_2)$, $\theta_1, \theta_2 > 0$. Un semplice calcolo (si veda l'*Esempio (1.7)*) mostra che il modello è esponenziale ed è possibile scrivere la densità di X nella forma canonica

$$f_{\theta'}(x) = e^{\langle \theta', t(x) \rangle - \chi(\theta')},$$

dove $x > 0$, $\theta' = (\theta'_1, \theta'_2) = (\theta_1 - 1, \theta_2)$, $t(x) = (\ln x, -x)$ e $\chi(\theta') = \ln \Gamma(\theta'_1 + 1) - (\theta'_1 + 1) \ln \theta'_2$. Dunque la densità del campione vale

$$f_{\theta'}(x_1, \ldots, x_n) = e^{\left\langle \theta', \left(\sum_{i=1}^n \ln x_i, -\sum_{i=1}^n x_i\right) \right\rangle - n\chi(\theta')},$$

dove $x_i > 0$, $i = 1, \ldots, n$. Supponendo il parametro θ_1 noto si ha che la statistica $T_2 = -\sum_{i=1}^n X_i$ è uno stimatore corretto ed efficiente di

$$\frac{\partial}{\partial \theta'_2} \left(n\chi(\theta') \right) = n \frac{\partial}{\partial \theta'_2} \left(\ln \Gamma(\theta'_1 + 1) - (\theta'_1 + 1) \ln \theta'_2 \right) = -n \frac{\theta_1}{\theta_2},$$

e dunque la statistica $T'_2 = \frac{1}{\theta_1}\, \overline{X}_n$ è uno stimatore corretto ed efficiente di $\frac{1}{\theta_2}$. ◄

ESEMPIO 4.12. ►

Sia $X \sim \mathtt{U_c}(0, \theta)$, $\theta > 0$. Usando i risultati dell'*Esempio (4.6)*, con $a = 0$ e $b = \theta$, si vede facilmente che la statistica

$$T_n = \frac{n+1}{n}\, X_{(n)}$$

è uno stimatore corretto di θ, con funzione di rischio pari a

$$\mathbf{R}_\theta(T_n) = \mathbf{V}_\theta(T_n) = \frac{1}{n(n+2)}\,\theta^2.$$

Nell'intervallo $(0, \theta)$ si ha $\ln f_\theta(x) = -\ln\theta$ e dunque, in base alla *Definizione (4.9)*, l'Informazione di Fisher di X vale

$$\mathcal{I}(\theta) = \mathbf{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\ln f_\theta \circ X\right)^2\right) = \frac{1}{\theta^2},$$

sicché, in base all'*equazione (4.15)*, l'Informazione di Fisher del campione vale $\mathcal{I}_{\mathbf{X}}(\theta) = n\,\mathcal{I}(\theta) = \frac{n}{\theta^2}$. Dal *Teorema (4.2)* si ricava che il limite inferiore della disuguaglianza FCR vale $\frac{1}{\mathcal{I}_{\mathbf{X}}(\theta)} = \frac{\theta^2}{n}$ e dunque

$$\mathbf{R}_\theta(T_n) = \frac{1}{n(n+2)}\,\theta^2 < \frac{\theta^2}{n}.$$

Tale risultato non deve sorprendere in quanto il modello in questione non è *regolare*: infatti il supporto di f_θ dipende da θ (si veda l'*Esempio (1.9)*). ◀

4.4 Sufficienza e completezza

Nell'ambito della Statistica Matematica la ricerca di "buoni" stimatori spesso inizia considerando quali candidati privilegiati le statistiche cosiddette *sufficienti*, le cui proprietà saranno illustrate tra breve. Da un punto di vista pratico, al termine di un esperimento il lavoro dello statistico consiste nell'interpretare i dati raccolti, così da trarre conclusioni valide circa il fenomeno studiato. Spesso l'archiviazione dei dati grezzi risulta costosa, e l'analisi dell'intero campione può richiedere tempi lunghi e risorse di calcolo ingenti. Ciò spinge l'analista a lavorare solamente con opportune statistiche su cui basare le proprie inferenze senza che vi sia perdita di "informazione". In molti problemi statistici è possibile mostrare che, ai fini della stima dei parametri incogniti, esistono statistiche, dette appunto "sufficienti", contenenti la stessa informazione dell'intero campione.

DEFINIZIONE 4.11 (Statistica sufficiente)**.** Sia $\mathbf{X}^n_{F_\theta}$, $\theta \in \Theta$, un campione. Una statistica S si dice *sufficiente* per θ (o per la famiglia $\{F_\theta\}_{\theta\in\Theta}$ o, ancora, per il modello $\mathfrak{M}$ soggiacente) se la distribuzione condizionale del campione dato l'evento $\{S = s\}$ è indipendente da θ (tranne al più per un insieme A di misura nulla, ovvero tale che $\mathbf{P}_\theta\{S \in A\} = 0$ per ogni $\theta \in \Theta$).

NOTA 4.5.

Se S è una statistica sufficiente per la famiglia $\{F_\theta\}_{\theta\in\Theta}$, allora essa è sufficiente anche per la sotto-famiglia $\{F_\theta\}_{\theta\in\Theta'}$, dove $\Theta' \subseteq \Theta$.

Generalizzando la definizione di statistica sufficiente è possibile introdurre quella di *statistiche congiuntamente sufficienti*.

DEFINIZIONE 4.12 (Statistiche congiuntamente sufficienti)**.** Sia $\mathbf{X}^n_{F_\theta}$, $\theta \in \Theta$, un campione. Una famiglia di statistiche $S_1, \ldots, S_m$ si dice *congiuntamente sufficiente* per θ (o per la famiglia $\{F_\theta\}_{\theta\in\Theta}$ o, ancora, per il modello $\mathfrak{M}$ soggiacente) se la distribuzione condizionale del campione dato l'evento $\{S_1 = s_1, \ldots, S_m = s_m\}$ è indipendente da θ (tranne al più per un insieme A di misura nulla, ovvero tale che $\mathbf{P}_\theta\{\boldsymbol{S} \in A\} = 0$ per ogni $\theta \in \Theta$).

OSSERVAZIONE 4.8.

Il campione $X_1, \ldots, X_n$ è sempre congiuntamente sufficiente in quanto la sua distribuzione condizionale, dato il campione medesimo, è indipendente da θ. Lo stesso vale per le Statistiche d'Ordine $X_{(1)}, \ldots, X_{(n)}$: infatti, i soli valori che possono essere assunti dal campione $X_1, \ldots, X_n$ sono una delle $n!$ permutazioni degli elementi dell'insieme $\{x_{(1)}, \ldots, x_{(n)}\}$, ma essendo il campionamento casuale ciascuna di tali permutazioni ha la stessa probabilità, pari a $1/n!$.

Purtroppo nessuna delle definizioni date è costruttiva, in quanto per decidere se una statistica sia o no sufficiente è necessario anzitutto individuarla e quindi procedere alla verifica (spesso tediosa se non del tutto impraticabile) che essa sia effettivamente sufficiente. Considerando v.a. reali, vale il seguente *Criterio di fattorizzazione* di uso frequente.

PROPOSIZIONE 4.6 (Criterio di fattorizzazione)**.** *Si consideri il campione* $\mathbf{X}_{F_\theta}^n$, $\theta \in \Theta$. *La statistica* $S = S(X_1, \ldots, X_n)$ *è sufficiente se, e solo se, è possibile fattorizzare la densità* f_θ *del campione nel prodotto*

$$f_\theta(x_1, \ldots, x_n) = h(x_1, \ldots, x_n)\, g_\theta(S(x_1, \ldots, x_n)), \tag{4.20}$$

dove h *è una funzione non negativa del campione che non dipende da* θ*, e* g_θ *è una funzione non negativa che dipende da* θ *e dal campione solamente tramite* S*.*

Ovviamente, sia θ sia S possono essere quantità vettoriali. Il *Criterio di fattorizzazione* può essere esteso ad una famiglia di statistiche congiuntamente sufficienti. Esso è utile per mostrare che una statistica è sufficiente, ma non serve invece a dimostrare che una statistica non è sufficiente: il fatto di non riuscire a fattorizzare la densità del campione non significa che essa non possa essere fattorizzata! Gli Esempi che seguono illustrano il calcolo di statistiche sufficienti per alcuni modelli noti.

ESEMPIO 4.13. ▶

Si consideri il campione $\mathbf{X}_{\mathtt{Be}(\theta)}^n$, $\theta \in (0,1)$. Si può prendere $\Omega = \{0,1\}^n$, scegliere come σ-algebra $\mathcal{F}$ la famiglia delle parti $\mathcal{P}(\Omega)$, e prendere come misura dominante μ la "misura del contare" μ_c su $(\Omega, \mathcal{F})$. Dunque la densità f_θ del campione è

$$f_\theta(\boldsymbol{x}) = \theta^{\sum_{i=1}^n x_i}\, (1-\theta)^{n-\sum_{i=1}^n x_i} \prod_{i=1}^n \mathbb{I}_{\{0,1\}}(x_i).$$

Segue dal *Criterio di fattorizzazione* che, prendendo $h = \prod_{i=1}^n \mathbb{I}_{\{0,1\}}(x_i)$ e $g_\theta(s) = \theta^s\, (1-\theta)^{n-s}$, la statistica $S = \sum_{i=1}^n X_i$ è sufficiente.

In alternativa, si consideri la probabilità condizionale $\mathbf{P}_\theta\{\boldsymbol{X} = \boldsymbol{x} \mid S = s\}$, che vale

$$\mathbf{P}_\theta\{X_1 = x_1, \ldots, X_n = x_n \mid S = s\} = \begin{cases} \frac{\mathbf{P}_\theta\{X_1 = x_1, \ldots, X_n = x_n, S = s\}}{\binom{n}{s}\, \theta^s\, (1-\theta)^{n-s}}, & \text{se } \sum_{i=1}^n x_i = s \\ 0, & \text{altrimenti} \end{cases},$$

in quanto $S = \sum_{i=1}^n X_i \sim \mathtt{Bi}(n, \theta)$. Perciò, se $\sum_{i=1}^n x_i = s$ allora

$$\mathbf{P}_\theta\{X_1 = x_1, \ldots, X_n = x_n \mid S = s\} = \frac{\theta^{\sum_{i=1}^n x_i}\, (1-\theta)^{n-\sum_{i=1}^n x_i}}{\binom{n}{s}\, \theta^s\, (1-\theta)^{n-s}} = \frac{1}{\binom{n}{s}},$$

che non dipende da θ, e dunque la statistica $S = \sum_{i=1}^n X_i$ è sufficiente. ◀

Il risultato dell'*Esempio (4.13)* mostra che, in termini operativi, al fine di stimare la probabilità di ottenere "Testa" in un gioco "Testa-o-Croce" (ovvero il parametro θ della soggiacente legge di Bernoulli), è sufficiente conoscere il numero totale di "Teste" presenti nel campione e non l'esatta sequenza con cui queste si presentano.

ESEMPIO 4.14. ▶

Si consideri il campione $\mathbf{X}_{\mathtt{e}(\theta)}^n$, $\theta > 0$. Si può prendere $\Omega = (\mathbb{R}_+)^n$, scegliere come σ-algebra $\mathcal{F}$ i Boreliani $\mathcal{B}(\Omega)$ di Ω, e prendere come misura dominante μ la restrizione della misura di Lebesgue n-dimensionale λ_n a $(\Omega, \mathcal{F})$. Dunque la densità f_θ del campione è

$$f_\theta(\boldsymbol{x}) = \theta^n\, e^{-\theta \sum_{i=1}^n x_i} \prod_{i=1}^n \mathbb{I}_{\mathbb{R}_+}(x_i).$$

Segue dal *Criterio di fattorizzazione* che, prendendo $h = \prod_{i=1}^n \mathbb{I}_{\mathbb{R}_+}(x_i)$ e $g_\theta(s) = \theta^n\, e^{-\theta s}$, la statistica $S = \sum_{i=1}^n X_i$ è sufficiente. ◀

I risultati degli Esempi precedenti scendono immediatamente dalla seguente proprietà del modello esponenziale (canonico) che, ancora una volta, gioca un ruolo privilegiato; la dimostrazione è ovvia.

PROPOSIZIONE 4.7. *In un modello esponenziale canonico nel quale la verosimiglianza $\mathbf{L}_\theta$ rispetto alla misura dominante μ sia*

$$\mathbf{L}_\theta = e^{\langle \theta, S\rangle - \chi(\theta)} \tag{4.21}$$

la statistica S è (congiuntamente) sufficiente.

Di statistiche sufficienti ne possono esistere più d'una; anzi, in generale, vi è abbondanza di statistiche sufficienti, come mostra il risultato che segue.

PROPOSIZIONE 4.8. *Se S è una statistica sufficiente per $\mathfrak{M}$ e se $\varphi\colon \mathbb{R} \to \mathbb{R}$ è misurabile e invertibile, allora anche $\varphi \circ S$ è una statistica sufficiente per lo stesso modello.*

La conoscenza di statistiche sufficienti per un modello statistico dominato facilita notevolmente la ricerca di stimatori ottimali. I Teoremi che seguono sono molto importanti nella ricerca di "buoni" stimatori e mostrano l'uso che si fa in tale ricerca delle statistiche sufficienti.

TEOREMA 4.4 (Rao–Blackwell)**.** *Siano $\boldsymbol{X}$ un campione, S una statistica sufficiente per il modello $\mathfrak{M}$ soggiacente, e $T \in \mathfrak{D}_c^2$ uno stimatore corretto di quadrato sommabile della funzione $\xi\colon \Theta \to \mathbb{R}^d$. Allora*

$$U = \mathbf{E}(T \mid S) \tag{4.22}$$

è uno stimatore corretto di ξ preferibile a T. Inoltre, U è migliore di T a meno che T non sia già funzione di S, e in tal caso $U = T$.

NOTA 4.6.

Si osservi che, in quanto speranza condizionale, U è una v.a. (si veda il *Teorema (0.2)* e la *Definizione (0.9)*).

Dunque, se si ha uno stimatore T e se si conosce una statistica sufficiente S, il *Teorema (4.4)* asserisce che $U = \mathbf{E}(T \mid S)$ è uno stimatore preferibile a T e suggerisce, quindi, di cercare i "buoni" stimatori di ξ nella classe delle funzioni $\varphi \circ S$, con φ misurabile.

La *Tabella (4.1)* mostra alcune statistiche sufficienti "naturali" per svariati modelli statistici. Si osservi come, nel caso di modelli appartenenti alla famiglia esponenziale (ad esempio i modelli *(A)–(C)* e *(E)–(G)*), sia sufficiente ricondursi alla forma canonica per calcolare le statistiche sufficienti.

Valgono inoltre i seguenti commenti.

- Modello *(A)*: per $n = 1$ la legge Binomiale $\mathtt{Bi}(1, p)$ è quella Bernoulliana $\mathtt{Be}(p)$.
- Modello *(C)*: per $r = 1$ la legge Binomiale Negativa $\mathtt{BN}(1, p)$ è quella Geometrica $\mathtt{g}(p)$.
- Modello *(F)*: per $\alpha = 1$ la legge Gamma $\Gamma(1, \lambda)$ è quella Esponenziale $\mathtt{e}(\lambda)$; per $\alpha = \frac{n}{2}$, $n \in \mathbb{N}$, e $\lambda = \frac{1}{2}$ la legge Gamma $\Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$ è quella Chi-quadro $\chi^2(n)$.
- Modello *(G)*: per $\alpha = \beta = 1$ la legge Beta $\beta(1, 1)$ è quella Uniforme $\mathtt{U_c}(0, 1)$.

Un altro concetto di importanza fondamentale nella costruzione di stimatori ottimali è quello di *completezza*.

DEFINIZIONE 4.13 (Statistica completa)**.** Una statistica S sufficiente per il modello $\mathfrak{M}$ si dice *completa* per $\mathfrak{M}$ se l'unico stimatore non distorto di 0 che sia funzione di S è la statistica identicamente nulla con probabilità 1.

ESEMPIO 4.15. ▶

Si riconsideri il campione $\mathbf{X}^n_{\mathtt{Be}(\theta)}$, $\theta \in (0, 1)$ dell'*Esempio (4.13)*. La statistica $S = \sum_{i=1}^n X_i$ è completa: infatti, se φ è una funzione sommabile, si ha

$$0 = \mathbf{E}_\theta(\varphi(S)) = \sum_{s=0}^n \varphi(s) \binom{n}{s} \theta^s (1-\theta)^{n-s} = (1-\theta)^n \sum_{s=0}^n \varphi(s) \binom{n}{s} \left(\frac{\theta}{1-\theta}\right)^s$$

se, e solo se, $\varphi(s) = 0$ per $s = 0, \ldots, n$, in quanto il polinomio in $\frac{\theta}{1-\theta}$ è nullo se, e solo se, lo sono tutti i suoi coefficienti. ◀

Anche per quanto concerne la completezza il modello esponenziale (canonico) gioca un ruolo privilegiato, come si evince dal risultato che segue.

Modello		Parametri	Densità	Stat. (Cong.) Suff.
(A)	$\mathtt{Bi}(n,p)$	$\dfrac{n \in \mathbb{N} \text{ noto}}{p \in (0,1)}$	$\binom{n}{x}\, p^x\, \frac{(1-p)^n}{(1-p)^x}\, \mathbb{I}_{\{0,\ldots,n\}}(x)$	$S = \sum_{i=1}^N X_i$
(B)	$\mathtt{P}(\lambda)$	$\lambda \in \mathbb{R}_+$	$e^{-\lambda}\, \frac{\lambda^x}{x!}\, \mathbb{I}_{\mathbb{N}}(x)$	$S = \sum_{i=1}^N X_i$
(C)	$\mathtt{BN}(r,p)$	$\dfrac{r \in \mathbb{N} \text{ noto}}{p \in (0,1)}$	$\binom{r-1+x}{x}\, p^r\, (1-p)^x\, \mathbb{I}_{\mathbb{N}}(x)$	$S = \sum_{i=1}^N X_i$
(D)	$\mathtt{U_d}(1,n)$	$n \in \mathbb{N}$	$\frac{1}{n}\mathbb{I}_{\{1,\ldots,n\}}(x)$	$S = X_{(N)}$
(E)	$\mathtt{N}(\mu,\sigma)$	$\dfrac{\mu \in \mathbb{R}}{\sigma \in \mathbb{R}_+}$	$\frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{1}{2\sigma^2}\,(x-\mu)^2}\, \mathbb{I}_{\mathbb{R}}(x)$	$S_1 = \sum_{i=1}^N X_i$ $S_2 = \sum_{i=1}^N X_i^2$
(F)	$\Gamma(\alpha,\lambda)$	$\dfrac{\alpha \in \mathbb{R}_+}{\lambda \in \mathbb{R}_+}$	$\frac{\lambda^\alpha}{\Gamma(\alpha)}\, x^{\alpha-1}\, e^{-\lambda x}\, \mathbb{I}_{\mathbb{R}_+}(x)$	$S_1 = \sum_{i=1}^N \ln X_i$ $S_2 = \sum_{i=1}^N X_i$
(G)	$\beta(\alpha,\beta)$	$\dfrac{\alpha \in \mathbb{R}_+}{\beta \in \mathbb{R}_+}$	$\frac{x^{\alpha-1}\,(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}\, \mathbb{I}_{(0,1)}(x)$	$S_1 = \sum_{i=1}^N \ln X_i$ $S_2 = \sum_{i=1}^N \ln(1-X_i)$
(H)	$\mathtt{U_c}(0,\theta)$	$\theta \in \mathbb{R}_+$	$\frac{1}{\theta}\mathbb{I}_{(0,\theta)}(x)$	$S = X_{(N)}$
(I)	$\mathtt{Pa}(\alpha,\xi)$	$\dfrac{\alpha \in \mathbb{R}_+}{\xi \in \mathbb{R}_+}$	$\alpha\, \xi^\alpha\, x^{-(\alpha+1)}\, \mathbb{I}_{(\xi,\infty)}(x)$	$S_1 = \prod_{i=1}^N X_i$ $S_2 = X_{(1)}$

Tabella 4.1: illustrazione delle statistiche (congiuntamente) sufficienti per alcuni dei principali modelli statistici; qui N indica l'ampiezza del campione.

PROPOSIZIONE 4.9. *In un modello esponenziale canonico, nel quale la verosimiglianza $\mathbf{L}_\theta$ rispetto alla misura dominante μ sia come nell'*equazione *(4.21), la statistica S è completa.*

ESEMPIO 4.16. ►
Si riconsideri il campione $\mathbf{X}_{\mathsf{e}(\theta)}^n$, $\theta > 0$, dell'*Esempio (4.14)*. Grazie alla *Proposizione (4.9)* si ottiene che la statistica $S = \sum_{i=1}^n X_i$ è completa. In alternativa, se φ è una funzione sommabile, si ha

$$0 = \mathbf{E}_\theta(\varphi(S)) = \int_0^\infty \varphi(s)\, \frac{\theta^n}{\Gamma(n)}\, s^{n-1}\, e^{-\theta s}\, \mathrm{d}s$$

se, e solo se, $\varphi(s) = 0$ per ogni $s > 0$, come scende dalle proprietà della *Trasformata di Laplace*. ◄

L'importanza dell'uso combinato delle statistiche sufficienti e complete è evidenziato dal seguente Teorema.

TEOREMA 4.5 (Lehman–Scheffé)**.** *Se S è una statistica sufficiente e completa per $\mathfrak{M}$, e se esiste uno stimatore $T \in \mathfrak{D}_c^2$ corretto e di quadrato sommabile della funzione $\xi\colon \Theta \to \mathbb{R}^d$, allora*

$$U = \mathbf{E}(T \mid S) \tag{4.23}$$

è uno stimatore di ξ ottimale nella classe $\mathfrak{D}_c^2$.

DEFINIZIONE 4.14 (UMVU)**.** Uno stimatore $U^* \in \mathfrak{D}_c^2$ della funzione ξ tale che sia $\mathbf{R}_\theta(U^*) \leq \mathbf{R}_\theta(U)$ per ogni $\theta \in \Theta$ e per ogni altro stimatore $U \in \mathfrak{D}_c^2$ di ξ si dice *Uniformemente di Varianza Minima Corretto* (l'acronimo anglosassone è UMVU, cioè "Uniformly Minimum Variance Unbiased").

Ora, grazie ai risultati enunciati, è immediato utilizzare le statistiche mostrate in *Tabella (4.1)* per ricavare stimatori ottimali (cioè di rischio minimo) dei parametri di numerosi modelli statistici. In particolare si osservi come, nel caso di modelli appartenenti alla famiglia esponenziale, sia sufficiente ricondursi alla forma canonica per calcolare delle statistiche sufficienti e complete (si vedano anche gli *Esempi (4.13)–(4.14)* e gli *Esempi (4.15)–(4.16)*).

Capitolo 5

TECNICHE DI STIMA

Nell'ambito della Statistica Matematica sono state ideate numerose tecniche per costruire stimatori di parametri (o di loro funzioni). Nel seguito si introducono alcune di tali tecniche, di fondamentale interesse per il Corso e di grande importanza nelle applicazioni.

5.1 Il Metodo dei Momenti

Uno dei metodi di stima più semplici è il cosiddetto *Metodo dei Momenti* (MM). Il principio su cui si basa è intuitivo e consiste nell'osservazione che, in una gran varietà di casi e di problemi, i parametri da stimare (o funzioni di questi) sono esprimibili come funzioni note di un numero finito di *momenti* della legge in questione.

ESEMPIO 5.1. ►
Se $X \sim \mathtt{P}(\theta)$, $\theta > 0$, allora $\mathbf{E}_\theta(X) = \theta$; dunque, θ è funzione del momento del primo ordine di X, e uno stimatore della speranza di X (quale la media campionaria) è anche uno stimatore di θ. ◄

Generalizzando quanto già visto per la media e la varianza campionarie nella *Sezione 4.1.1*, è possibile definire i *momenti campionari*.

DEFINIZIONE 5.1 (Momenti e Momenti Campionari)**.** Dato un campione $\mathbf{X}_{\mu_\theta}^n$, $\theta \in \Theta$, si definiscono *momenti* della legge μ_θ (o *momenti della popolazione*) di ordine $r \in \mathbb{N}$ le quantità

$$m_r(\theta) = \mathbf{E}_\theta(X^r)\,. \tag{5.1}$$

Analogamente, si definiscono *momenti campionari* (o *empirici*) di ordine $r \in \mathbb{N}$ le quantità

$$\overline{m}_r = \frac{1}{n}\sum_{i=1}^n X_i^r. \tag{5.2}$$

Dato un campione $\mathbf{X}_{\mu_\theta}^n$, $\theta \in \Theta$, nel seguito supporremo che esistano e che siano finiti i momenti di μ_θ sino all'ordine $k \in \mathbb{N}$ incluso, cioè che sia, per ogni $\theta \in \Theta$,

$$\mathbf{E}_\theta(|X^r|) < \infty, \quad r = 1, \ldots, k. \tag{5.3}$$

NOTA 5.1.
I momenti campionari $\overline{m}_r$, a differenza dei momenti m_r, esistono per ogni ordine $r \in \mathbb{N}$; in particolare, $\overline{m}_1 = \overline{X}_n$.

Nella sua semplicità, il risultato che segue è di fondamentale importanza.

PROPOSIZIONE 5.1. *I momenti campionari sono stimatori corretti dei rispettivi momenti, nell'ipotesi che questi ultimi esistano.*

DIM. Evidentemente si ha, per ogni $\theta \in \Theta$,

$$\mathbf{E}_\theta(\overline{m}_r) = \frac{1}{n}\sum_{i=1}^n \mathbf{E}_\theta(X_i^r) = \frac{1}{n}\sum_{i=1}^n \mathbf{E}_\theta(X^r) = \mathbf{E}_\theta(X^r) = m_r(\theta),$$

da cui la tesi. □

Si può ricorrere al MM quando si voglia stimare la funzione $\xi\colon \Theta \to \mathbb{R}^d$ e sia possibile scrivere quest'ultima come funzione dei primi k momenti di μ_θ, che si suppone esistano finiti. Si noti, tuttavia, che proprio quest'ultima richiesta rappresenta una delle debolezze del MM.

ESEMPIO 5.2. ►
Se $X \sim \mathtt{e}(\theta)$, $\theta > 0$, e $\xi(\theta) = 1/\theta$, allora $\xi(\theta) = \mathbf{E}_\theta(X) = m_1(\theta)$ e $\overline{m}_1 = \overline{X}_n$ è uno stimatore di $\xi(\theta)$. ◄

DEFINIZIONE 5.2 (Metodo dei Momenti)**.** Sia $h\colon \mathbb{R}^k \to \mathbb{R}$ una funzione misurabile tale che

$$\xi(\theta) = h\left(m_1(\theta), \ldots, m_k(\theta)\right).$$

Il *Metodo dei Momenti* consiste nello stimare $\xi(\theta)$ mediante la statistica T_n data da

$$\begin{aligned} T_n\left(X_1, \ldots, X_n\right) &= h\left(\overline{m}_1(\theta), \ldots, \overline{m}_k(\theta)\right) \\ &= h\left(\frac{1}{n}\sum_{i=1}^n X_i, \ldots, \frac{1}{n}\sum_{i=1}^n X_i^k\right). \end{aligned} \tag{5.4}$$

Gli esempi che seguono illustrano pregi e difetti del MM.

ESEMPIO 5.3 (Legge $\mathtt{N}\left(\boldsymbol{\theta_1}, \boldsymbol{\theta_2^2}\right)$)**.** ►
Sia dato il campione $\mathbf{X}^n_{\mathtt{N}\left(\theta_1, \theta_2^2\right)}$; per stimare θ_1 e θ_2^2 con il MM si può procedere nel seguente modo. Poiché $\theta_1 = \mathbf{E}_\theta(X) = m_1$ e $\theta_2^2 = \mathbf{V}_\theta(X) = \mathbf{E}_\theta\left(X^2\right) - \mathbf{E}_\theta(X)^2 = m_2 - m_1^2$, usiamo gli stimatori

$$\begin{aligned} T_1 &= \overline{m}_1 = \overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i, \\ T_2 &= \overline{m}_2 - \overline{m}_1^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \overline{X}_n^2 = \frac{1}{n}\sum_{i=1}^n \left(X_i - \overline{X}_n\right)^2 = \overline{\sigma}_n^2. \end{aligned}$$

Evidentemente T_1 è corretto, essendo la media campionaria, mentre $T_2 = \overline{\sigma}_n^2$ è distorto, essendo la varianza campionaria non corretta. ◄

OSSERVAZIONE 5.1.
Il risultato dell'*Esempio (5.3)* può essere generalizzato. Sia infatti $\mathbf{X}^n_{\mu_\theta}$, $\theta \in \Theta$, un campione di legge μ_θ avente varianza θ_2^2. Il MM propone sempre come stimatore di θ_2^2 la statistica

$$T = \overline{m}_2 - \overline{m}_1^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2 = \frac{1}{n}\sum_{i=1}^n \left(X_i - \overline{X}_n\right)^2 = \overline{\sigma}_n^2$$

che, per quanto visto nella *Sezione 4.1.1*, è uno stimatore *distorto* della varianza. Si ricordi comunque il risultato dell'*Esempio (4.7)* circa il rischio dello stimatore $\overline{\sigma}_n^2$.

ESEMPIO 5.4 (Legge $\mathtt{P}(\boldsymbol{\theta})$)**.** ►
Sia dato il campione $\mathbf{X}^n_{\mathtt{P}(\theta)}$, $\theta > 0$; per stimare θ con il MM si può procedere nel seguente modo. Come è noto, $m_1 = \theta$ e $m_2 = \theta + \theta^2$. Volendo stimare θ con il momento del primo ordine usiamo $\overline{m}_1 = \overline{X}_n$. Se invece si stima il parametro con il momento del secondo ordine occorre risolvere rispetto a θ l'equazione $\theta^2 + \theta - m_2 = 0$. L'unica soluzione positiva (la sola che interessi, essendo $\theta > 0$) è $\tilde{\theta} = -\frac{1}{2} + \frac{1}{2}\sqrt{1 + 4m_2}$ e dunque lo stimatore del MM è

$$T = -\frac{1}{2} + \frac{1}{2}\sqrt{1 + 4\overline{m}_2},$$

che non è particolarmente maneggevole. ◄

NOTA 5.2.
Il risultato dell'*Esempio (5.4)* suggerisce di usare, quando possibile, i momenti di ordine più basso. Questa strategia è "prudente", nel senso che non è sempre detto che esistano i momenti di ordine elevato e, in tal caso, $\overline{m}_r$, per $r \gg 1$, potrebbe risultare essere (come già osservato nella *Nota (5.1)*) lo stimatore di una quantità inesistente!

ESEMPIO 5.5 (Legge $\mathtt{U_c}(\mathbf{0}, \boldsymbol{\theta})$). ►
Sia dato il campione $\mathbf{X}^n_{\mathtt{U_c}(0,\theta)}$, $\theta > 0$; per stimare θ con il MM si può procedere nel seguente modo: essendo $m_1 = \frac{\theta}{2}$, useremo $T = 2\overline{m}_1 = 2\overline{X}_n$. ◄

OSSERVAZIONE 5.2.
Si noti che lo stimatore T calcolato nell'*Esempio (5.5)* assume valori nell'intervallo $I_T = (0, 2\theta)$ anziché nell'intervallo $(0, \theta)$. Usando i risultati dell'*Esempio (4.6)* e dell'*Esempio (4.12)* si vede facilmente che la statistica $T_n = \frac{n+1}{n} X_{(n)}$ è uno stimatore corretto di θ che assume valori nell'intervallo $I_{T_n} = (0, \frac{n+1}{n}\theta)$: dunque, per $n > 1$ (cioè per campioni non banali) si ha $I_{T_n} \subset I_T$. Il rischio di T è pari a $\mathbf{R}_\theta(T) = \mathbf{V}_\theta(T) = \frac{1}{3n}\theta^2$, mentre il rischio di T_n è pari a $\mathbf{R}_\theta(T_n) = \mathbf{V}_\theta(T_n) = \frac{1}{n(n+2)}\theta^2$: dunque, T_n è preferibile a T.

ESEMPIO 5.6 (Legge $\mathtt{e}(\boldsymbol{\theta})$). ►
Sia dato il campione $\mathbf{X}^n_{\mathtt{e}(\theta)}$, $\theta > 0$; per stimare θ con il MM si può procedere nel seguente modo: essendo $m_1 = \frac{1}{\theta}$, useremo $T = \frac{1}{\overline{X}_n}$. ◄

ESEMPIO 5.7 (Legge $\mathtt{Bi}(\boldsymbol{n}, \boldsymbol{\theta})$). ►
Sia dato il campione $\mathbf{X}^N_{\mathtt{Bi}(n,\theta)}$, con $n \in \mathbb{N}$ e $\theta \in (0, 1)$; per stimare n e θ con il MM si può procedere nel seguente modo. Poiché

$$\mathbf{E}_\theta(X) = m_1 = n\theta,$$
$$\mathbf{E}_\theta\left(X^2\right) = m_2 = n\theta(1-\theta) + (n\theta)^2,$$

è necessario risolvere le seguenti equazioni in n e θ

$$\overline{m}_1 = \frac{1}{N}\sum_{i=1}^{N} X_i = n\theta,$$
$$\overline{m}_2 = \frac{1}{N}\sum_{i=1}^{N} X_i^2 = n\theta(1-\theta) + (n\theta)^2.$$

Usiamo quindi gli stimatori

$$T_1\left(X_1, \ldots, X_N\right) = \frac{\overline{X}_N}{T_2\left(X_1, \ldots, X_N\right)},$$
$$T_2\left(X_1, \ldots, X_N\right) = \frac{\overline{X}_N^2}{\overline{X}_N + \overline{X}_N^2 - \sum_{i=1}^{N} X_i^2/N},$$

dove T_1 è lo stimatore di θ e T_2 è lo stimatore di n. ◄

5.2 Stimatori di Massima Verosimiglianza

Il metodo della *Massima Verosimiglianza* è molto usato nell'ambito della Statistica Matematica, sia perché intuitivo sia perché di facile applicazione; gli stimatori costruiti usando questa tecnica sono detti *stimatori di Massima Verosimiglianza* (SMV). Il principio su cui si basa è semplice e consiste nell'individuare il valore del parametro (o dei parametri) che massimizzi la funzione di verosimiglianza del campione. In altre parole, si cerca il valore del parametro (o dei parametri) che renda "più probabile" il fatto di avere osservato proprio il campione in esame (cioè la *moda* in funzione del parametro incognito).

DEFINIZIONE 5.3 (Stimatori di Massima Verosimiglianza)**.** Sia assegnato un modello statistico dominato $\mathfrak{M}$ e sia $\mathbf{L}\colon \Theta \times \Omega \to \mathbb{R}_+$ la funzione di verosimiglianza, dove Θ è un aperto di $\mathbb{R}^d$. Uno stimatore $T\colon \Omega \to \mathbb{R}^d$, $d \geq 1$, del parametro $\theta \in \Theta$ si dice di *Massima Verosimiglianza* se si ha, per ogni $\omega \in \Omega$,

$$\mathbf{L}(T(\omega), \omega) = \sup_{\theta \in \Theta} \mathbf{L}(\theta, \omega). \tag{5.5}$$

Lo SMV, se esiste ed è unico, è di solito indicato con $\hat{\theta}$ o $\hat{\theta}_n$, dove n è l'ampiezza del campione.

OSSERVAZIONE 5.3 (Equazione di Massima Verosimiglianza)**.**
Se la funzione di verosimiglianza $\theta \mapsto \mathbf{L}(\theta, \omega)$ risulta essere differenziabile rispetto a θ per ogni $\omega \in \Omega$, allora $\hat{\theta}$ è soluzione dell'*Equazione di Massima Verosimiglianza*

$$\left[\frac{\partial}{\partial\theta}\mathbf{L}(\theta,\omega)\right]_{\theta=\hat{\theta}} = 0. \tag{5.6}$$

Se si considera un campione $\mathbf{X}_{\mu_\theta}^n$, con densità f_θ rispetto a μ_θ, la verosimiglianza del campione è la funzione $\mathbf{L}\colon \Theta \times \mathbb{R}^n \to \mathbb{R}_+$ data da

$$\mathbf{L}_\theta(\boldsymbol{x}) = \prod_{i=1}^n f_\theta(x_i),$$

dove $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Perciò l'*equazione (5.6)* diviene

$$\left[\frac{\partial}{\partial\theta}\prod_{i=1}^n f_\theta(x_i)\right]_{\theta=\hat{\theta}} = 0. \tag{5.7}$$

Se poi risulta $f_\theta(\boldsymbol{x}) > 0$ per ogni punto di un insieme che *non dipende* da θ (modello *regolare*), può essere più facile calcolare il massimo di $\ln \mathbf{L}_\theta$ anziché di $\mathbf{L}_\theta$ (i punti di massimo non cambiano, essendo la funzione logaritmo strettamente crescente). Si può allora scrivere l'*Equazione di Massima Verosimiglianza* nella forma

$$\left[\frac{\partial}{\partial\theta}\ln\prod_{i=1}^n f_\theta(x_i)\right]_{\theta=\hat{\theta}} = \left[\sum_{i=1}^n \frac{\partial}{\partial\theta}\ln f_\theta(x_i)\right]_{\theta=\hat{\theta}} = 0, \tag{5.8}$$

che spesso risulta più maneggevole dell'*equazione (5.7)*; si noti tuttavia che può essere comunque necessario ricorrere a metodi numerici per risolvere tale equazione.

Il Teorema che segue stabilisce un legame tra SMV e *stimatori efficienti* (si veda la *Definizione (4.10)*).

TEOREMA 5.1 (Efficienza degli SMV)**.** *Se Θ è un aperto di $\mathbb{R}$ e se per il campione $(X_1, \ldots, X_n)$ esiste uno stimatore efficiente T di θ, allora T è anche uno* SMV*, cioè $\hat{\theta} = T$.*

DIM. Nel *Teorema (4.3)* si è visto che se T è uno stimatore efficiente di θ allora vale, per $\boldsymbol{x} \in \mathbb{R}^n$,

$$\frac{\partial}{\partial\theta}\ln\mathbf{L}_\theta(\boldsymbol{x}) = h(\theta)\,(t(\boldsymbol{x}) - \theta)\,.$$

Ponendo uguale a zero quest'ultima espressione, l'*Equazione di Massima Verosimiglianza* fornisce $\hat{\theta} = T$. Tale soluzione corrisponde effettivamente ad un massimo di $\mathbf{L}_\theta$: infatti la derivata seconda è

$$\frac{\partial^2}{\partial\theta^2}\ln\mathbf{L}_\theta(\boldsymbol{x}) = h'(\theta)\,(t(\boldsymbol{x}) - \theta) - h(\theta),$$

che calcolata per $\theta = \hat{\theta}$ vale $-h(\hat{\theta})$. Poiché T è efficiente (e quindi corretto) si ha, per ogni $\theta \in \Theta$,

$$\frac{1}{\mathbf{V}_\theta(T)} = \mathcal{I}_{\mathbf{X}}(\theta) = -\mathbf{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\ln\mathbf{L}_\theta\right) = -h'(\theta)\mathbf{E}_\theta\Big(\hat{\theta} - \theta\Big) + h(\theta) = h(\theta)$$

onde, per ogni $\theta \in \Theta$, $h(\theta) = \frac{1}{\mathbf{V}_\theta(T)} > 0$. □

NOTA 5.3.
Il *Teorema (5.1)* non asserisce che uno SMV sia necessariamente efficiente — si veda, ad esempio, [24].

Nel seguito si illustra con alcuni esempi il metodo della *Massima Verosimiglianza*.

ESEMPIO 5.8. ►

Si consideri il campione $\mathbf{X}^n_{\mathrm{P}(\theta)}$, $\theta > 0$, e si voglia trovare uno SMV di θ. La densità, rispetto alla "misura del contare" μ_c, è $f_\theta(x) = e^{-\theta}\frac{\theta^x}{x!}\mathbb{I}_{\mathbb{N}}(x)$ e dunque

$$\frac{\partial}{\partial\theta}\ln\prod_{i=1}^n f_\theta(x_i) = -n + \frac{\sum_{i=1}^n x_i}{\theta}.$$

Risolvendo l'*Equazione di Massima Verosimiglianza* si ottiene $\hat\theta = \frac{1}{n}\sum_{i=1}^n x_i = \overline{X}_n$. ◄

ESEMPIO 5.9. ►

Si consideri il campione $\mathbf{X}^n_{\mathrm{e}(\theta)}$, $\theta > 0$, e si voglia trovare uno SMV di θ. La densità, rispetto alla misura di Lebesgue λ, è $f_\theta(x) = \theta e^{-\theta x}\mathbb{I}_{\mathbb{R}_+}(x)$ e dunque

$$\frac{\partial}{\partial\theta}\ln\prod_{i=1}^n f_\theta(x_i) = \frac{n}{\theta} - \sum_{i=1}^n x_i.$$

Risolvendo l'*Equazione di Massima Verosimiglianza* si ottiene $\hat\theta = \frac{n}{\sum_{i=1}^n x_i} = \frac{1}{\overline{X}_n}$. ◄

ESEMPIO 5.10. ►

Si consideri il campione $\mathbf{X}^n_{\mathrm{N}\left(\theta_1,\theta_2^2\right)}$ e si voglia trovare uno SMV di θ_1 e di θ_2^2. La densità, rispetto alla misura di Lebesgue λ, è $f_\theta(x) = \frac{1}{\sqrt{2\pi}\theta_2}\, e^{-\frac{(x-\theta_1)^2}{2\theta_2^2}}$, $x \in \mathbb{R}$, e dunque

$$\ln f_\theta(x) = -\frac{(x-\theta_1)^2}{2\theta_2^2} - \ln\theta_2 - \frac{1}{2}\ln(2\pi).$$

Le *Equazioni di Massima Verosimiglianza* sono

$$\begin{cases} \displaystyle\sum_{i=1}^n \frac{\partial}{\partial\theta_1}\ln f_\theta(x_i) = \sum_{i=1}^n \frac{(x_i-\theta_1)}{\theta_2^2} = 0, \\ \displaystyle\sum_{i=1}^n \frac{\partial}{\partial\theta_2}\ln f_\theta(x_i) = \sum_{i=1}^n \frac{(x_i-\theta_1)^2}{\theta_2^3} - \frac{n}{\theta_2} = 0. \end{cases}$$

La prima equazione fornisce $\hat\theta_1 = \overline{X}_n$; la seconda equazione, tenuto conto della soluzione della prima, fornisce $\hat\theta_2^2 = \frac{1}{n}\sum_{i=1}^n \left(X_i - \overline{X}_n\right)^2$. Dunque uno SMV dei due parametri è $\hat\theta = \left(\overline{X}_n, \overline{\sigma}_n^2\right)$. ◄

ESEMPIO 5.11. ►

Si consideri il campione $\mathbf{X}^n_\mu$, dove μ rappresenta la legge GEV o la legge GP (si veda il *Capitolo 3* nelle parti riguardanti lo studio dei massimi), e si voglia trovare uno SMV dei rispettivi parametri $\boldsymbol{\theta} = \{a, b, \gamma\}$ o $\boldsymbol{\theta} = \{b, \gamma\}$.

Nel caso GEV la densità, rispetto alla misura di Lebesgue λ, è (si veda l'*equazione (3.11)*)

$$f_{\boldsymbol{\theta}}(x) = \exp\left\{-\left[1+\gamma\,\frac{x-a}{b}\right]^{-1/\gamma}\right\}\,\frac{1}{b}\,\left[1+\gamma\,\frac{x-a}{b}\right]^{-1/\gamma-1}$$

sull'insieme $\left\{x\colon 1+\gamma\,\frac{x-a}{b} > 0\right\}$, e dunque

$$\ln f_{\boldsymbol{\theta}}(x) = -\left(1+\gamma\,\frac{x-a}{b}\right)^{-1/\gamma} - \ln b - \left(1+\frac{1}{\gamma}\right)\,\ln\left(1+\gamma\,\frac{x-a}{b}\right).$$

Purtroppo le *Equazioni di Massima Verosimiglianza* corrispondenti risultano essere alquanto complicate e, in generale, non sono risolvibili analiticamente (tranne nel caso speciale $\gamma = 0$); perciò, nella pratica, è necessario adottare procedure numeriche.

Nel caso GP la densità, rispetto alla misura di Lebesgue λ, è (si veda l'*equazione (3.28)*)

$$f_{\boldsymbol{\theta}}(x) = \frac{1}{b}\,\left[1+\gamma\,\frac{x}{b}\right]^{-1/\gamma-1}$$

sull'insieme $\{x\colon x > 0 \text{ e } 1 + \gamma x/b > 0\}$, e dunque

$$\ln f_{\boldsymbol{\theta}}(x) = -\ln b - \left(1 + \frac{1}{\gamma}\right) \ln\left(1 + \gamma\,\frac{x}{b}\right).$$

Purtroppo anche in questo caso le *Equazioni di Massima Verosimiglianza* corrispondenti risultano essere alquanto complicate e, in generale, non sono risolvibili analiticamente (tranne nel caso speciale $\gamma = 0$); perciò, nella pratica, è necessario adottare procedure numeriche. ◀

Il risultato che segue sancisce il legame stretto esistente tra gli SMV e le statistiche sufficienti; in generale ciò non vale per altri stimatori (quali, ad esempio, quelli calcolati con il MM).

PROPOSIZIONE 5.2. *Se S è una statistica sufficiente per la famiglia $\{F_\theta\}_{\theta\in\Theta}$, e inoltre esiste unico uno* SMV *T di θ, allora T è funzione di S.*

DIM. Poiché S è una statistica sufficiente, grazie al Criterio di fattorizzazione la verosimiglianza $\mathbf{L}_\theta$ del campione vale $\mathbf{L}_\theta(\boldsymbol{x}) = h(\boldsymbol{x})\, g_\theta(S(\boldsymbol{x}))$. Dunque, massimizzare $\mathbf{L}_\theta$ rispetto a θ equivale a massimizzare g_θ, che è una funzione di S solamente. □

NOTA 5.4.

La *Proposizione (5.2)* non asserisce che lo SMV T sia esso stesso una statistica sufficiente (sebbene, nella pratica, spesso sia questo il caso) — si veda, ad esempio, [24].

Dunque, gli SMV possono essere candidati ideali, almeno dal punto di vista della riduzione del Rischio. Il Teorema che segue illustra un'ulteriore proprietà importante degli SMV.

TEOREMA 5.2 (Invarianza degli SMV)**.** *Sia $\theta \in \Theta \subseteq \mathbb{R}^d$, $d \geq 1$, e sia $\xi\colon \Theta \to \Delta$ una funzione misurabile, dove Δ è un intervallo in $\mathbb{R}^k$, $1 \leq k \leq d$. Se $\hat\theta$ è uno* SMV *di θ, allora $\xi(\hat\theta)$ è uno* SMV *di $\xi(\theta)$.*

ESEMPIO 5.12. ▶

Si consideri il campione $\mathbf{X}^n_{\mathtt{Be}(\theta)}$, $\theta \in [0,1]$, e si voglia trovare uno SMV della funzione $\xi(\theta) = \theta(1-\theta)$, ovvero della varianza di X. Si noti che $\Delta = [0, 1/4]$ e che ξ non è invertibile. Uno SMV $\hat\theta$ di θ si calcola facilmente. La densità, rispetto alla "misura del contare" μ_c, è $f_\theta(x) = \theta^x(1-\theta)^{1-x}\mathbb{I}_{\{0,1\}}(x)$ e dunque

$$\frac{\partial}{\partial\theta} \ln \prod_{i=1}^n f_\theta(x_i) = \frac{\sum_{i=1}^n x_i}{\theta} - \frac{n - \sum_{i=1}^n x_i}{1-\theta}.$$

Risolvendo l'*Equazione di Massima Verosimiglianza* si ottiene $\hat\theta = \frac{1}{n}\sum_{i=1}^n x_i = \overline{X}_n$. Ne scende che uno SMV $\hat\xi$ della funzione $\xi(\theta)$ è $\xi(\hat\theta) = \overline{X}_n(1-\overline{X}_n)$. ◀

Gli Esempi che seguono mostrano che non è sempre detto che uno SMV esista, può non essere effettivamente un massimo e può non essere univocamente determinato.

ESEMPIO 5.13 (Uno SMV può non esistere)**.** ▶

Si consideri il campione $\mathbf{X}^n_{\mathtt{Be}(\theta)}$, $\theta \in (0,1)$, e si voglia trovare uno SMV di θ. Per quanto visto nell'*Esempio (5.12)* uno SMV di θ è $\hat\theta = \overline{X}_n$. Se però si dovesse osservare il campione $(x_1 = 0, \ldots, x_n = 0)$ o $(x_1 = 1, \ldots, x_n = 1)$ si avrebbe, rispettivamente, $\overline{x}_n = 0$ o $\overline{x}_n = 1$, entrambi valori non ammissibili per θ. Dunque uno SMV può non esistere. ◀

ESEMPIO 5.14 (Uno SMV può non essere un massimo)**.** ▶

Si consideri il campione $\mathbf{X}^n_{\mu_\theta}$, dove μ_θ è la legge Uniforme su $[0,\theta)$, $\theta > 0$, e si voglia trovare uno SMV di θ. La densità, rispetto alla misura di Lebesgue λ, è $f_\theta(x) = \frac{1}{\theta}\,\mathbb{I}_{[0,\theta)}(x)$ e dunque la verosimiglianza del campione è

$$\mathbf{L}_\theta(\boldsymbol{x}) = \frac{1}{\theta^n}\,\mathbb{I}_{[0,\theta)}\left(x_{(n)}\right),$$

dove $\boldsymbol{x} \in \mathbb{R}^n$. Per quanto visto nell'*Osservazione (5.3)*, poiché il supporto di f_θ dipende da θ non è possibile fare ricorso all'*equazione (5.8)*; d'altro canto, l'uso delle derivate come mostrato nell'*equazione (5.7)* è possibile solo nel senso (generalizzato) delle Distribuzioni. Tuttavia, essendo evidentemente

$$\sup_{\theta > x_{(n)}} \mathbf{L}_\theta(\boldsymbol{x}) = \frac{1}{x^n_{(n)}},$$

si ha che uno SMV di θ è $\hat\theta = X_{(n)}$, che però genera un estremo superiore e non un massimo. ◀

ESEMPIO 5.15 (Uno SMV può non essere unico)**.** ►
Si consideri il campione $\mathbf{X}^n_{\mathtt{U}_c\left(\theta-\frac{1}{2},\theta+\frac{1}{2}\right)}$, $\theta \in \mathbb{R}$, e si voglia trovare uno SMV di θ. La densità, rispetto alla misura di Lebesgue λ, è $f_\theta(x) = \mathbb{I}_{(\theta-\frac{1}{2},\theta+\frac{1}{2})}(x)$ e dunque la verosimiglianza del campione è

$$\mathbf{L}_\theta(\boldsymbol{x}) = \mathbb{I}_{\left(\theta-\frac{1}{2},\theta+\frac{1}{2}\right)^n}(\boldsymbol{x}),$$

dove $\boldsymbol{x} \in \mathbb{R}^n$. Come nell'*Esempio (5.14)*, anche in questo caso non è possibile fare ricorso alle tecniche di derivazione dell'*Osservazione (5.3)*. Tuttavia, evidentemente $\mathbf{L}_\theta$ assume solamente i valori 0 e 1; in particolare, quest'ultimo si ottiene per tutti i valori di θ tali che

$$\theta - \frac{1}{2} < x_{(1)} \leq x_{(n)} < \theta + \frac{1}{2}.$$

Dunque $\mathbf{L}_\theta$ è massima ogni qual volta siano soddisfatte le disuguaglianze

$$x_{(n)} - \frac{1}{2} < \theta < x_{(1)} + \frac{1}{2},$$

sicché è uno SMV di θ ogni v.a. $\hat{\theta}$ tale che

$$X_{(n)} - \frac{1}{2} < \hat{\theta} < X_{(1)} + \frac{1}{2},$$

e dunque $\hat{\theta}$ può non essere unico. Ad esempio, per $0 < c < 1$, la statistica

$$T_c(X_1, \ldots, X_n) = X_{(n)} - \frac{1}{2} + c\left(1 + X_{(1)} - X_{(n)}\right)$$

assume sempre valori nell'intervallo $\left(X_{(n)} - \frac{1}{2}, X_{(1)} + \frac{1}{2}\right)$ e dunque è uno SMV di θ. In particolare, tradizionalmente si assume $c = \frac{1}{2}$ e si usa

$$T_{\frac{1}{2}} = \frac{X_{(n)} + X_{(1)}}{2}.$$

◄

Chiudiamo questa Sezione prendendo in esame alcune proprietà ottimali degli SMV; per le condizioni di regolarità usate nei risultati che seguono si consultino, ad esempio, [24, 6, 28]. Si noti che il metodo della *Massima Verosimiglianza* spesso ammette solo giustificazioni asintotiche, nel senso che gli stimatori calcolati con tale tecnica risultano validi (o competitivi rispetto ad altri) solo per campioni di ampiezza considerevole.

DEFINIZIONE 5.4 (Efficienza asintotica)**.** Una successione di stimatori $\{T_n\}$ è *asintoticamente efficiente* per la funzione $\xi(\theta)$ se

$$\lim_{n\to\infty} \frac{\mathbf{V}_\theta(T_n)}{\left(\frac{\partial}{\partial\theta}\xi(\theta)\right)^2 / \mathcal{I}_{\boldsymbol{X}}(\theta)} = 1, \tag{5.9}$$

ovvero la varianza di T_n tende al limite inferiore della disuguaglianza FCR.

TEOREMA 5.3 (Consistenza ed efficienza asintotica degli SMV)**.** *Si consideri il campione $\mathbf{X}^n_{\mu_\theta}$ e sia $\hat{\theta}_n$ uno* SMV *di θ. Se la funzione $\theta \mapsto \xi(\theta)$ è continua, sotto opportune condizioni di regolarità sulla funzione di verosimiglianza $\mathbf{L}_\theta$ si ha che, per ogni $\theta \in \Theta$ e $\epsilon > 0$,*

$$\lim_{n\to\infty} \mathbf{P}_\theta\left\{\left|\xi(\hat{\theta}_n) - \xi(\theta)\right| \geq \epsilon\right\} = 0, \tag{5.10}$$

ovvero $\xi(\hat{\theta})$ è uno stimatore consistente di $\xi(\theta)$. Inoltre, $\left\{\xi(\hat{\theta}_n)\right\}$ è una successione di stimatori asintoticamente efficiente per $\xi(\theta)$.

TEOREMA 5.4 (Normalità asintotica degli SMV)**.** *Si consideri il campione $\mathbf{X}^n_{\mu_\theta}$ e sia $\hat{\theta}_n$ uno* SMV *di θ. Sotto opportune condizioni di regolarità sulla funzione di verosimiglianza $\mathbf{L}_\theta$ si ha che, per $n \gg 1$,*

$$\hat{\theta}_n \sim \mathtt{N}\left(\theta, \sigma^2\right), \tag{5.11}$$

dove $\sigma^2 = 1/\mathcal{I}_{\boldsymbol{X}}(\theta)$.

I risultati sopra esposti chiariscono alcune delle ragioni della scelta degli SMV quali stimatori dei parametri o di loro funzioni. Non si deve però dimenticare che tali proprietà sono asintotiche, ovvero valide per campioni di ampiezza considerevole, e dunque nelle applicazioni (dove spesso la numerosità delle osservazioni è assai ridotta) non sempre è lecito (o possibile) fare ricorso agli SMV.

Capitolo 6

CAMPIONI GAUSSIANI

Il caso Gaussiano riveste grande importanza in numerose questioni della Statistica Matematica, senza dubbio in conseguenza del ruolo che il *Teorema del Limite Centrale* assegna alla legge Normale (o Gaussiana). In particolare i campioni (vettori aleatori) Gaussiani giocano un ruolo privilegiato.

Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbf{P})$ si consideri il campione $\mathbf{X}^n_{\mathtt{N}(0,\sigma^2)}$. La *funzione caratteristica*[1] *(f.c.)* $\boldsymbol{\varphi_X}$ del vettore aleatorio $\boldsymbol{X} = (X_1, \ldots, X_n)$ è data da, per $\boldsymbol{t} \in \mathbb{R}^n$,

$$\begin{aligned} \boldsymbol{\varphi_X}(\boldsymbol{t}) &= \mathbf{E}\Big(e^{\mathrm{i}\langle \boldsymbol{t}, \boldsymbol{X}\rangle}\Big) \\ &= \mathbf{E}\left(\prod_{k=1}^{n} e^{\mathrm{i}t_k X_k}\right) \\ &= \prod_{k=1}^{n} e^{-\frac{1}{2}\sigma^2 t_k^2} \\ &= e^{-\frac{1}{2}\sigma^2 \|\boldsymbol{t}\|^2}, \end{aligned} \tag{6.1}$$

dove $\|\boldsymbol{t}\| = \sqrt{\langle \boldsymbol{t}, \boldsymbol{t}\rangle} = \sqrt{t_1^2 + \cdots + t_n^2}$ denota il *modulo* di $\boldsymbol{t}$. Il risultato che segue è fondamentale.

PROPOSIZIONE 6.1 (Invarianza)**.** *Si consideri il campione* $\mathbf{X}^n_{\mathtt{N}(0,\sigma^2)}$ *e sia* $\boldsymbol{A}$ *una matrice ortogonale* $n \times n$*. Allora anche il vettore aleatorio* $\boldsymbol{Y} = \boldsymbol{AX}$ *è un campione Gaussiano di legge* $\mathtt{N}\big(0, \sigma^2\big)$.

DIM. Useremo il fatto che $\boldsymbol{A}^{-1} = \boldsymbol{A}^\intercal$. Calcolando la legge di $\boldsymbol{Y}$ si ottiene, per $\boldsymbol{t} \in \mathbb{R}^n$,

$$\begin{aligned} \boldsymbol{\varphi_Y}(\boldsymbol{t}) &= \mathbf{E}\Big(e^{\mathrm{i}\langle \boldsymbol{t}, \boldsymbol{Y}\rangle}\Big) \\ &= \mathbf{E}\Big(e^{\mathrm{i}\langle \boldsymbol{t}, \boldsymbol{AX}\rangle}\Big) \\ &= \mathbf{E}\Big(e^{\mathrm{i}\langle \boldsymbol{A}^\intercal\boldsymbol{t}, \boldsymbol{X}\rangle}\Big), \end{aligned}$$

che rappresenta il valore di $\boldsymbol{\varphi_X}$ nel punto $\boldsymbol{t}' = \boldsymbol{A}^\intercal\boldsymbol{t} \in \mathbb{R}^n$, cioè $\boldsymbol{\varphi_X}\left(\boldsymbol{t}'\right)$. Usando l'*equazione (6.1)* si ha

$$\boldsymbol{\varphi_Y}(\boldsymbol{t}) = \boldsymbol{\varphi_X}\left(\boldsymbol{A}^\intercal\boldsymbol{t}\right) = e^{-\frac{1}{2}\sigma^2\|\boldsymbol{A}^\intercal\boldsymbol{t}\|^2} = e^{-\frac{1}{2}\sigma^2\|\boldsymbol{t}\|^2},$$

essendo

$$\|\boldsymbol{A}^\intercal\boldsymbol{t}\|^2 = \langle \boldsymbol{A}^\intercal\boldsymbol{t}, \boldsymbol{A}^\intercal\boldsymbol{t}\rangle = \langle \boldsymbol{AA}^\intercal\boldsymbol{t}, \boldsymbol{t}\rangle = \left\langle \boldsymbol{AA}^{-1}\boldsymbol{t}, \boldsymbol{t}\right\rangle = \langle \boldsymbol{t}, \boldsymbol{t}\rangle = \|\boldsymbol{t}\|^2,$$

da cui la tesi. □

[1] La *funzione caratteristica* $\boldsymbol{\varphi}_X$ di una v.a. X è data da, per $t \in \mathbb{R}$,

$$\boldsymbol{\varphi}_X(t) = \mathbf{E}\Big(e^{\mathrm{i}tX}\Big) = \int_{\mathbb{R}} e^{\mathrm{i}tx}\ \mathrm{d}F_X(x),$$

dove "i" è l'*unità immaginaria*. In termini rigorosi, $\boldsymbol{\varphi}_X$ è la *Trasformata di Fourier-Stieltjes* della legge di X: essa esiste per qualsiasi v.a. e ne individua univocamente la legge.

NOTA 6.1 (Invarianza della legge Gaussiana)**.**
La *Proposizione (6.1)* mostra che legge Gaussiana è invariante rispetto a trasformazioni ortogonali.

Il Teorema che segue è di grande importanza.

TEOREMA 6.1 (Cochran)**.** *Si consideri la decomposizione $\mathbb{R}^n = E_1 \oplus \cdots \oplus E_k$ nella somma diretta di $1 \leq k \leq n$ sottospazi ortogonali, ciascuno di dimensione r_i, $i = 1, \ldots, k$, con $r_1 + \cdots + r_k = n$. Sia $\mathbf{X}^n_{\mathtt{N}(0,\sigma^2)}$ un campione e siano π_i, $i = 1, \ldots, k$, le proiezioni ortogonali di $\mathbb{R}^n$ su E_i. Allora i k vettori aleatori $\pi_i(\boldsymbol{X})$ sono indipendenti.*

DIM. Basta supporre che sia $k = 2$; la dimostrazione nel caso generale si ottiene per iterazione. Sia allora $\mathbb{R}^n = E_1 \oplus E_2$. Supponiamo dapprima che E_1 sia il sottospazio generato dai primi r vettori della base canonica $(c_1, \ldots, c_n)$, mentre E_2 sia generato dai rimanenti $s = n - r$. Di conseguenza si ha

$$\pi_1(\boldsymbol{X}) = (\underbrace{X_1, \ldots, X_r}_{r \text{ elementi}}, \underbrace{0, \ldots, 0}_{s \text{ elementi}}),$$
$$\pi_2(\boldsymbol{X}) = (\underbrace{0, \ldots, 0}_{r \text{ elementi}}, \underbrace{X_{r+1}, \ldots, X_n}_{s \text{ elementi}}).$$

Dunque $\pi_1(\boldsymbol{X})$ e $\pi_2(\boldsymbol{X})$ sono indipendenti, essendo $\boldsymbol{X}$ un campione. In generale siano $(e_1, \ldots, e_r)$ una base ortonormale per E_1 e $(e_{r+1}, \ldots, e_n)$ una base ortonormale per E_2. Essendo $(e_1, \ldots, e_n)$ una base ortonormale per $\mathbb{R}^n$ esiste una trasformazione ortogonale che porta $(e_1, \ldots, e_n)$ in $(c_1, \ldots, c_n)$; dunque è sufficiente usare la *Proposizione (6.1)*, che assicura l'invarianza della legge Gaussiana rispetto a trasformazioni ortogonali, per ricondursi al caso precedente. □

OSSERVAZIONE 6.1.
Dalla *Proposizione (6.1)* e dal *Teorema (6.1)* scende immediatamente che le componenti $(Y_1, \ldots, Y_n)$ di $\boldsymbol{Y}$ sono *indipendenti* e tutte di legge $\mathtt{N}\big(0, \sigma^2\big)$.

Il risultato che segue è un importante corollario del *Teorema di Cochran*.

PROPOSIZIONE 6.2. *Sia $\mathbf{X}^n_{\mathtt{N}(\mu,\sigma^2)}$ un campione; allora la media campionaria $\overline{X}_n$ e la varianza campionaria corretta $\overline{S}^2_n$ sono indipendenti.*

DIM. Sia E_1 il sottospazio unidimensionale generato dal vettore unitario $\left(\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}\right)$ e sia $E_2 = E_1^{\perp}$ il sottospazio complementare (ortogonale a E_1) di dimensione $n - 1$. Si ponga

$$Y_i = \frac{X_i - \mu}{\sigma}$$

per $i = 1, \ldots, n$, di modo che $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ sia un campione $\mathbf{Y}^n_{\mathtt{N}(0,1)}$. Dunque

$$\pi_1(\boldsymbol{Y}) = \sum_{i=1}^n \frac{Y_i}{\sqrt{n}} = \frac{\sqrt{n}}{n} \sum_{i=1}^n Y_i = \sqrt{n}\, \overline{Y}_n = \sqrt{n} \left(\frac{\overline{X}_n - \mu}{\sigma} \right).$$

Inoltre

$$\begin{aligned}
\|\pi_2(\boldsymbol{Y})\|^2 &= \|\boldsymbol{Y}\|^2 - \|\pi_1(\boldsymbol{Y})\|^2 \quad \bigstar \textit{ Teorema di Pitagora} \text{ generalizzato} \\
&= \sum_{i=1}^n Y_i^2 - n\, \overline{Y}_n^2 \\
&= \sum_{i=1}^n Y_i^2 + n\, \overline{Y}_n^2 - 2\, \overline{Y}_n \sum_{i=1}^n Y_i \\
&= \sum_{i=1}^n \big(Y_i - \overline{Y}_n\big)^2 \\
&= \frac{1}{\sigma^2} \sum_{i=1}^n \big(X_i - \overline{X}_n\big)^2 \\
&= \frac{n-1}{\sigma^2}\, \overline{S}^2_n.
\end{aligned}$$

Il *Teorema (6.1)* assicura che $\pi_1(\boldsymbol{Y})$ e $\pi_2(\boldsymbol{Y})$ sono indipendenti; dunque lo sono anche $\overline{X}_n$ e $\overline{S}_n^2$ (in quanto funzioni misurabili delle proiezioni), da cui la tesi. □

Alla luce della loro importanza per il Corso, di seguito studiamo alcune leggi legate a quella Gaussiana. Il calcolo delle densità corrispondenti si può ottenere tramite l'usuale *Trasformazione di leggi*. Per ulteriori dettagli si consultino, ad esempio, [24, 12, 17, 1]. Nel seguito adotteremo la stessa notazione usata in precedenza.

6.1 Legge Chi-quadro

La legge *Chi-quadro* gioca un ruolo fondamentale nella stima della varianza di campioni Gaussiani.

DEFINIZIONE 6.1 (Legge Chi-quadro)**.** Sia $Y \sim \mathtt{N}(0,1)$; allora la v.a.

$$X = Y^2 \tag{6.2}$$

ha legge *Chi-quadro a un grado di libertà* e si scrive $X \sim \chi^2(1)$.

PROPOSIZIONE 6.3. *Sia $X \sim \chi^2(1)$; allora X ha densità f_X data da*

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{x}} e^{-\frac{1}{2}x} \mathbb{I}_{(0,\infty)}(x). \tag{6.3}$$

Inoltre $\mathbf{E}(X) = 1$ e $\mathbf{V}(X) = 2$.

NOTA 6.2.

Si osservi che $X \sim \Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$.

DEFINIZIONE 6.2. Siano $\{Y_i\}$, $i = 1, \ldots, n$, v.a. i.i.d. di legge $\mathtt{N}(0,1)$; allora la v.a.

$$X = \sum_{i=1}^{n} Y_i^2 \tag{6.4}$$

ha legge *Chi-quadro a n gradi di libertà* e si scrive $X \sim \chi^2(n)$.

PROPOSIZIONE 6.4. *Sia $X \sim \chi^2(n)$; allora X ha densità f_X data da*

$$f_X(x) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{1}{2}x} \mathbb{I}_{(0,\infty)}(x). \tag{6.5}$$

Inoltre $\mathbf{E}(X) = n$ e $\mathbf{V}(X) = 2n$.

NOTA 6.3.

Si osservi che $X \sim \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$ e che la somma di n v.a. indipendenti di legge $\chi^2(1)$ (le n v.a. Y_i^2 dell'*equazione (6.4)*) è una v.a. di legge $\chi^2(n)$.

In *Figura (6.1)* si mostra il grafico della densità di una v.a. di legge $\chi^2(n)$ per diversi valori del parametro n.
Il risulato che segue è un Corollario del *Teorema di Cochran*.

PROPOSIZIONE 6.5. *Sia $\mathbf{X}^n_{\mathtt{N}(0,\sigma^2)}$ un campione; allora la v.a.*

$$X = \|\pi_i(\boldsymbol{X})\|^2 \tag{6.6}$$

per $i = 1, \ldots, k$, ha legge $\sigma^2\chi^2(r_i) \sim \Gamma\left(\frac{r_i}{2}, \frac{1}{2\sigma^2}\right)$.

Combinando opportunamente i risultati precedenti si perviene a quanto segue.

PROPOSIZIONE 6.6. *Sia $\mathbf{X}^n_{\mathtt{N}(0,\sigma^2)}$ un campione; allora la v.a.*

$$X = \|\pi_2(\boldsymbol{Y})\|^2 = \frac{n-1}{\sigma^2}\overline{S}_n^2 \tag{6.7}$$

ha legge $\chi^2(n-1)$.

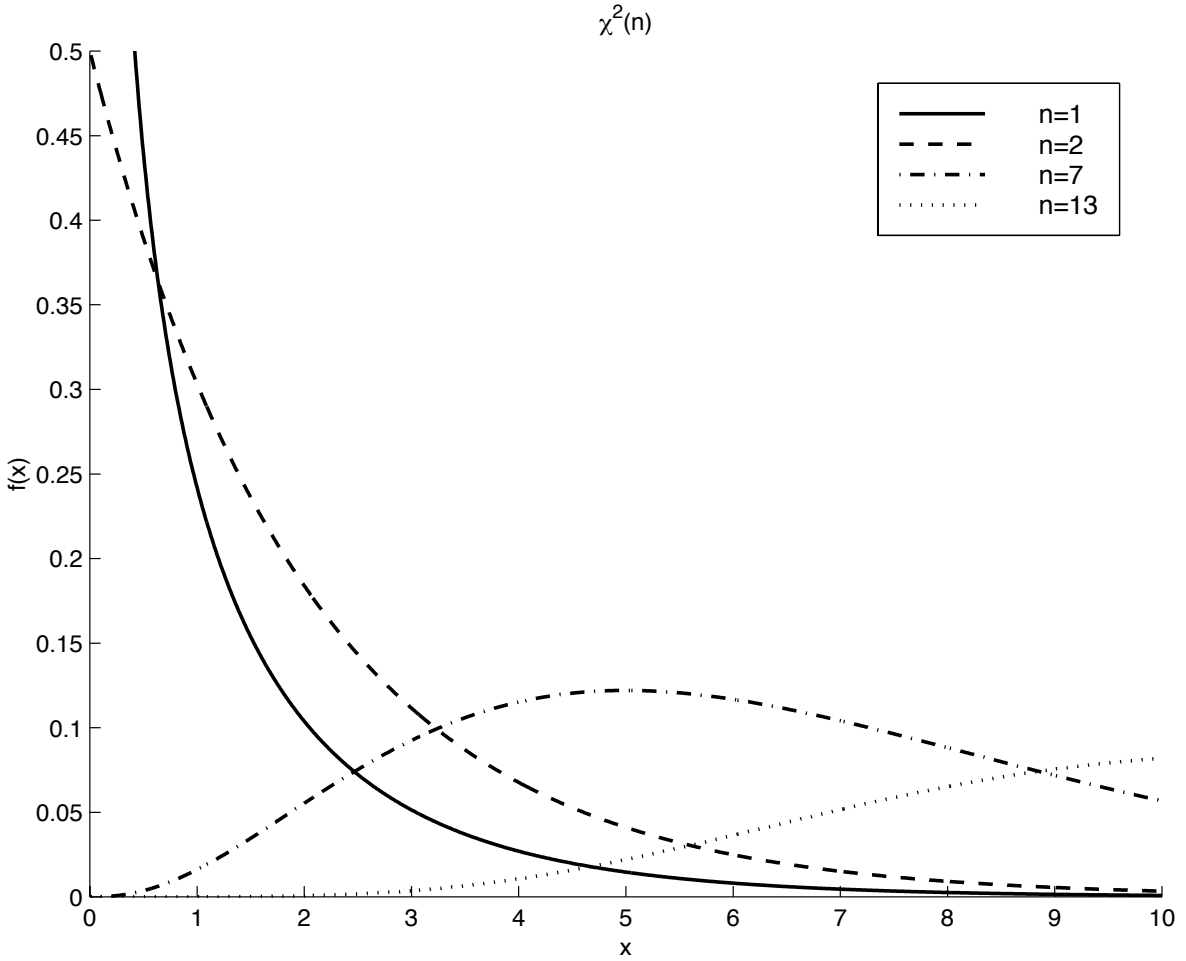

Figura 6.1: grafico della densità di una v.a. di legge $\chi^2(n)$ per diversi valori del parametro n.

6.2 Legge t-Student

La legge *t-Student* gioca un ruolo fondamentale nella stima della speranza di campioni Gaussiani.

DEFINIZIONE 6.3 (Legge t-Student)**.** Siano $U \sim \mathtt{N}(0,1)$ e $V \sim \chi^2(n)$, $n > 0$, v.a. indipendenti; allora la v.a.

$$X = \frac{U}{\sqrt{V/n}} \tag{6.8}$$

ha legge *t-Student a n gradi di libertà* e si scrive $X \sim \mathtt{t}(n)$.

PROPOSIZIONE 6.7. *Sia $X \sim \mathtt{t}(n)$; allora X ha densità f_X data da, per $x \in \mathbb{R}$,*

$$f_X(x) = \frac{1}{\mathrm{B}\left(\frac{n}{2}, \frac{1}{2}\right)} \frac{1}{\sqrt{n}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}. \tag{6.9}$$

Inoltre $\mathbf{E}(X) = 0$ se $n > 1$ e $\mathbf{V}(X) = \frac{n}{n-2}$ se $n > 2$.

OSSERVAZIONE 6.2.

La legge $\mathtt{t}(1)$ corrisponde a una legge *Cauchy* (standard) $\mathtt{C}(0,1)$ che non ha momenti di ordine $k \geq 1$. Dall'*equa-*

zione (6.9) è possibile dedurre che, per $n \to \infty$, la densità f_X approssima quella della legge Normale standard[2] (si veda anche la *Figura (6.2)*). Inoltre il momento del primo ordine esiste solo per $n > 1$ e quello del secondo ordine esiste solo per $n > 2$ — d'altro canto l'andamento delle code $\mathbf{P}\{|X| > x\}$ è di tipo *algebrico* da cui si deduce anche che, per n "piccolo" e x "grande", $\mathbf{P}\{|X| > x\} \geq \mathbf{P}\{|Z| > x\}$, dove $Z \sim \mathtt{N}(0, 1)$, e le code tendono a zero più "lentamente" rispetto al caso Gaussiano.

Infine si noti che f_X è una funzione *pari* e che quindi X è una v.a. *simmetrica*: ciò spiega il fatto che $\mathbf{E}(X) = 0$ (quando la speranza esiste) e più in generale il fatto che i momenti di ordine *dispari* (quando esistono) sono nulli, ciò che avviene anche per la legge Normale standard.

In *Figura (6.2)* si mostra il grafico della densità di una v.a. di legge $\mathtt{t}(n)$ per diversi valori del parametro n; per confronto si mostra anche la densità Normale standard.

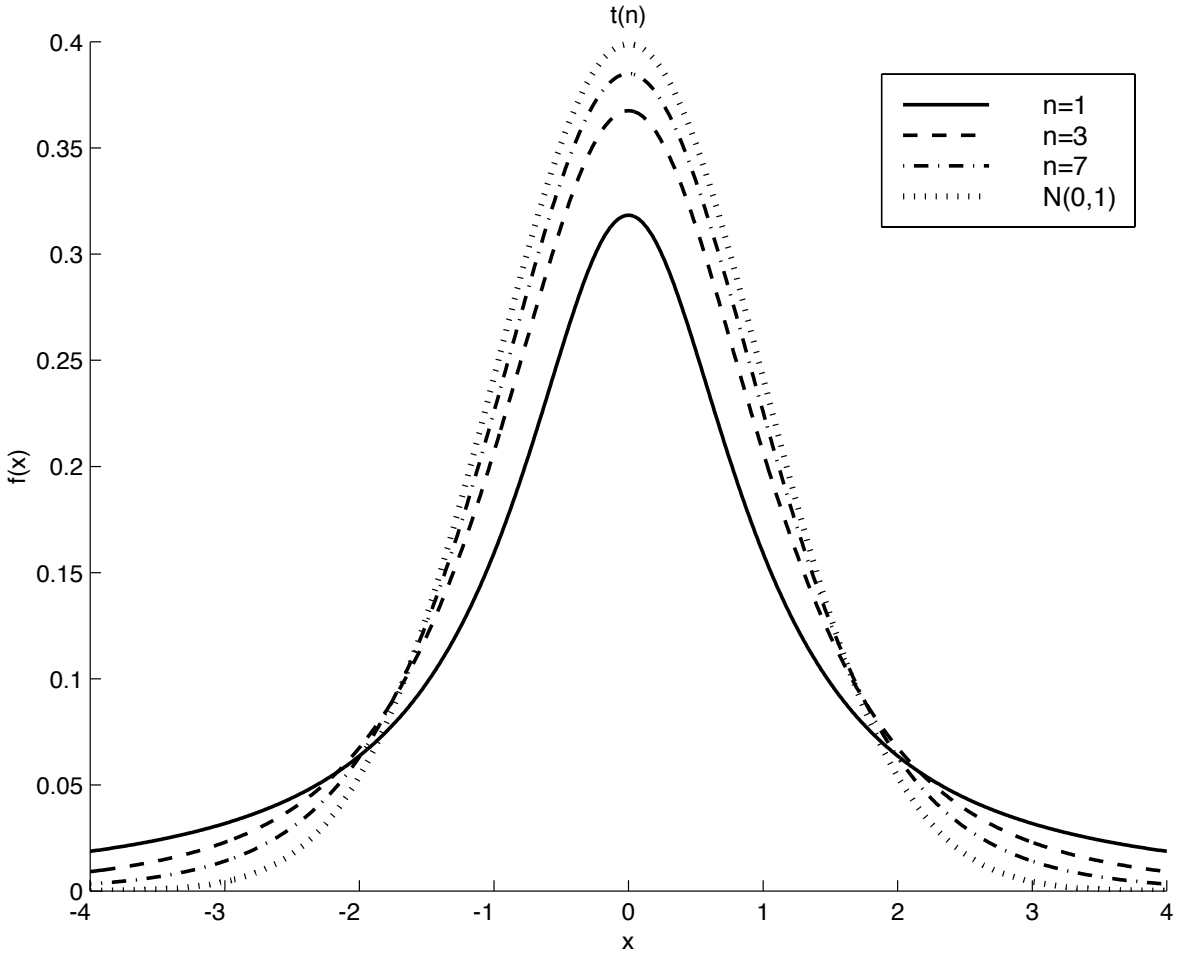

Figura 6.2: grafico della densità di una v.a. di legge $\mathtt{t}(n)$ per diversi valori del parametro n.

Il risulato che segue è un Corollario del *Teorema di Cochran*.

PROPOSIZIONE 6.8. *Sia $\mathbf{X}^n_{\mathtt{N}(\mu,\sigma^2)}$ un campione; allora la v.a.*

$$X = \sqrt{n}\,\frac{\overline{X}_n - \mu}{\overline{S}_n} \tag{6.10}$$

[2] Infatti, per $n \gg 1$, vale $\Gamma\left(\frac{n+1}{2}\right) / \left(\sqrt{\frac{n}{2}}\,\Gamma\left(\frac{n}{2}\right)\right) \approx 1$, come si ricava dalla *formula di Stirling* che fornisce l'approssimazione $\Gamma(t) \approx \sqrt{2\pi}\,e^{-t}\,t^{t-1/2}$ per $t \in \mathbb{R}_+$.

ha legge $\mathtt{t}(n-1)$.

DIM. Usando la *Proposizione (6.2)* e i risultati precedenti si ha

$$\sqrt{n}\,\frac{\overline{X}_n-\mu}{\overline{S}_n}=\sqrt{n}\,\frac{\overline{X}_n-\mu}{\frac{\sigma}{\sqrt{n-1}}\,\|\pi_2(\boldsymbol{Y})\|}=\sqrt{n}\,\frac{\overline{Y}_n}{\|\pi_2(\boldsymbol{Y})\|}\,\sqrt{n-1}=\frac{\pi_1(\boldsymbol{Y})}{\sqrt{\frac{\|\pi_2(\boldsymbol{Y})\|^2}{n-1}}},$$

che ha legge $\mathtt{t}(n-1)$. □

6.3 Legge di Fisher-Snedecor

La legge di *Fisher-Snedecor* gioca un ruolo fondamentale nel confronto delle varianze di campioni Gaussiani.

DEFINIZIONE 6.4 (Legge di Fisher-Snedecor)**.** Siano $U\sim\chi^2(m)$ e $V\sim\chi^2(n)$, $m,n>0$, v.a. indipendenti; allora la v.a.

$$X=\frac{U/m}{V/n} \tag{6.11}$$

ha legge di *Fisher-Snedecor* di parametri (m,n) e si scrive $X\sim\mathtt{F}(m,n)$.

PROPOSIZIONE 6.9. *Sia $X\sim\mathtt{F}(m,n)$; allora X ha densità f_X data da*

$$f_X(x)=\frac{\left(\frac{m}{n}\right)^{\frac{m}{2}}}{\mathrm{B}\left(\frac{m}{2},\frac{n}{2}\right)}\,\frac{x^{\frac{m-2}{2}}}{\left(1+\frac{m}{n}x\right)^{\frac{m+n}{2}}}\,\mathbb{I}_{(0,\infty)}(x). \tag{6.12}$$

Inoltre $\mathbf{E}(X)=\frac{n}{n-2}$ *se* $n>2$ *e* $\mathbf{V}(X)=\frac{2(m+n-2)n^2}{m(n-4)(n-2)^2}$ *se* $n>4$.

OSSERVAZIONE 6.3.

Si noti che $\mathtt{F}(m,n)$ non è simmetrica nei parametri m e n, cioè in generale $\mathtt{F}(m,n)\neq\mathtt{F}(n,m)$. L'esistenza della speranza è legata solamente al (numero di gradi di libertà del) denominatore nell'*equazione (6.12)*, essendo

$$\mathbf{E}(X)=\mathbf{E}\left(\frac{U/m}{V/n}\right)=\frac{n}{m}\,\mathbf{E}(U)\;\mathbf{E}\left(\frac{1}{V}\right)=n\,\mathbf{E}\left(\frac{1}{V}\right).$$

Infine è chiaro dall'*equazione (6.12)* che il momento del primo ordine esiste solo per $n>2$ e quello del secondo ordine solo per $n>4$ — d'altro canto l'andamento della coda $\mathbf{P}\{X>x\}$ è di tipo *algebrico*.

In *Figura (6.3)* si mostra il grafico della densità di una v.a. di legge $\mathtt{F}(m,n)$ per diversi valori dei parametri m,n.

Combinando opportunamente i risultati precedenti si perviene a quanto segue.

PROPOSIZIONE 6.10. *Siano $\mathbf{Y}^m_{\mathtt{N}\left(0,\sigma_1^2\right)}$ e $\mathbf{Z}^n_{\mathtt{N}\left(0,\sigma_2^2\right)}$ due campioni Gaussiani indipendenti; allora la v.a.*

$$X=\frac{\overline{S}^2_{m,Y}/\sigma_1^2}{\overline{S}^2_{n,Z}/\sigma_2^2} \tag{6.13}$$

ha legge $\mathtt{F}(m-1,n-1)$.

Il risultato della *Proposizione (6.10)* giustifica il fatto che in Statistica Matematica la legge di *Fisher-Snedecor* sia anche detta *legge del rapporto delle varianze*.

ESEMPIO 6.1. ►

Sia $X\sim\mathtt{F}(1,n)$, $n>0$; per calcolarne la legge si può procedere nel seguente modo. Il calcolo diretto (ponendo $m=1$ nell'*equazione (6.12)*) non fornisce indicazioni. Ragioniamo allora come segue: siano $U\sim\chi^2(1)$ e $V\sim\chi^2(n)$ v.a. indipendenti e sia $Z\sim\mathtt{N}(0,1)$ indipendente da V; poiché $X\sim\mathtt{F}(1,n)$ si ha

$$X\sim\frac{U}{V/n}\sim\frac{Z^2}{\left(\sqrt{V/n}\right)^2}=\left(\frac{Z}{\sqrt{V/n}}\right)^2\sim\left(\mathtt{t}(n)\right)^2,$$

e quindi la legge di X è quella del quadrato di una v.a. con legge $\mathtt{t}(n)$. ◄

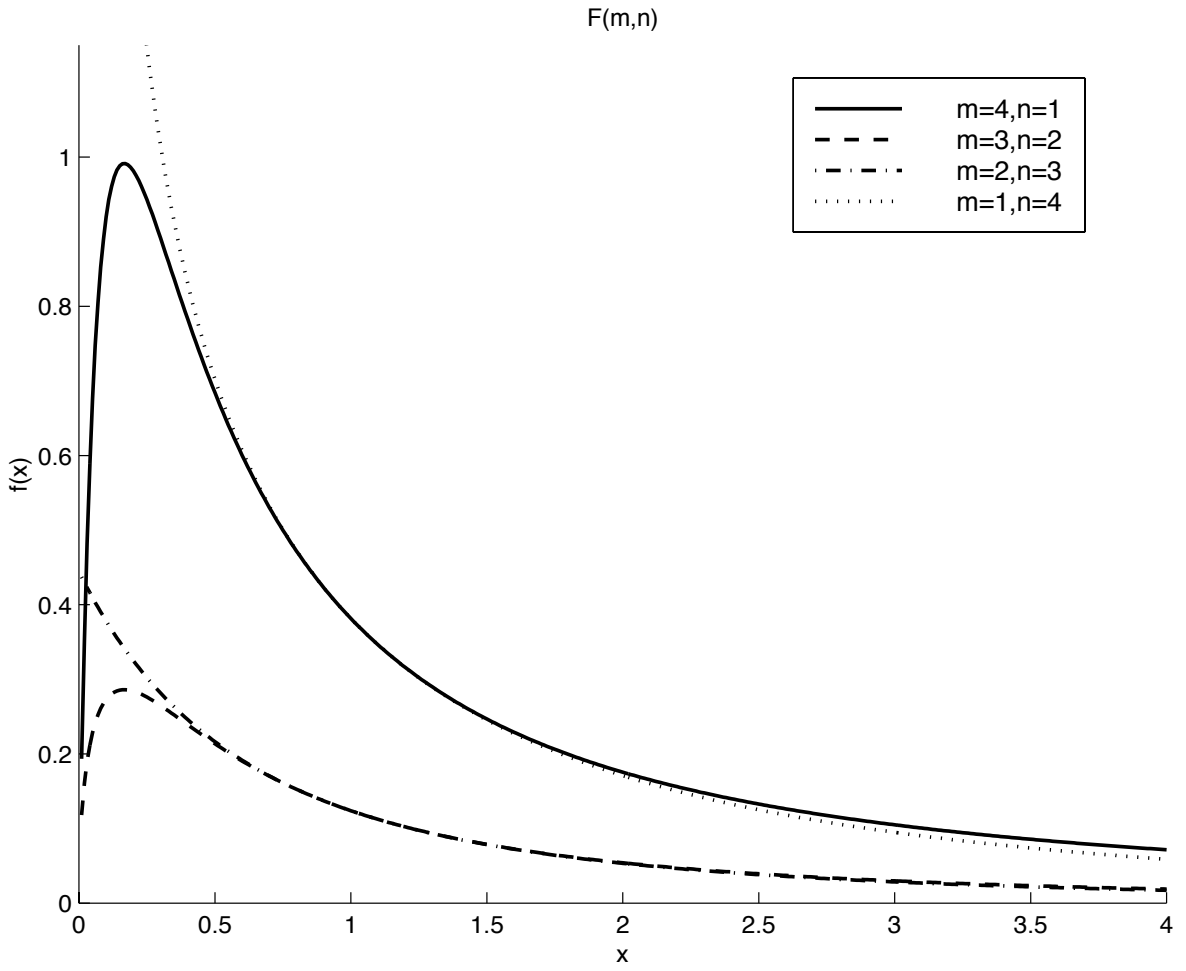

Figura 6.3: grafico della densità di una v.a. di legge $\mathtt{F}(m,n)$ per diversi valori dei parametri m, n.

ESEMPIO 6.2. ►

Sia $X \sim \mathtt{F}(1,1)$; per calcolarne la legge si può procedere nel seguente modo. Sfruttando il risultato dell'*Esempio (6.1)* e ponendo $n = 1$ otteniamo che la legge di X è quella del quadrato di una v.a. con legge $\mathtt{t}(1)$, cioè la legge del quadrato di una v.a. con legge $\mathtt{C}(0,1)$ in base all'*Osservazione (6.2)*. ◄

ESEMPIO 6.3. ►

Sia $X \sim \mathtt{F}(m,n)$, $m, n > 0$; è possibile calcolare la legge della v.a. $Y = 1/X$ semplicemente ricordando la *Definizione (6.4)* e usando l'*equazione (6.11)*, ottenendo $Y \sim \mathtt{F}(n,m)$ ◄

ESEMPIO 6.4 (Quantili della legge Fisher-Snedecor)**.** ►

Siano $X \sim \mathtt{F}(m,n)$ e $Y \sim \mathtt{F}(n,m)$, $m, n > 0$, e sia $p \in (0,1)$; allora tra i quantili di X e Y vale la relazione $x_p = \frac{1}{y_{1-p}}$. Infatti, usando i risultati della *Sezione 0.3* e dell'*Esempio (6.3)*, si ha

$$p = \mathbf{P}\{X \leq x_p\} = \mathbf{P}\left\{\frac{1}{X} \geq \frac{1}{x_p}\right\} = \mathbf{P}\left\{Y \geq \frac{1}{x_p}\right\},$$

e quindi $\mathbf{P}\left\{Y \leq \frac{1}{x_p}\right\} = 1 - p$; essendo anche $1 - p = \mathbf{P}\{Y \leq y_{1-p}\}$ avremo che $\frac{1}{x_p} = y_{1-p}$. Dunque, in pratica è sufficiente tabulare la legge $\mathtt{F}(m,n)$ per alcuni valori dei parametri m, n e per valori $p \leq \frac{1}{2}$ oppure $p \geq \frac{1}{2}$; usando la relazione data è poi possibile ottenere la tabulazione completa della legge in questione. ◄

Capitolo 7

VERIFICA DI IPOTESI

Nelle applicazioni accade spesso che l'oggetto del problema sia rappresentato non tanto dalla stima (puntuale) dei parametri delle leggi in gioco (che è stata trattata nelle Parti precedenti), quanto piuttosto dalla necessità di scegliere tra diverse *ipotesi statistiche* formulate circa i fenomeni oggetto di studio. In buona sostanza si tratta di scegliere se rifiutare o accettare le ipotesi stesse — *ma non di giudicare se siano Vere o False!* La *verifica di ipotesi* gioca un ruolo fondamentale nell'ambito della Statistica Matematica e, come vedremo nel seguito, è strettamente connessa al problema della stima affrontato in precedenza. I due esempi che seguono illustrano l'argomento.

1. Si consideri una serie di lanci indipendenti di una moneta, ad esempio ai fini di una scommessa. Solitamente si ritiene che la moneta sia equilibrata, ma come è possibile verificare che tale ipotesi sia corretta? Intuitivamente una procedura ragionevole consiste nell'effettuare un certo numero di lanci e calcolare la percentuale p di teste osservate. Evidentemente se p non è "troppo" diversa da $1/2$ allora l'ipotesi formulata può essere accettata.

2. Un commerciante dichiara che la durata media delle batterie in vendita è (almeno) pari a τ ore; come è possibile verificare che tale affermazione sia veritiera? Intuitivamente una procedura ragionevole consiste nell'osservare un campione di tali batterie, diciamo n, e misurare la durata di ciascuna. Se la vita media delle stesse (stimata, ad esempio, tramite $\overline{X}_n$) risulta essere "troppo" inferiore a τ allora l'affermazione del commerciante può essere messa in dubbio.

La verifica di ipotesi è stata creata per risolvere problemi del tipo appena esposto; il linguaggio e le procedure appena descritte sono solo qualitative e necessitano di una formalizzazione matematica. Nel seguito assumeremo sempre che vi sia un modello statistico dominato $\mathfrak{M} = \big(\Omega, \mathcal{F}, \{\mathbf{P}_\theta\}_{\theta\in\Theta}\big)$ soggiacente; senza perdita di generalità considereremo un campione $\mathbf{X}^n_{f_\theta}$, con $\theta \in \Theta \subseteq \mathbb{R}^d$, $d \geq 1$, dove la densità f_θ sarà supposta essere nota a meno del parametro θ. Assumeremo inoltre che Θ contenga almeno due punti e useremo in maniera indifferenziata la notazione "ω" o "$\boldsymbol{x}$" per indicare una realizzazione (osservazione) del campione in esame.

DEFINIZIONE 7.1 (Ipotesi nulla e alternativa)**.** Una *ipotesi statistica (parametrica)* è un'affermazione circa il parametro incognito θ. Data la partizione $\{\Theta_0, \Theta_1\}$ di Θ, cioè $\Theta = \Theta_0 \cup \Theta_1$ e $\Theta_0 \cap \Theta_1 = \emptyset$, si dice *ipotesi nulla* l'affermazione $\mathcal{H}_0 \colon \theta \in \Theta_0$ e *ipotesi alternativa* l'affermazione $\mathcal{H}_1 \colon \theta \in \Theta_1$.

DEFINIZIONE 7.2 (Ipotesi semplice e composta)**.** Se Θ_0 contiene un solo punto allora $\mathcal{H}_0$ si dice *ipotesi semplice*, altrimenti *ipotesi composta*; lo stesso vale per $\mathcal{H}_1$.

NOTA 7.1.

Se un'ipotesi è semplice allora la legge f_θ del campione risulta essere completamente specificata sotto quell'ipotesi.

ESEMPIO 7.1. ►

Sia $X \sim \mathtt{N}\big(\mu, \sigma^2\big)$, con $\mu \in \mathbb{R}$ e $\sigma^2 > 0$; dunque $\Theta = \mathbb{R} \times \mathbb{R}_+$. Siano $\mu_0 \in \mathbb{R}$ e $\sigma_0^2 > 0$ costanti note. L'ipotesi nulla $\mathcal{H}_0 \colon \mu \leq \mu_0, \sigma^2 > 0$ è composta, così come lo è l'ipotesi alternativa $\mathcal{H}_1 \colon \mu > \mu_0, \sigma^2 > 0$: infatti in nessun caso la legge di X risulta essere specificata. Parimenti sono composte le ipotesi $\mathcal{H}_0 \colon \mu = \mu_0, \sigma^2 > 0$ e $\mathcal{H}_1 \colon \mu \neq \mu_0, \sigma^2 > 0$. Se $\sigma^2 = \sigma_0^2$ allora l'ipotesi nulla $\mathcal{H}_0 \colon \mu = \mu_0$ è semplice, mentre l'ipotesi alternativa $\mathcal{H}_1 \colon \mu \neq \mu_0$ risulta essere composta. ◄

La verifica di ipotesi può essere descritta come segue: (*1*) dato un campione $\mathbf{X}_{f_\theta}^n$, e (*2*) osservata una sua realizzazione $\boldsymbol{x} = (x_1, \ldots, x_n)$, (*3*) si costruisca una *regola decisionale* (o *test*) che permetta di rifiutare o accettare l'ipotesi nulla. Si noti che nel primo caso si accetta l'ipotesi alternativa, mentre nel secondo caso la si rifiuta. La costruzione della regola decisionale corrisponde all'individuazione di una partizione di $\mathbb{R}^n$, cioè $\mathbb{R}^n = C \cup C^{\complement}$, di modo che

- se $\boldsymbol{x} \in C$ allora si rifiuta $\mathcal{H}_0$ e si accetta $\mathcal{H}_1$, cioè l'ipotesi $X \sim f_\theta$ con $\theta \in \Theta_1$;
- se $\boldsymbol{x} \in C^{\complement}$ allora si rifiuta $\mathcal{H}_1$ e si accetta $\mathcal{H}_0$, cioè l'ipotesi $X \sim f_\theta$ con $\theta \in \Theta_0$.

In ogni caso l'osservazione del campione è fondamentale al fine di rifiutare o accettare l'ipotesi nulla.

DEFINIZIONE 7.3 (Regione critica e di accettazione)**.** Sia $\mathbf{X}_{f_\theta}^n$ un campione e sia $\boldsymbol{x} = (x_1, \ldots, x_n)$ una sua realizzazione. Si dice *regione critica* (o *regione di rifiuto*) l'insieme misurabile $C \subset \mathbb{R}^n$ tale che se $\boldsymbol{x} \in C$ allora $\mathcal{H}_0$ è rifiutata con probabilità 1. L'insieme $C^{\complement}$ è detto invece *regione di accettazione*.

DEFINIZIONE 7.4 (Funzione test)**.** Sia $\tau\colon \mathbb{R}^n \to [0,1]$ una funzione misurabile; allora τ è detta *funzione test*. Se τ è la funzione indicatrice di un Boreliano $C \in \mathcal{B}(\mathbb{R}^n)$ allora C è detto *regione critica del test* τ.

Un esempio di funzione test è dato dalla funzione $\tau(\boldsymbol{x}) = \alpha$ per ogni $\boldsymbol{x} \in \mathbb{R}^n$, dove $0 \leq \alpha \leq 1$. Se C è la regione critica di un test $\tau(\boldsymbol{x}) = \mathbb{I}_C(\boldsymbol{x})$ si ha che

- se $\boldsymbol{x} \in C$ allora $\tau(\boldsymbol{x}) = 1$ e si rifiuta $\mathcal{H}_0$ accettando $\mathcal{H}_1$;
- se $\boldsymbol{x} \in C^{\complement}$ allora $\tau(\boldsymbol{x}) = 0$ e si accetta $\mathcal{H}_0$ rifiutando $\mathcal{H}_1$.

In pratica si accetta l'ipotesi $\mathcal{H}_{\tau(\boldsymbol{x})}$.

NOTA 7.2.

In effetti la verifica di ipotesi può essere formalizzata come un *problema decisionale* ove l'insieme delle *azioni* è $\mathfrak{A} = \{0, 1\}$: l'azione $a = 0$ corrisponde ad accettare l'ipotesi $\mathcal{H}_0$ (o, equivalentemente, a rifiutare l'alternativa $\mathcal{H}_1$), l'azione $a = 1$ corrisponde a rifiutare l'ipotesi $\mathcal{H}_0$ (o, equivalentemente, ad accettare l'alternativa $\mathcal{H}_1$).

Nella verifica di ipotesi si introduce la *funzione perdita* (o *costo*) $L\colon \Theta \times \mathfrak{A} \to \mathbb{R}$ data da

$$L(\theta, a) = a\,\mathbb{I}_{\Theta_0}(\theta) + (1-a)\,\mathbb{I}_{\Theta_1}(\theta) = \begin{cases} a, & \text{se } \theta \in \Theta_0 \\ 1-a, & \text{se } \theta \in \Theta_1 \end{cases}. \tag{7.1}$$

Se $\theta \in \Theta_0$ e si compie l'azione $a = 0$ la perdita è nulla; lo stesso vale se $\theta \in \Theta_1$ e si compie l'azione $a = 1$.

DEFINIZIONE 7.5 (Decisione)**.** Una *decisione* è una funzione misurabile $\delta\colon \mathbb{R}^n \to \mathfrak{A}$.

Evidentemente, una decisione dipende dall'osservazione di un campione $\boldsymbol{X}$. Sia ora C una regione critica e $\delta(\boldsymbol{X}) = \mathbb{I}_C(\boldsymbol{X})$ una decisione; allora il *rischio* $\mathbf{R}$ (o *perdita media*) insito nella decisione δ è

$$\begin{aligned} \mathbf{R}_\theta(\delta(\boldsymbol{X})) &= \mathbf{E}_\theta(L(\theta, \delta(\boldsymbol{X}))) \\ &= \mathbf{E}_\theta(\delta(\boldsymbol{X})\,\mathbb{I}_{\Theta_0}(\theta) + (1-\delta(\boldsymbol{X}))\,\mathbb{I}_{\Theta_1}(\theta)) \\ &= \mathbf{E}_\theta(\mathbb{I}_C(\boldsymbol{X})\,\mathbb{I}_{\Theta_0}(\theta) + (1-\mathbb{I}_C(\boldsymbol{X}))\,\mathbb{I}_{\Theta_1}(\theta)) \\ &= \mathbf{P}_\theta\{\boldsymbol{X} \in C\}\,\mathbb{I}_{\Theta_0}(\theta) + (1-\mathbf{P}_\theta\{\boldsymbol{X} \in C\})\,\mathbb{I}_{\Theta_1}(\theta), \end{aligned}$$

cioè

$$\mathbf{R}_\theta(\delta(\boldsymbol{X})) = \mathbf{E}_\theta(L(\theta, \delta(\boldsymbol{X}))) = \begin{cases} \mathbf{P}_\theta\{\boldsymbol{X} \in C\}, & \text{se } \theta \in \Theta_0 \\ \mathbf{P}_\theta\{\boldsymbol{X} \notin C\}, & \text{se } \theta \in \Theta_1 \end{cases}. \tag{7.2}$$

Nella verifica di ipotesi due sono gli errori che si possono commettere:

Errore di TIPO I: rifiutare l'ipotesi nulla $\mathcal{H}_0$ quando questa sia corretta (vera);

Errore di TIPO II: accettare l'ipotesi nulla $\mathcal{H}_0$ quando questa sia sbagliata (falsa).

Formalmente[1] si ha la seguente definizione.

DEFINIZIONE 7.6 (Errore di TIPO I e TIPO II)**.** Si commette un errore di TIPO I rifiutando $\mathcal{H}_0$ quando questa sia vera. Si commette un errore di TIPO II accettando $\mathcal{H}_0$ quando questa sia falsa.

Nella pratica solitamente si cerca di strutturare la procedura di verifica delle ipotesi in modo che un errore di TIPO I sia "più grave" (abbia conseguenze "peggiori") di un errore di TIPO II. Si considerino ad esempio le seguenti situazioni:

- $\mathcal{H}_0$*: il paziente è affetto da tumore* contro $\mathcal{H}_1$*: il paziente non è affetto da tumore*;
- $\mathcal{H}_0$*: il carburante è in esaurimento* contro $\mathcal{H}_1$*: il carburante non è in esaurimento*.

Evidentemente un errore di TIPO I (cioè il mancato riconoscimento dell'effettiva presenza del tumore o dell'effettiva mancanza di carburante) può portare al decesso del malato o ad una fermata del mezzo; al contrario un errore di TIPO II (cioè il dichiarare malato un paziente sano oppure il non riconoscere la presenza di una riserva sufficiente di carburante) può avere conseguenze di minore entità. Poiché la verifica di ipotesi avviene essenzialmente tramite l'osservazione di campioni statistici, è possibile definire la probabilità di commettere un errore di TIPO I o di TIPO II.

DEFINIZIONE 7.7 (Probabilità di errore)**.** Sia C la regione critica di un test. Allora, se $\boldsymbol{X}$ è il vettore delle osservazioni,

- $\mathbf{P}_\theta\{C\} = \mathbf{P}_\theta\{\boldsymbol{X} \in C\}$, $\theta \in \Theta_0$, rappresenta la probabilità di commettere un errore di TIPO I,
- $\mathbf{P}_\theta\left\{C^{\complement}\right\} = \mathbf{P}_\theta\left\{\boldsymbol{X} \in C^{\complement}\right\}$, $\theta \in \Theta_1$, rappresenta la probabilità di commettere un errore di TIPO II.

Naturalmente sarebbe auspicabile individuare una regione critica tale che entrambe le probabilità di errore siano nulle; tuttavia tale situazione non è (quasi) mai realizzabile in pratica. Si può però introdurre un *ordinamento parziale*[2] sull'insieme dei test.

DEFINIZIONE 7.8 (Ordinamento)**.** Si dice che

1. un test di regione critica C_1 è *preferibile* ad un altro test di regione critica C_2 se

$$\text{per ogni } \theta \in \Theta_0 \text{ si ha } \mathbf{P}_\theta\{C_1\} \leq \mathbf{P}_\theta\{C_2\}, \tag{7.3a}$$

$$\text{per ogni } \theta \in \Theta_1 \text{ si ha } \mathbf{P}_\theta\{C_1\} \geq \mathbf{P}_\theta\{C_2\}; \tag{7.3b}$$

2. un test di regione critica C_1 è *migliore* di un altro test di regione critica C_2 se è preferibile ad esso e almeno una delle due disuguaglianze delle *equazioni (7.3)* è stretta;
3. un test è *ammissibile* se non ne esistono di migliori;
4. un test è *ottimale* se è preferibile ad ogni altro test.

Poiché l'errore di TIPO I è in genere considerato più grave di quello di TIPO II, nella pratica si cerca di limitare la probabilità del primo ad un preassegnato livello $0 \leq \alpha \leq 1$ fissato, per tentare poi di minimizzare la probabilità del secondo. A tale fine è necessario introdurre le seguenti definizioni.

DEFINIZIONE 7.9 (Livello di un test)**.** Sia $\mathbf{X}_{f_\theta}^n$ un campione. La funzione test τ è detta *test dell'ipotesi* $\mathcal{H}_0 : \theta \in \Theta_0$ contro l'alternativa $\mathcal{H}_1 : \theta \in \Theta_1$ di *livello* $0 \leq \alpha \leq 1$ se si ha, per ogni $\theta \in \Theta_0$,

$$\mathbf{E}_\theta(\tau(X_1, \ldots, X_n)) = \mathbf{P}_\theta\{\text{rifiutare } \mathcal{H}_0\} \leq \alpha. \tag{7.4}$$

Il test τ è detto *test per il problema* $(\alpha, \Theta_0, \Theta_1)$. Nel seguito indicheremo con $\mathfrak{T}_\alpha$ la classe di tutti i test di livello α per il problema $(\alpha, \Theta_0, \Theta_1)$.

[1] Si ricordi la **Regola del RiVAF**: **Ri**fiutare **V**era, **A**ccettare **F**alsa.

[2] L'ordinamento è *parziale* in quanto non sempre due test sono confrontabili.

OSSERVAZIONE 7.1.

Se $\tau = \mathbb{I}_C$, dove C è una regione critica, e $\theta \in \Theta_0$, allora si ha

$$\mathbf{E}_\theta(\tau(\boldsymbol{X})) = \mathbf{P}_\theta\{\boldsymbol{X} \in C\}\,.$$

Inoltre si noti che se $\tau_1, \tau_2 \in \mathfrak{T}_\alpha$ e $0 \leq c \leq 1$ allora anche $\tau = c\,\tau_1 + (1-c)\,\tau_2$ è un test di livello α per il problema $(\alpha, \Theta_0, \Theta_1)$: dunque la classe $\mathfrak{T}_\alpha$ è *convessa*.

Usando la notazione della *Definizione (7.4)* e della *Definizione (7.9)* è possibile introdurre le seguenti nozioni.

DEFINIZIONE 7.10 (Funzione di potenza)**.** La funzione $\boldsymbol{\beta}_\tau \colon \Theta \to [0,1]$ data da

$$\boldsymbol{\beta}_\tau(\theta) = \mathbf{E}_\theta(\tau(X_1, \ldots, X_n)) = \mathbf{P}_\theta\{\text{rifiutare } \mathcal{H}_0\} \tag{7.5}$$

è detta *funzione di potenza* del test τ.

NOTA 7.3.

Un test τ è di livello $0 \leq \alpha \leq 1$ se si ha, per ogni $\theta \in \Theta_0$, $\boldsymbol{\beta}_\tau(\theta) \leq \alpha$.

DEFINIZIONE 7.11 (Ampiezza di un test)**.** Sia τ un test di regione critica C. La quantità

$$\sup_{\theta \in \Theta_0} \boldsymbol{\beta}_\tau(\theta) = \sup_{\theta \in \Theta_0} \mathbf{P}_\theta\{C\} = \sup_{\theta \in \Theta_0} \mathbf{P}_\theta\{\text{rifiutare } \mathcal{H}_0\} \tag{7.6}$$

è detta *ampiezza* (o *taglia*) del test τ.

La potenza di un test ne rappresenta una caratteristica importante nella verifica di ipotesi.

NOTA 7.4.

Attraverso la funzione di potenza è possibile introdurre una classificazione (*ordinamento parziale*) sull'insieme dei test.

DEFINIZIONE 7.12 (Test più potente (PP))**.** Un test $\tau_1 \in \mathfrak{T}_\alpha$ dell'ipotesi $\mathcal{H}_0 \colon \theta \in \Theta_0$ contro l'alternativa $\mathcal{H}_1 \colon \theta \in \Theta_1$ è *più potente* (PP) di un test $\tau_2 \in \mathfrak{T}_\alpha$ se per ogni $\theta \in \Theta_1$ si ha

$$\boldsymbol{\beta}_{\tau_1}(\theta) \geq \boldsymbol{\beta}_{\tau_2}(\theta)\,. \tag{7.7}$$

DEFINIZIONE 7.13 (Test uniformemente più potente (UPP))**.** Un test $\tau^* \in \mathfrak{T}_\alpha$ dell'ipotesi $\mathcal{H}_0 \colon \theta \in \Theta_0$ contro l'alternativa $\mathcal{H}_1 \colon \theta \in \Theta_1$ è *uniformemente più potente* (UPP) se per ogni test $\tau \in \mathfrak{T}_\alpha$ e per ogni $\theta \in \Theta_1$ si ha

$$\boldsymbol{\beta}_{\tau^*}(\theta) \geq \boldsymbol{\beta}_\tau(\theta)\,. \tag{7.8}$$

NOTA 7.5.

Un test UPP non è necessariamente ammissibile, in quanto potrebbe non essere confrontabile con altri test.

Se le ipotesi $\mathcal{H}_0, \mathcal{H}_1$ sono composte la verifica di ipotesi si riduce a costruire un test che sia UPP per il problema $(\alpha, \Theta_0, \Theta_1)$. Spesso, purtroppo, un tale test non esiste e dunque, come vedremo meglio in seguito, si dovranno porre opportune restrizioni alla classe $\mathfrak{T}_\alpha$.

Riassumendo quanto visto finora, la verifica di ipotesi procede nei termini seguenti. Sia τ un test; ad ogni $\boldsymbol{x} \in \mathbb{R}^n$ si associa, attraverso τ, la probabilità di rifiutare l'ipotesi $\mathcal{H}_0 \colon X \sim f_\theta, \theta \in \Theta_0$, se $\boldsymbol{x}$ è osservato; la restrizione imposta dall'*equazione (7.4)* equivale a dire che se l'ipotesi $\mathcal{H}_0$ fosse vera allora il test τ la rifiuterebbe con probabilità non superiore a $0 \leq \alpha \leq 1$. Contemporaneamente si cerca di fare in modo che $\boldsymbol{\beta}_\tau(\theta)$ sia massima per $\theta \in \Theta_1$. Usando la notazione introdotta finora è possibile distinguere due tipi di test.

DEFINIZIONE 7.14 (Test deterministico e aleatorio)**.** Se $\tau = \mathbb{I}_C$, dove C è una regione critica, allora τ è detto *test deterministico*, altrimenti è detto *test aleatorio* (o *casuale*).

Sia C una regione critica e $\tau = \mathbb{I}_C$ un test deterministico. Allora

$$\tau(x) = \begin{cases} 1, & \text{se } x \in C \text{ e si } \textit{rifiuta } \mathcal{H}_0 \text{ con probabilità } 1 \\ 0, & \text{se } x \notin C \text{ e si } \textit{accetta } \mathcal{H}_0 \text{ con probabilità } 1 \end{cases}.$$

Se invece τ è un test aleatorio è possibile rifiutare o accettare $\mathcal{H}_0$ con probabilità esattamente pari a $0 < \gamma < 1$. Formalmente in un test aleatorio l'insieme delle azioni non è più l'insieme discreto $\{0, 1\}$ bensì tutto l'intervallo $\mathfrak{A} = [0, 1]$; un test siffatto è dunque una funzione (decisione) $\tau \colon \Omega \to \mathfrak{A}$ misurabile rispetto agli spazi di misura $(\Omega, \mathcal{F})$ e $(\mathfrak{A}, \mathcal{B}(\mathfrak{A}))$. In un test aleatorio la verifica di ipotesi procede come segue:

- se $\tau(\omega) = 1$ allora si rifiuta $\mathcal{H}_0$ con probabilità 1;
- se $\tau(\omega) = 0$ allora si accetta $\mathcal{H}_0$ con probabilità 1;
- se $\tau(\omega) = \gamma$, $0 < \gamma < 1$, allora si rifiuta $\mathcal{H}_0$ con probabilità γ.

Nell'ultimo dei casi appena descritti si ricorre ad un'estrazione dalla legge $\mathtt{Be}(\gamma)$ che sia *indipendente* dall'osservazione ω: si rifiuta $\mathcal{H}_0$ se il risultato è 1 (ciò che avviene con probabilità γ), mentre la si accetta se esso è 0 (ciò che avviene con probabilità $1 - \gamma$). Il *costo* dell'azione $a \in \mathfrak{A} = [0, 1]$ è

$$L(\theta, a) = a\, \mathbb{I}_{\Theta_0}(\theta) + (1 - a)\, \mathbb{I}_{\Theta_1}(\theta),$$

sicché il rischio $\mathbf{R}$ risulta essere pari a

$$\mathbf{R}_\theta(\tau) = \mathbf{E}_\theta(L(\theta, \tau)) = \begin{cases} \mathbf{E}_\theta(\tau)\,, & \text{se } \theta \in \Theta_0 \\ 1 - \mathbf{E}_\theta(\tau)\,, & \text{se } \theta \in \Theta_1 \end{cases}.$$

NOTA 7.6.

Sostituendo $\mathbf{R}_\theta$ a $\mathbf{P}_\theta$ nella *Definizione (7.8)* è possibile introdurre una classificazione (*ordinamento parziale*) sull'insieme dei test aleatori.

★ **Il problema della verifica di ipotesi** può essere riformulato come segue: *sia $\mathbf{X}^n_{f_\theta}$, $\theta \in \Theta$, un campione e sia $0 \leq \alpha \leq 1$ fissata; data una realizzazione $\boldsymbol{x} \in \mathbb{R}^n$ del campione si trovi un test τ tale che $\boldsymbol{\beta}_\tau(\theta) \leq \alpha$ per ogni $\theta \in \Theta_0$ e $\boldsymbol{\beta}_\tau(\theta)$ sia massima per $\theta \in \Theta_1$.*

ESEMPIO 7.2. ►

Si consideri il campione $\mathbf{X}^n_{\mathtt{N}(\mu,1)}$, dove $\mu \in \Theta = \{\mu_0, \mu_1\}$, $\mu_0 < \mu_1$, e si formulino le ipotesi

$$\mathcal{H}_0 \colon X \sim f_{\Theta_0} \sim \mathtt{N}(\mu_0, 1) \quad \text{contro} \quad \mathcal{H}_1 \colon X \sim f_{\Theta_1} \sim \mathtt{N}(\mu_1, 1)\,.$$

Evidentemente sia $\mathcal{H}_0$ sia $\mathcal{H}_1$ sono ipotesi semplici. Intuitivamente si accetta $\mathcal{H}_0$ se $\overline{X}_n$ (che è uno stimatore della speranza di X) è "più vicina" a μ_0 che non a μ_1; in altri termini si rifiuta $\mathcal{H}_0$ se $\overline{X}_n > r \in \mathbb{R}$, dove la costante r può essere determinata sulla base del livello del test. Sotto $\mathcal{H}_0$ si ha $\overline{X}_n \sim \mathtt{N}(\mu_0, 1/n)$; fissato $0 < \alpha < 1$ vale

$$\mathbf{P}\{\text{errore di TIPO I}\} = \mathbf{P}_{\mathcal{H}_0}\{\text{rifiutare } \mathcal{H}_0\} = \mathbf{P}_{\mathcal{H}_0}\left\{\overline{X}_n > r\right\} = \mathbf{P}_{\mathcal{H}_0}\left\{\frac{\overline{X}_n - \mu_0}{1/\sqrt{n}} > \frac{r - \mu_0}{1/\sqrt{n}}\right\} = \alpha,$$

e dunque, essendo (sotto $\mathcal{H}_0$) $\frac{\overline{X}_n - \mu_0}{1/\sqrt{n}} \sim \mathtt{N}(0, 1)$, si ha $\frac{r - \mu_0}{1/\sqrt{n}} = z_{1-\alpha}$, cioè $r = \mu_0 + z_{1-\alpha}/\sqrt{n}$, dove $z_{1-\alpha}$ rappresenta il quantile di ordine $1 - \alpha$ (si veda la *Sezione 0.3*) della legge $\mathtt{N}(0, 1)$. Dunque un test τ di livello α può essere costruito come, per $\boldsymbol{x} \in \mathbb{R}^n$,

$$\tau(x_1, \ldots, x_n) = \begin{cases} 1 \text{ (si rifiuti } \mathcal{H}_0), & \text{se } \overline{x}_n > \mu_0 + z_{1-\alpha}/\sqrt{n} \\ 0 \text{ (si accetti } \mathcal{H}_0), & \text{altrimenti} \end{cases},$$

dove $\overline{x}_n = \frac{1}{n}\sum_{i=1}^n x_i$. La v.a. $\overline{X}_n$ è detta *statistica test*; il test τ proposto è deterministico, con regione critica data da $C = \{\boldsymbol{x} \in \mathbb{R}^n \colon \overline{x}_n > \mu_0 + z_{1-\alpha}/\sqrt{n}\}$. Si osservi come, al diminuire di α, il quantile $z_{1-\alpha}$ aumenti e, di conseguenza, aumenti la soglia critica di rifiuto per la statistica test $\overline{X}_n$. In altri termini, la diminuzione della

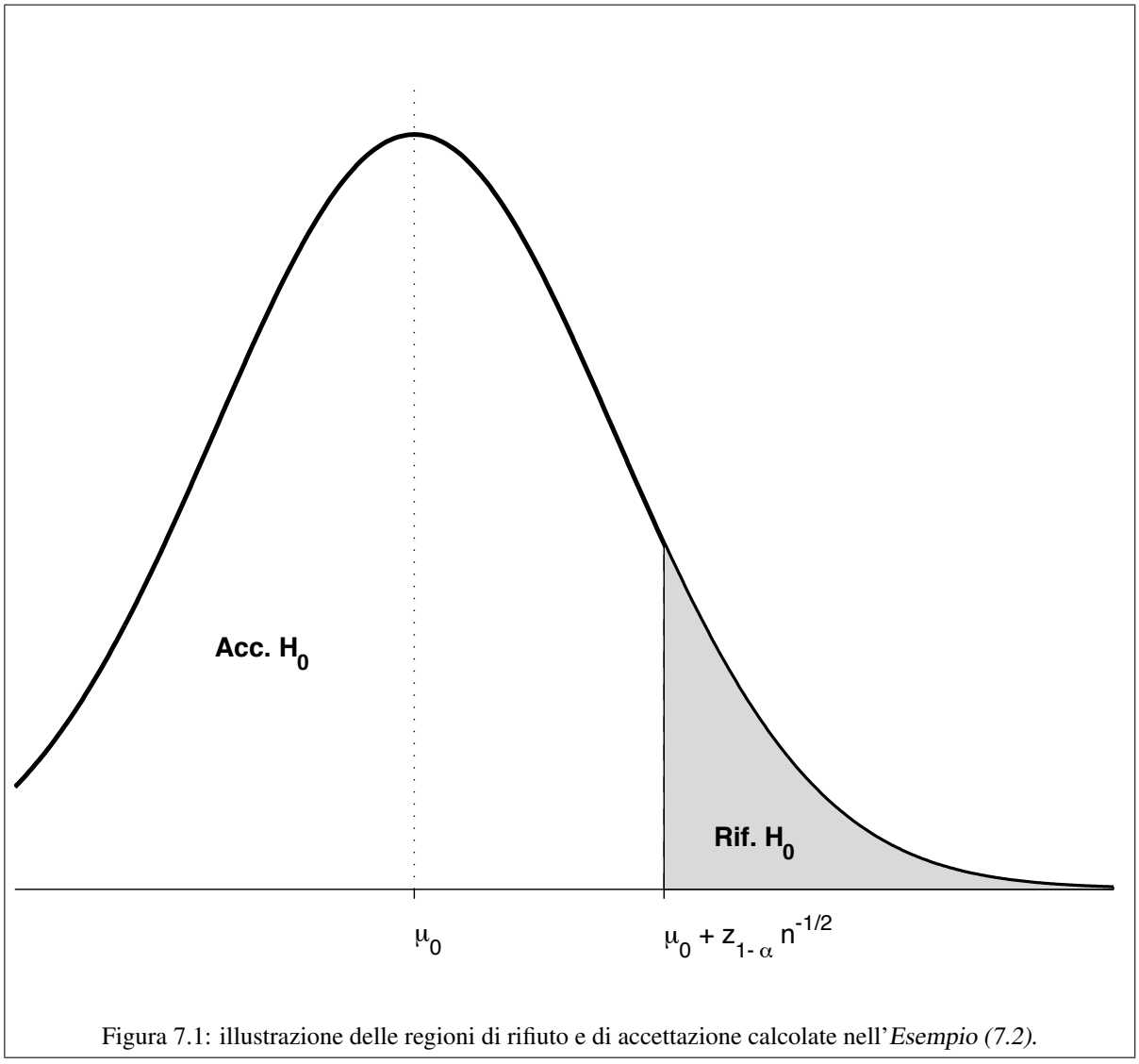

Figura 7.1: illustrazione delle regioni di rifiuto e di accettazione calcolate nell'*Esempio (7.2).*

probabilità di errore di TIPO I è ottenuta "riducendo" la regione critica C, rendendo così più "difficile" il rifiuto dell'ipotesi $\mathcal{H}_0$.

La *Figura (7.1)* mostra le regioni di rifiuto e di accettazione calcolate in precedenza; la curva corrisponde alla densità della v.a. $\overline{X}_n \sim \mathtt{N}(\mu_0, 1/n)$. Si noti che l'area (ombreggiata) soprastante la regione di rifiuto corrisponde ad una probabilità pari a α.

Presa $\mu = \mu_1$ la potenza del test τ vale

$$\begin{aligned}
\boldsymbol{\beta}_\tau(\mu_1) &= \mathbf{P}_{\mathcal{H}_1}\{\text{rifiutare } \mathcal{H}_0\} \\
&= \mathbf{P}_{\mathcal{H}_1}\left\{\overline{X}_n > \mu_0 + \frac{z_{1-\alpha}}{\sqrt{n}}\right\} \\
&= \mathbf{P}_{\mathcal{H}_1}\left\{\frac{\overline{X}_n - \mu_1}{1/\sqrt{n}} > (\mu_0 - \mu_1)\sqrt{n} + z_{1-\alpha}\right\} \\
&= \mathbf{P}\big\{Z > z_{1-\alpha} - \sqrt{n}(\mu_1 - \mu_0)\big\}\,,
\end{aligned}$$

dove $Z \sim \mathtt{N}(0,1)$. In particolare, essendo $\mu_1 > \mu_0$, si ha $\boldsymbol{\beta}_\tau(\mu_1) > \alpha$. Inoltre

$$\mathbf{P}_{\mathcal{H}_1}\{\text{accettare } \mathcal{H}_0\} = 1 - \boldsymbol{\beta}_\tau(\mu_1) = \mathbf{P}\big\{Z \leq z_{1-\alpha} - \sqrt{n}(\mu_1 - \mu_0)\big\}\,,$$

che fornisce la probabilità di errore di TIPO II. Per quanto già osservato, al diminuire di α il quantile $z_{1-\alpha}$ aumenta e, di conseguenza, aumenta la probabilità di errore di TIPO II. In altri termini, non è possibile ridurre contemporaneamente sia la probabilità di errore di TIPO I sia quella di TIPO II, che rappresentano due esigenze conflittuali. ◀

ESEMPIO 7.3. ▶

Si consideri il campione $\mathbf{X}^5_{\mathtt{Be}(p)}$, dove $p \in \Theta = (0,1)$, e si formulino le ipotesi

$$\mathcal{H}_0 \colon X \sim f_{\Theta_0} \sim \mathtt{Be}(1/2) \quad \text{contro} \quad \mathcal{H}_1 \colon X \sim f_{\Theta_1} \sim \mathtt{Be}(p)\,, \text{ con } p \neq 1/2.$$

L'ipotesi $\mathcal{H}_0$ è semplice mentre $\mathcal{H}_1$ è composta. Intuitivamente si accetta $\mathcal{H}_0$ se $\overline{X}_5$ (che è uno stimatore della speranza di X e quindi di p) è "sufficientemente" vicina a $\frac{1}{2}$; in altri termini si accetta $\mathcal{H}_0$ se $\left|\overline{X}_5 - \frac{1}{2}\right| \leq r \in \mathbb{R}_+$, dove la costante r può essere determinata sulla base del livello del test che, ad esempio, fissiamo pari a $\alpha = 0.1$. Dunque è necessario risolvere l'equazione in r data da

$$0.1 = \mathbf{P}_{\mathcal{H}_0}\{\text{rifiutare } \mathcal{H}_0\} = \mathbf{P}_{\mathcal{H}_0}\left\{\left|\overline{X}_5 - \frac{1}{2}\right| > r\right\}$$

o, equivalentemente,

$$0.9 = \mathbf{P}_{\mathcal{H}_0}\left\{\left|\overline{X}_5 - \frac{1}{2}\right| \leq r\right\} = \mathbf{P}_{\mathcal{H}_0}\left\{-5r \leq U - \frac{5}{2} \leq 5r\right\} = \mathbf{P}_{\mathcal{H}_0}\left\{\frac{5-10r}{2} \leq U \leq \frac{5+10r}{2}\right\},$$

dove $U = \sum_{i=1}^5 X_i$. Essendo (sotto $\mathcal{H}_0$) $U \sim \mathtt{Bi}\left(5, \frac{1}{2}\right)$, la densità di U vale

$$\mathbf{P}_{\mathcal{H}_0}\{U = u\} = \binom{5}{u}\left(\frac{1}{2}\right)^u\left(1-\frac{1}{2}\right)^{5-u}, \quad u = 0,1,2,3,4,5,$$

come mostrato nella tabella che segue.

$\boldsymbol{u = \sum_{i=1}^5 x_i}$	**0**	**1**	**2**	**3**	**4**	**5**
$\mathbf{P_{\mathcal{H}_0}\{U = u\}}$	0.03125	0.15625	0.31250	0.31250	0.15625	0.03125

Dunque non è possibile scegliere r in modo tale che il vincolo sia soddisfatto esattamente. Ad esempio, l'ipotesi $\mathcal{H}_0$ può essere rifiutata se si osserva $U \in \{0,5\}$ (o $\overline{X}_5 \in \{0,1\}$), ma in questo caso l'ampiezza del test sarebbe

$$\alpha' = 0.03125 + 0.03125 = 0.0625 < 0.1.$$

Parimenti, l'ipotesi $\mathcal{H}_0$ può essere rifiutata se si osserva $U \in \{0,1,4,5\}$ (o $\overline{X}_5 \in \left\{0, \frac{1}{5}, \frac{4}{5}, 1\right\}$), ma in questo caso l'ampiezza del test sarebbe

$$\alpha'' = 0.03125 + 0.15625 + 0.15625 + 0.03125 = 0.375 > 0.1.$$

Al fine di ottenere un test di ampiezza esattamente pari a $\alpha = 0.1$ si ricorre ad un *test aleatorio*: anziché rifiutare (o accettare) $\mathcal{H}_0$ con probabilità 1 quando $U \in \{1,4\}$, si rifiuta $\mathcal{H}_0$ con probabilità $\gamma \in (0,1)$, cioè

$$\tau(\boldsymbol{x}) = \begin{cases} 1, & \text{se } u \in \{0,5\} \\ \gamma, & \text{se } u \in \{1,4\} \\ 0, & \text{altrimenti} \end{cases},$$

dove γ è soluzione dell'equazione

$$\alpha = 0.1 = \mathbf{P}_{\mathcal{H}_0}\{U \in \{0,5\}\} + \gamma\,\mathbf{P}_{\mathcal{H}_0}\{U \in \{1,4\}\} = \mathbf{E}_{\mathcal{H}_0}(\tau)\,.$$

Dunque $\gamma = \frac{0.0375}{0.3125} = 0.12$, e si verifica facilmente che $\mathbf{E}_{\mathcal{H}_0}(\tau) = 0.1$, cioè che τ è un test aleatorio di ampiezza esattamente pari a $\alpha = 0.1$. La potenza del test τ è data da, per $p \neq \frac{1}{2}$,

$$\boldsymbol{\beta}_\tau(p) = \mathbf{P}_p\{U \in \{0,5\}\} + \gamma\,\mathbf{P}_p\{U \in \{1,4\}\}\,,$$

che può essere calcolata esplicitamente. La *Figura (7.2)* mostra sia la funzione di potenza β_τ sia la probabilità di errore di TIPO II al variare del parametro $p \in (0,1)$ — ovviamente è da escludere il punto $p = \frac{1}{2}$.

◀

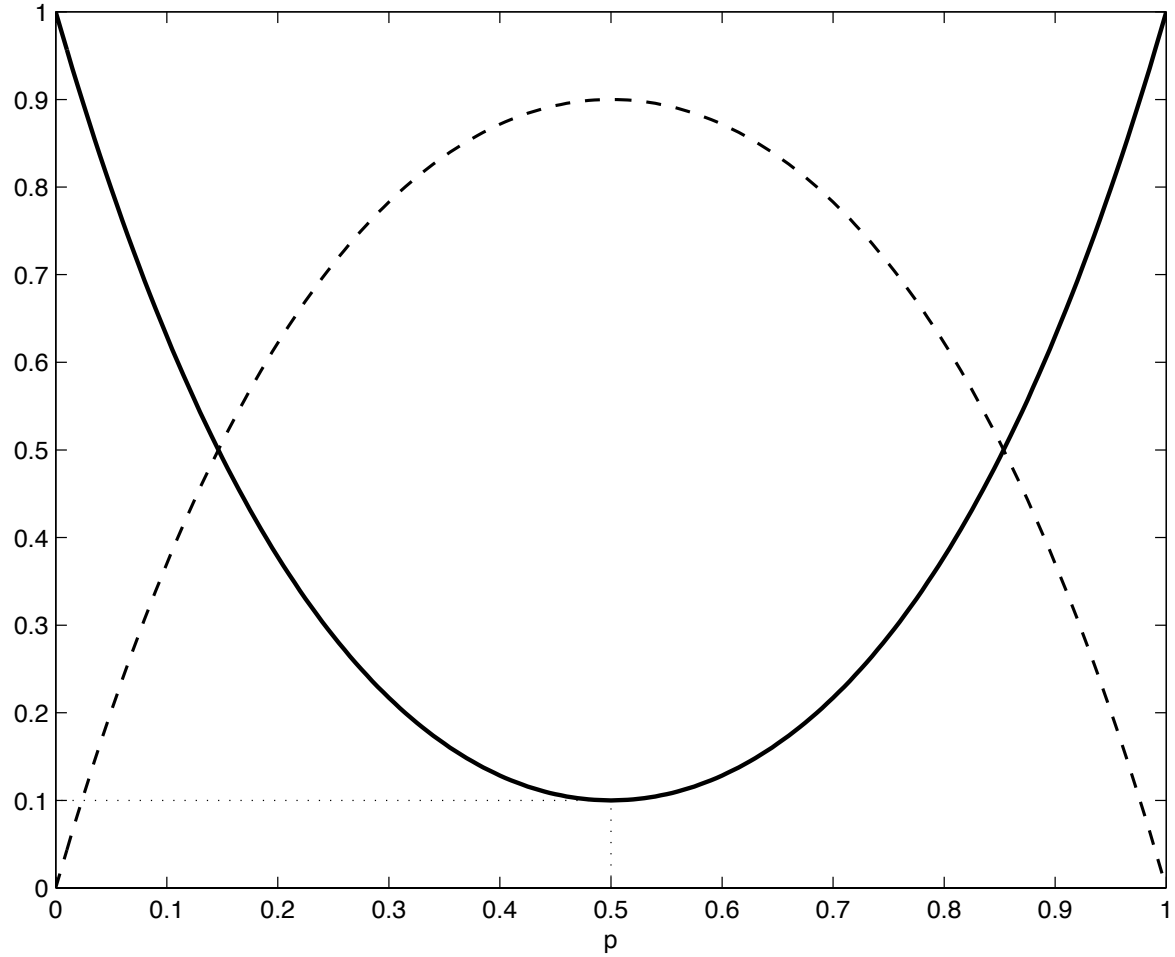

Figura 7.2: grafico della funzione di potenza β_τ (linea *continua*) e della probabilità di errore di TIPO II (linea *tratteggiata*) dell'*Esempio (7.3).*

7.1 Il Lemma di Neyman-Pearson

In questa Sezione si dimostra il *Lemma di Neyman-Pearson* (1933), che fornisce un metodo per costruire un test UPP in presenza di ipotesi $\mathcal{H}_0, \mathcal{H}_1$ *semplici*. Il modello statistico soggiacente è $\mathfrak{M} = \left(\Omega, \mathcal{F}, \{\mathbf{P}_\theta\}_{\theta=\theta_0,\theta_1}\right)$, che è sicuramente dominato (ad esempio dalla misura $\mu = \mathbf{P}_{\theta_0} + \mathbf{P}_{\theta_1}$). Dunque le funzioni

$$f_{\theta_0} = \frac{\mathrm{d}\mathbf{P}_{\theta_0}}{\mathrm{d}\mu} \quad \text{e} \quad f_{\theta_1} = \frac{\mathrm{d}\mathbf{P}_{\theta_1}}{\mathrm{d}\mu}$$

rappresentano versioni della densità. Per semplicità nel seguito useremo la notazione "$0, 1$" al posto di "θ_0, θ_1".

LEMMA 7.1 (Neyman-Pearson)**.** *Sia $r > 0$. Allora l'insieme $C = \{f_1 > r f_0\}$ è la regione critica di un test τ di livello (ampiezza) $\alpha = \mathbf{P}_0\{C\}$ dell'ipotesi $\mathcal{H}_0 \colon \theta = \theta_0$ contro l'alternativa $\mathcal{H}_1 \colon \theta = \theta_1$ che risulta essere sia ammissibile sia* UPP *in $\mathfrak{T}_\alpha$.*

DIM. Sia C^* un'altra regione critica; vale allora la disuguaglianza, per ogni $\omega \in \Omega$,

$$(\mathbb{I}_{C^*}(\omega) - \mathbb{I}_C(\omega))\,(f_1(\omega) - r f_0(\omega)) \leq 0. \tag{7.9}$$

Infatti se $\omega \in C^* \cap C$, oppure se $\omega \notin C^*$ e $\omega \notin C$, è nullo il primo membro dell'*equazione (7.9)*, sicché questa vale. Se, invece, $\omega \in C^*$ ma $\omega \notin C$, oppure se $\omega \in C$ ma $\omega \notin C^*$, allora, per definizione di C, l'*equazione (7.9)*

vale. Integrando rispetto a μ si ottiene

$$\mathbf{P}_1\{C^*\} - \mathbf{P}_1\{C\} \leq r\,(\mathbf{P}_0\{C^*\} - \mathbf{P}_0\{C\})\,. \tag{7.10}$$

Tale disuguaglianza implica che il test τ di regione critica C sia *ammissibile*, cioè non esiste alcun test τ^* di regione critica C^* tale che valgano le due disuguaglianze

$$\mathbf{P}_0\{C^*\} \leq \mathbf{P}_0\{C\}\,, \tag{7.11a}$$

$$\mathbf{P}_1\{C^*\} \geq \mathbf{P}_1\{C\}\,, \tag{7.11b}$$

e che almeno una delle due sia stretta: infatti, se fosse stretta la prima allora l'*equazione (7.10)* darebbe $\mathbf{P}_1\{C^*\} < \mathbf{P}_1\{C\}$, violando l'*equazione (7.11b)*; parimenti, se fosse stretta la seconda allora l'*equazione (7.10)* darebbe $\mathbf{P}_0\{C^*\} > \mathbf{P}_0\{C\}$ violando l'*equazione (7.11a)*.

Si consideri ora su $\mathfrak{T}_\alpha$ un test τ^* di regione critica C^* di livello α, tale cioè che $\mathbf{P}_0\{C^*\} \leq \alpha = \mathbf{P}_0\{C\}$. Il secondo membro dell'*equazione (7.10)* è allora negativo o nullo, sicché è tale anche il primo, cioè $\mathbf{P}_1\{C^*\} \leq \mathbf{P}_1\{C\}$: ciò significa che il test τ di regione critica C è più potente del test τ^* di regione critica C^*. □

OSSERVAZIONE 7.2.

Il *Lemma di Neyman-Pearson* continua a valere anche se

1. $r = 0$, e in tal caso $C = \{f_1 > 0\}$,
2. $r = \infty$, e in tal caso $C = \{f_0 = 0\} \cap \{f_1 > 0\}$,

dove si è usata la convenzione $0 \cdot \infty = 0$.

ESEMPIO 7.4. ►

Si consideri il problema trattato nell'*Esempio (7.2)*. Il *Lemma di Neyman-Pearson* suggerisce di procedere alla verifica delle ipotesi utilizzando il test

$$\tau(\boldsymbol{x}) = \begin{cases} 1, & \text{se } \boldsymbol{\varrho}(\boldsymbol{x}) > r \\ 0, & \text{se } \boldsymbol{\varrho}(\boldsymbol{x}) < r \end{cases},$$

con $\boldsymbol{x} \in \mathbb{R}^n$, dove

$$\boldsymbol{\varrho}(\boldsymbol{x}) = \frac{f_1(\boldsymbol{x})}{f_0(\boldsymbol{x})} = \frac{e^{-\frac{1}{2}\sum_{i=1}^n (x_i-\mu_1)^2}}{e^{-\frac{1}{2}\sum_{i=1}^n (x_i-\mu_0)^2}}$$

e la costante r è soluzione dell'equazione $\mathbf{E}_0(\tau(\boldsymbol{X})) = \alpha$. Poiché

$$\boldsymbol{\varrho}(\boldsymbol{x}) = e^{(\mu_1-\mu_0)\sum_{i=1}^n x_i + \frac{n}{2}(\mu_0^2-\mu_1^2)}$$

si ha $\boldsymbol{\varrho}(\boldsymbol{x}) > r$ se, e solo se, $\sum_{i=1}^n x_i > r'$. La costante r' è soluzione dell'equazione

$$\mathbf{P}_{\mathcal{H}_0}\{\text{rifiutare } \mathcal{H}_0\} = \mathbf{P}_0\left\{\sum_{i=1}^n X_i > r'\right\} = \mathbf{P}_0\left\{\frac{\sum_{i=1}^n X_i - n\mu_0}{\sqrt{n}} > \frac{r' - n\mu_0}{\sqrt{n}}\right\} = \alpha,$$

e dunque $r' = n\mu_0 + z_{1-\alpha}\sqrt{n}$, una soluzione equivalente a quella dell'*Esempio (7.2)*. Perciò il test τ così costruito risulta essere sia ammissibile sia UPP in $\mathfrak{T}_\alpha$. ◄

Come illustrato nell'*Esempio (7.3)*, talora può essere necessario utilizzare un test aleatorio per ottenere esattamente il livello α desiderato. Infatti, tranne nel caso in cui la legge del modello soggiacente sia continua o, equivalentemente, risulti $\mathbf{P}_0\{f_1 = tf_0\} = 0$ per ogni $t > 0$, può non esistere una regione critica del tipo $C = \{f_1 > rf_0\}$ che risolva esattamente l'equazione $\mathbf{P}_0\{C\} = \alpha$. A tale proposito i *test di Neyman-Pearson* giocano un ruolo importante nella verifica di ipotesi semplici (e non solo).

DEFINIZIONE 7.15 (Test di Neyman-Pearson)**.** Si dice *test di Neyman-Pearson* un test τ per il quale esiste una costante $r \in (0, \infty)$ tale che, preso $\omega \in \Omega$,

$$\tau(\omega) = 1 \quad \text{se } f_1(\omega) > rf_0(\omega), \tag{7.12a}$$

$$\tau(\omega) = 0 \quad \text{se } f_1(\omega) < rf_0(\omega). \tag{7.12b}$$

Evidentemente i test dell'*Esempio (7.2)* e dell'*Esempio (7.4)* sono test di Neyman-Pearson. Il Teorema che segue è fondamentale nella verifica di ipotesi.

TEOREMA 7.1 (Neyman-Pearson)**.** *Valgono le seguenti affermazioni.*

1. *Ogni test di Neyman-Pearson $\tau \in \mathfrak{T}_\alpha$ è* UPP *tra i test di livello $0 < \alpha < 1$.*
2. *Per ogni $0 < \alpha < 1$ si possono determinare due costanti r e γ tali che il test*
$$\tau = \mathbb{I}_{\{f_1 > r f_0\}} + \gamma\, \mathbb{I}_{\{f_1 = r f_0\}} \tag{7.13}$$
abbia ampiezza pari a α.
3. *Ogni test ammissibile è un test di Neyman-Pearson.*

DIM. Dimostriamo separatamente le singole affermazioni.

1. Sia τ un test di Neyman-Pearson e sia r la costante della *Definizione (7.15)*. Se τ' è un altro test vale allora la disuguaglianza, per ogni $\omega \in \Omega$,
$$(\tau'(\omega) - \tau(\omega))\,(f_1(\omega) - r f_0(\omega)) \leq 0. \tag{7.14}$$
Infatti se ω è tale che $f_1(\omega) > r f_0(\omega)$ allora si ha, dall'*equazione (7.12a)*, $\tau(\omega) = 1$, onde il secondo fattore a primo membro dell'*equazione (7.14)* è strettamente positivo mentre il primo è negativo. Se, invece, $f_1(\omega) < r f_0(\omega)$ allora, dall'*equazione (7.12b)*, $\tau(\omega) = 0$, onde il secondo fattore a primo membro dell'*equazione (7.14)* è strettamente negativo mentre il primo è positivo. Infine l'*equazione (7.14)* è banalmente verificata se $f_1(\omega) = r f_0(\omega)$. Si scriva ora tale equazione nella forma
$$\tau'(\omega) f_1(\omega) - \tau(\omega) f_1(\omega) \leq r\ (\tau'(\omega) f_0(\omega) - \tau(\omega) f_0(\omega))\,.$$
Integrando rispetto a μ si ottiene
$$\mathbf{E}_1(\tau') - \mathbf{E}_1(\tau) \leq r\ (\mathbf{E}_0(\tau') - \mathbf{E}_0(\tau))\,. \tag{7.15}$$
Di qui si procede esattamente come nel *Lemma di Neyman-Pearson*.

2. Si ponga
$$r = \inf\{t > 0 \colon \mathbf{P}_0\{f_1 > t f_0\} \leq \alpha\}\,.$$
Se si ha $\alpha = \mathbf{P}_0\{f_1 > r f_0\}$ (cioè se la funzione $s \mapsto \mathbf{P}_0\{f_1 > s f_0\}$ è continua nel punto $s = r$) allora si scelga $\gamma = 0$; in questo caso l'*equazione (7.13)* fornisce $\alpha = \mathbf{E}_0(\tau)$. Se, invece, è $\alpha > \mathbf{P}_0\{f_1 > r f_0\}$ allora la funzione $s \mapsto \mathbf{P}_0\{f_1 > s f_0\}$ presenta una discontinuità nel punto $s = r$, che corrisponde ad un salto di ampiezza pari a $\mathbf{P}_0\{f_1 = r f_0\} > 0$; si scelga allora
$$\gamma = \frac{\alpha - \mathbf{P}_0\{f_1 > r f_0\}}{\mathbf{P}_0\{f_1 = r f_0\}},$$
così che l'*equazione (7.13)* dia, con tale scelta di r e γ,
$$\mathbf{E}_0(\tau) = \mathbf{P}_0\{f_1 > r f_0\} + \gamma\, \mathbf{P}_0\{f_1 = r f_0\} = \alpha.$$

3. Sia τ un test di Neyman-Pearson e sia τ' un test ammissibile. Come al punto *(1)* si dimostra che, per ogni $\omega \in \Omega$, $(\tau'(\omega) - \tau(\omega))\,(f_1(\omega) - r f_0(\omega)) \leq 0$; da qui, integrando rispetto a μ, si riottiene $\mathbf{E}_1(\tau') - \mathbf{E}_1(\tau) \leq r\ (\mathbf{E}_0(\tau') - \mathbf{E}_0(\tau))$, cioè l'*equazione (7.15)*. Poiché τ' è ammissibile si ha
$$\mathbf{E}_0(\tau') \leq \mathbf{E}_0(\tau)\,, \tag{7.16a}$$
$$\mathbf{E}_1(\tau') \geq \mathbf{E}_1(\tau)\,. \tag{7.16b}$$
L'*equazione (7.15)* dimostra che le *equazioni (7.16)* sono possibili solo se entrambe sono uguaglianze. Dunque $\mathbf{E}_0(\tau') = \mathbf{E}_0(\tau)$ e $\mathbf{E}_1(\tau') = \mathbf{E}_1(\tau)$. Perciò nell'*equazione (7.15)* vale il segno di uguaglianza e quindi vale (μ-q.o.) l'uguaglianza
$$(\tau'(\omega) - \tau(\omega))\,(f_1(\omega) - r f_0(\omega)) = 0. \tag{7.17}$$
Se $\omega \in \{f_1 > r f_0\}$ tale equazione fornisce $\tau' = \tau = 1$ in quanto τ è un test di Neyman-Pearson. Analogamente, se $\omega \in \{f_1 < r f_0\}$ tale equazione fornisce $\tau' = \tau = 0$. In definitiva anche τ' è un test di Neyman-Pearson.

Le affermazioni sono dunque dimostrate. □

OSSERVAZIONE 7.3 (Test corretto)**.**

In un test di Neyman-Pearson τ si ha

$$\mathbf{E}_1(\tau) \geq \mathbf{E}_0(\tau)\,. \tag{7.18}$$

Si ponga infatti $\alpha = \mathbf{E}_0(\tau)$ e si consideri il test di funzione $\tau' = \alpha$ (costante). Poiché $\mathbf{E}_0(\tau') = \mathbf{E}_1(\tau') = \alpha$, grazie alla disuguaglianza dell'*equazione (7.15)* si ha $\mathbf{E}_1(\tau) \geq \mathbf{E}_1(\tau') = \alpha = \mathbf{E}_0(\tau)$. Un test per cui valga la disuguaglianza dell'*equazione (7.18)* si dice *corretto* — si veda la *Definizione (7.19)*.

In termini pratici, siano $0 < \alpha < 1$ e $\boldsymbol{\varrho}(\boldsymbol{x}) = \frac{f_1(\boldsymbol{x})}{f_0(\boldsymbol{x})}$, dove $\boldsymbol{x} \in \mathbb{R}^n$. Si consideri il test

$$\tau(\boldsymbol{x}) = \begin{cases} 1, & \text{se } \boldsymbol{\varrho}(\boldsymbol{x}) > r \\ \gamma, & \text{se } \boldsymbol{\varrho}(\boldsymbol{x}) = r \\ 0, & \text{se } \boldsymbol{\varrho}(\boldsymbol{x}) < r \end{cases}, \tag{7.19}$$

dove $0 \leq \gamma \leq 1$; allora è possibile ottenere $\mathbf{E}_0(\tau(\boldsymbol{X})) = \alpha$. Infatti, essendo $\mathbf{P}_0\{f_0(\boldsymbol{X}) = 0\} = 0$, si ha

$$\begin{aligned} \mathbf{E}_0(\tau(\boldsymbol{X})) &= \mathbf{P}_0\{\boldsymbol{\varrho}(\boldsymbol{X}) > r\} + \gamma\, \mathbf{P}_0\{\boldsymbol{\varrho}(\boldsymbol{X}) = r\} \\ &= 1 - \mathbf{P}_0\{\boldsymbol{\varrho}(\boldsymbol{X}) \leq r\} + \gamma\, \mathbf{P}_0\{\boldsymbol{\varrho}(\boldsymbol{X}) = r\}\,. \end{aligned}$$

Si tratta ora di determinare r e γ affinché sia $\mathbf{E}_0(\tau(\boldsymbol{X})) = \alpha$, cioè

$$\mathbf{P}_0\{\boldsymbol{\varrho}(\boldsymbol{X}) \leq r\} - \gamma\, \mathbf{P}_0\{\boldsymbol{\varrho}(\boldsymbol{X}) = r\} = 1 - \alpha.$$

Si noti che $F_0(r) = \mathbf{P}_0\{\boldsymbol{\varrho}(\boldsymbol{X}) \leq r\}$ è in effetti una *funzione di ripartizione*. Se esiste r_0 tale che $F_0(r_0) = 1 - \alpha$ allora si sceglie $r = r_0$ e $\gamma = 0$. Altrimenti, per la continuità a destra di F_0, esiste r_0 tale che

$$F_0^-(r_0) = \mathbf{P}_0\{\boldsymbol{\varrho}(\boldsymbol{X}) < r_0\} \leq 1 - \alpha < \mathbf{P}_0\{\boldsymbol{\varrho}(\boldsymbol{X}) \leq r_0\} = F_0(r_0),$$

ovvero la funzione F_0 ha un salto in corrispondenza di r_0. In questo caso si sceglie $r = r_0$ e

$$\gamma = \frac{\mathbf{P}_0\{\boldsymbol{\varrho}(\boldsymbol{X}) \leq r_0\} - (1 - \alpha)}{\mathbf{P}_0\{\boldsymbol{\varrho}(\boldsymbol{X}) = r_0\}} = \frac{\alpha - \mathbf{P}_0\{\boldsymbol{\varrho}(\boldsymbol{X}) > r_0\}}{\mathbf{P}_0\{\boldsymbol{\varrho}(\boldsymbol{X}) = r_0\}}.$$

Evidentemente un test τ siffatto è di livello α.

NOTA 7.7.

La funzione $\boldsymbol{\varrho}$ appena introdotta rappresenta un rapporto di densità (o, equivalentemente, di verosimiglianze). Intuitivamente si è portati a rifiutare $\mathcal{H}_0$ se $f_1(\boldsymbol{x}) > r\, f_0(\boldsymbol{x})$ per un opportuno valore r, cioè se l'ipotesi $\mathcal{H}_1$ è più "verosimile"; il test τ appena costruito agisce in tale maniera.

ESEMPIO 7.5. ▶

Si consideri il campione $\mathbf{X}^5_{\mathtt{Be}(p)}$, dove $p \in \Theta = \left\{\frac{1}{2}, \frac{3}{4}\right\}$, e si formulino le ipotesi

$$\mathcal{H}_0 \colon X \sim f_{\Theta_0} \sim \mathtt{Be}(1/2) \quad \text{contro} \quad \mathcal{H}_1 \colon X \sim f_{\Theta_1} \sim \mathtt{Be}(3/4)\,.$$

A differenza di quanto visto nell'*Esempio (7.3)* sia $\mathcal{H}_0$ sia $\mathcal{H}_1$ sono ora ipotesi semplici; dunque è possibile applicare il *Teorema (7.1)*; come nell'*Esempio (7.3)* useremo la statistica $U = \sum_{i=1}^5 X_i$, avente legge $U \sim \mathtt{Bi}(5, p)$, con $p = p_0 = \frac{1}{2}$ o $p = p_1 = \frac{3}{4}$. Nella tabella seguente sono riportati i valori delle variabili di interesse (adottando una notazione semplificata).

$u = \sum_{i=1}^{5} x_i$	0	1	2	3	4	5
$f_0 = f_{p=p_0}$	$\frac{1}{32}$	$\frac{5}{32}$	$\frac{10}{32}$	$\frac{10}{32}$	$\frac{5}{32}$	$\frac{1}{32}$
$f_1 = f_{p=p_1}$	$\frac{1}{1024}$	$\frac{5\cdot 3}{1024}$	$\frac{10\cdot 3^2}{1024}$	$\frac{10\cdot 3^3}{1024}$	$\frac{5\cdot 3^4}{1024}$	$\frac{3^5}{1024}$
$\varrho = f_1/f_0$	$\frac{1}{32}$	$\frac{3}{32}$	$\frac{3^2}{32}$	$\frac{3^3}{32}$	$\frac{3^4}{32}$	$\frac{3^5}{32}$
r	$\frac{1}{32}$	$\frac{3}{32}$	$\frac{9}{32}$	$\frac{27}{32}$	$\frac{81}{32}$	$\frac{243}{32}$
$\mathbf{P}_0\{\varrho < r\}$	0	$\frac{1}{32}$	$\frac{6}{32}$	$\frac{16}{32}$	$\frac{26}{32}$	$\frac{31}{32}$
$\mathbf{P}_0\{\varrho = r\}$	$\frac{1}{32}$	$\frac{5}{32}$	$\frac{10}{32}$	$\frac{10}{32}$	$\frac{5}{32}$	$\frac{1}{32}$
$\mathbf{P}_0\{\varrho \leq r\}$	$\frac{1}{32}$	$\frac{6}{32}$	$\frac{16}{32}$	$\frac{26}{32}$	$\frac{31}{32}$	1

Come nell'*Esempio (7.3)*, scegliamo $\alpha = 0.1$; si tratta ora di determinare r e γ affinché sia $\mathbf{E}_0(\tau) = \alpha$, dove τ è dato dall'*equazione (7.13)*, cioè

$$\mathbf{P}_0\{\varrho \leq r\} - \gamma\, \mathbf{P}_0\{\varrho = r\} = 1 - \alpha.$$

Usando il *Teorema (7.1)* cerchiamo anzitutto di calcolare r tramite la relazione

$$r = \inf\{t > 0 \colon \mathbf{P}_0\{\varrho > t\} \leq \alpha\}.$$

La *Figura (7.3)* mostra l'andamento della funzione $\mathbf{P}_0\{\varrho > t\}$ al variare di t: evidentemente non esiste alcun valore $r = r_0$ tale che sia esattamente $\mathbf{P}_0\{\varrho > r_0\} = \alpha$.

Tuttavia, preso $r = r_0 = \frac{81}{32}$, si ha

$$\mathbf{P}_0\{\varrho < r_0\} = \frac{26}{32} = 0.8125 \leq 1 - \alpha = 0.9 < \mathbf{P}_0\{\varrho \leq r_0\} = \frac{31}{32} = 0.96875$$

Dunque, prendendo

$$\gamma = \frac{\mathbf{P}_0\{\varrho \leq r_0\} - (1-\alpha)}{\mathbf{P}_0\{\varrho = r_0\}} = \frac{22}{50} = 0.44$$

è possibile ottenere un test di ampiezza esattamente pari a α.

Si noti come, utilizzando un test τ' con $\alpha' = 0.05$, si ottenga $r' = r_0 = \frac{81}{32}$ e $\gamma' = 0.12$: ciò significa che ora, nel caso in cui si osservi $U = 4$, si rifiuterà l'ipotesi $\mathcal{H}_0$ con probabilità $\gamma' < \gamma$, ovvero la si accetterà con maggiore facilità. Analogamente, utilizzando un test τ'' con $\alpha'' = 0.01$, si ottiene $r'' = \frac{243}{32}$ e $\gamma'' = 0.32$: dunque, nel caso in cui si osservi $U = 5$, si rifiuterà l'ipotesi $\mathcal{H}_0$ con probabilità γ''. Evidentemente, considerando i diversi test τ, τ', τ'' proposti, la probabilità di commettere un errore di TIPO I diminuisce ($\alpha > \alpha' > \alpha''$), ma contemporaneamente aumenta quella di commettere un errore di TIPO II: infatti, la potenza del test τ è data da

$$\boldsymbol{\beta}_\tau(p_1) = \mathbf{E}_{p_1}(\tau) = \mathbf{P}_{p_1}\{\varrho > r_0\} + \gamma\, \mathbf{P}_{p_1}\{\varrho = r_0\} = \frac{3^5}{1024} + \frac{22}{50}\,\frac{5\cdot 3^4}{1024} \approx 0.41 \geq \mathbf{E}_{p_0}(\tau) = 0.1.$$

Invece, considerando il test τ', si ha $\boldsymbol{\beta}_{\tau'}(p_1) \approx 0.29 \geq \mathbf{E}_0(\tau')$, che è meno potente del test τ. Analogamente, considerando il test τ'' si ha $\boldsymbol{\beta}_{\tau''}(p_1) \approx 0.08 \geq \mathbf{E}_0(\tau'')$, che è meno potente del test τ'. Si noti il fatto che tutti i test sono *corretti*. ◀

OSSERVAZIONE 7.4.

Nell'*Esempio (7.5)*, comunque si scelgano p_0 e p_1, con $0 < p_0 < p_1 < 1$, si ha che il valore di γ non dipende da p_1. Ne segue che il test appena descritto risulta essere UPP nella verifica delle ipotesi $\mathcal{H}_0 \colon p = p_0$ contro $\mathcal{H}_1 \colon p > p_0$; evidentemente $\mathcal{H}_1$ non è un'ipotesi *semplice*, ma *composta* — si veda l'*Osservazione (7.6)*.

7.2 Rapporto di verosimiglianza monotono

Nelle applicazioni raramente si presentano situazioni in cui siano semplici tanto l'ipotesi nulla quanto quella alternativa. Tuttavia la teoria di Neyman-Pearson risulta comunque interessante per i riflessi che ha su alcuni particolari tipi di test definiti di seguito.

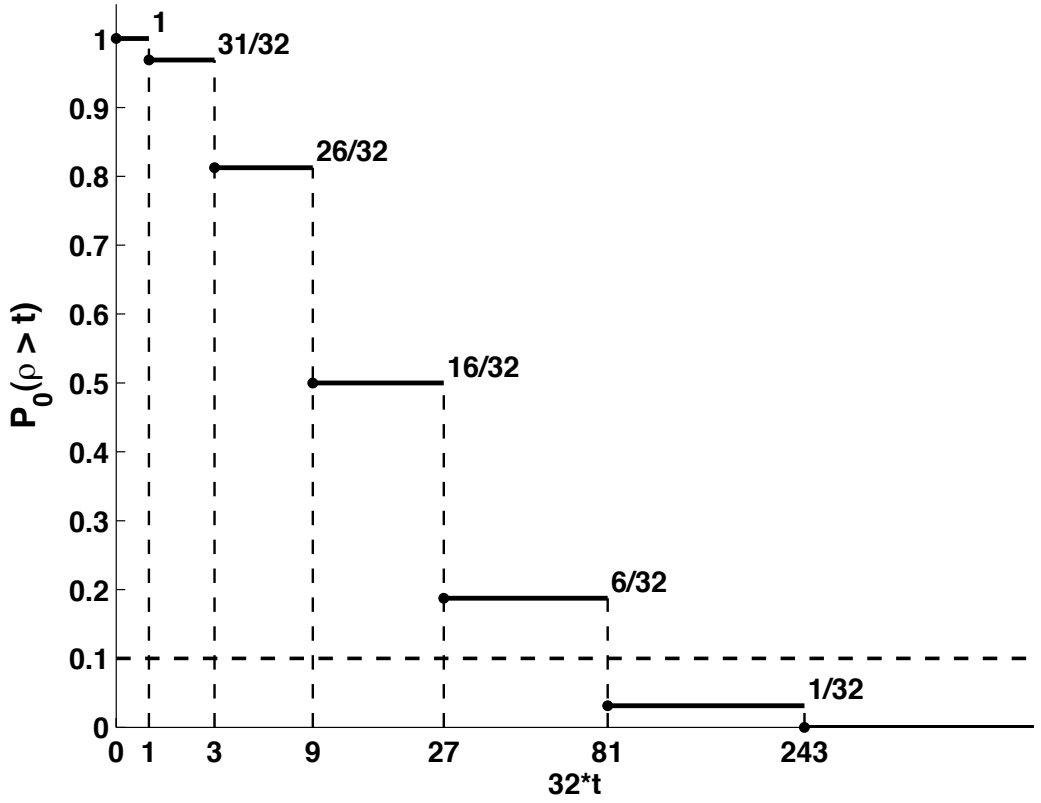

Figura 7.3: grafico della funzione $\mathbf{P}_0\{\varrho > t\}$ dell'*Esempio (7.5)* al variare di t; la scala dell'asse orizzontale è stata alterata onde consentire una migliore lettura.

DEFINIZIONE 7.16 (Test unilatero)**.** Sia Θ un intervallo di $\mathbb{R}$, limitato o no, e sia $\theta' \in \mathring{\Theta}$. Si dice *unilatero* ogni test dell'ipotesi $\mathcal{H}_0 \colon \theta \leq \theta'$ contro l'alternativa $\mathcal{H}_1 \colon \theta > \theta'$ o dell'ipotesi $\mathcal{H}_0 \colon \theta \geq \theta'$ contro l'alternativa $\mathcal{H}_1 \colon \theta < \theta'$. Parimenti si dicono *unilatere* le ipotesi formulate in un test unilatero.

In generale non è possibile costruire un test UPP per questo problema di verifica di ipotesi. Ci limiteremo allora a considerare una classe particolare di modelli statistici (sufficientemente ampia da includere la famiglia esponenziale trattata nella *Sezione 1.2*) per la quale dimostreremo che un test UPP esiste per ipotesi unilatere.

DEFINIZIONE 7.17 (Rapporto di verosimiglianza monotono)**.** Un modello statistico dominato $\mathfrak{M}$ si dice avere *rapporto di verosimiglianza monotono* (RVM) se esiste una v.a. T tale che, per ogni coppia $\theta_1, \theta_2 \in \Theta \subseteq \mathbb{R}$, $\theta_1 < \theta_2$, esista una funzione monotona $\boldsymbol{\varrho}_{\theta_1,\theta_2} \colon \mathbb{R} \to \overline{\mathbb{R}}_+$ tale che

$$\frac{\mathbf{L}(\theta_2)}{\mathbf{L}(\theta_1)} = \boldsymbol{\varrho}_{\theta_1,\theta_2} \circ T \quad \mu\text{-q.o.}, \tag{7.20}$$

dove μ è la misura dominante. Il RVM è *crescente* o *decrescente* se $\boldsymbol{\varrho}_{\theta_1,\theta_2}$ è tale.

Evidentemente ci limitiamo ad uno spazio dei parametri $\Theta \subseteq \mathbb{R}$ unidimensionale ($d = 1$).

ESEMPIO 7.6 (Modello esponenziale)**.** ▶ Si consideri un modello esponenziale canonico di verosimiglianza $\mathbf{L}_\theta = e^{\theta T - \chi(\theta)}$ (si veda la *Sezione 1.2*). Per $\theta_1 < \theta_2$ il rapporto di verosimiglianza è

$$\frac{\mathbf{L}(\theta_2)}{\mathbf{L}(\theta_1)} = e^{(\theta_2 - \theta_1)T}\, e^{-(\chi(\theta_2) - \chi(\theta_1))}.$$

Dunque basta prendere

$$\boldsymbol{\varrho}_{\theta_1,\theta_2}(t) = e^{(\theta_2-\theta_1)t}\, e^{-(\chi(\theta_2)-\chi(\theta_1))}$$

per verificare che il modello è a RVM crescente. ◀

OSSERVAZIONE 7.5.

Per quanto visto nell'*Osservazione (1.3)*, se $\Omega = \mathbb{R}^n$ e se la legge del campione è nella forma

$$f_\theta(x) = s(x)\, e^{p(\theta)q(x)+r(\theta)}$$

rispetto alla misura dominante (rispettivamente, la "misura del contare" μ_c o la misura di Lebesgue λ) allora il campione è di classe esponenziale. La verosimiglianza del campione $\boldsymbol{X}$ vale, per $\boldsymbol{x} \in \mathbb{R}^n$,

$$\mathbf{L}_\theta(\boldsymbol{x}) = \left(\prod_{i=1}^n s(x_i)\right) e^{p(\theta)\sum_{i=1}^n q(x_i)+nr(\theta)}.$$

Dunque, per $\theta_1 < \theta_2$ il rapporto di verosimiglianza vale

$$\boldsymbol{\varrho}_{\theta_1,\theta_2}(\boldsymbol{x}) = e^{(p(\theta_2)-p(\theta_1))\sum_{i=1}^n q(x_i)}\, e^{n(r(\theta_2)-r(\theta_1))}.$$

Ponendo $T = \sum_{i=1}^n q \circ X_i$ è facile verificare che il modello è a RVM crescente (decrescente) se, e solo se, la funzione $\theta \mapsto p(\theta)$ è crescente (decrescente). Naturalmente, attraverso un'opportuna riparametrizzazione, è possibile ricondursi ad un modello esponenziale canonico a RVM crescente.

ESEMPIO 7.7. ►

Per quanto visto nell'*Esempio (1.9)* il modello del campione $\mathbf{X}^n_{\mathfrak{U}_c(0,\theta)}$, $\theta > 0$, non è esponenziale. La verosimiglianza del campione vale, per $\boldsymbol{x} \in \mathbb{R}^n$,

$$\mathbf{L}_\theta(\boldsymbol{x}) = \frac{1}{\theta^n} \prod_{i=1}^n \mathbb{I}_{(0,\theta)}(x_i) = \frac{1}{\theta^n}\, \mathbb{I}_{(0,\theta)}(x_{(n)})$$

rispetto alla misura di Lebesgue λ. Posto $T = X_{(n)}$ si ha, per $\theta_1 < \theta_2$,

$$\boldsymbol{\varrho}_{\theta_1,\theta_2}(\boldsymbol{X}) = \left(\frac{\theta_1}{\theta_2}\right)^n \frac{\mathbb{I}_{(0,\theta_2)}(T)}{\mathbb{I}_{(0,\theta_1)}(T)},$$

sicché il rapporto di verosimiglianza $\boldsymbol{\varrho}_{\theta_1,\theta_2}$

- non è definito se $T \leq 0$ o $T \geq \theta_2$ (tuttavia gli insiemi $\{T \leq 0\}$ e $\{T \geq \theta_2\}$ hanno probabilità nulla sia rispetto a $\mathbf{P}_{\theta_1}$ sia rispetto a $\mathbf{P}_{\theta_2}$);
- vale $\left(\frac{\theta_1}{\theta_2}\right)^n$ se $0 < T < \theta_1$;
- vale $+\infty$ se $\theta_1 \leq T < \theta_2$.

Si tratta perciò di un modello a RVM crescente. ◀

ESEMPIO 7.8. ►

Si consideri il campione $\mathbf{X}^1_{\mathcal{C}(1,\theta)}$, $\theta \in \mathbb{R}$. La verosimiglianza del campione vale, per $x \in \mathbb{R}$,

$$\mathbf{L}_\theta(x) = \frac{1}{\pi}\, \frac{1}{1+(x-\theta)^2}$$

rispetto alla misura di Lebesgue λ, sicché il rapporto di verosimiglianza $\boldsymbol{\varrho}_{\theta_1,\theta_2}$ vale, per $\theta_1 < \theta_2$,

$$\boldsymbol{\varrho}_{\theta_1,\theta_2}(x) = \frac{1+(x-\theta_1)^2}{1+(x-\theta_2)^2} \xrightarrow{x\to\pm\infty} 1.$$

Dunque $\boldsymbol{\varrho}_{\theta_1,\theta_2}$ *non* soddisfà alle condizioni della *Definizione (7.17)*. ◀

La teoria dei test unilateri si basa sul seguente risultato che scende dal *Teorema di Neyman-Pearson* — per la notazione si veda il *Teorema (7.1)*.

PROPOSIZIONE 7.1. *Sia $\mathfrak{M}$ un modello statistico dominato con* RVM *crescente. Allora il test*

$$\tau = \mathbb{I}_{\{T>r\}} + \gamma\, \mathbb{I}_{\{T=r\}}, \tag{7.21}$$

con $r \in \mathbb{R}$ e $\gamma \in [0,1]$, è, per ogni coppia $\theta_1, \theta_2 \in \Theta$, $\theta_1 < \theta_2$, un test di Neyman-Pearson dell'ipotesi $\mathcal{H}_0 \colon \theta = \theta_1$ contro l'alternativa $\mathcal{H}_1 \colon \theta = \theta_2$. Lo stesso vale usando il test

$$\tau = \mathbb{I}_{\{T<r\}} + \gamma\, \mathbb{I}_{\{T=r\}} \tag{7.22}$$

nel caso in cui il RVM *sia decrescente.*

DIM. Si ponga $s = \boldsymbol{\varrho}_{\theta_1,\theta_2}(r)$. Se ω è tale per cui $T(\omega) \geq r$ allora $\left(\boldsymbol{\varrho}_{\theta_1,\theta_2} \circ T\right)(\omega) \geq s$; dunque

$$\left\{\left(\boldsymbol{\varrho}_{\theta_1,\theta_2} \circ T\right) < s\right\} \subseteq \{T < r\},$$

sicché $\tau(\omega) = 0$ se $\left(\boldsymbol{\varrho}_{\theta_1,\theta_2} \circ T\right)(\omega) < s$. Analogamente, se ω è tale per cui $T(\omega) \leq r$ allora $\left(\boldsymbol{\varrho}_{\theta_1,\theta_2} \circ T\right)(\omega) \leq s$; dunque

$$\left\{\left(\boldsymbol{\varrho}_{\theta_1,\theta_2} \circ T\right) > s\right\} \subseteq \{T > r\},$$

sicché $\tau(\omega) = 1$ se $\left(\boldsymbol{\varrho}_{\theta_1,\theta_2} \circ T\right)(\omega) > s$. Si noti che può accadere che sia $s = 0$ oppure $s = +\infty$. □

Il risultato che segue è importante nella verifica di ipotesi — si veda a tale proposito l'*Osservazione (7.3)*.

PROPOSIZIONE 7.2. *Se il rapporto di verosimiglianza è monotono (rispettivamente,* crescente *o* decrescente*) allora è* crescente *la funzione $\theta \mapsto \mathbf{E}_\theta(\tau) = \boldsymbol{\beta}_\tau(\theta)$, cioè*

$$\mathbf{E}_{\theta_1}(\tau) \leq \mathbf{E}_{\theta_2}(\tau) \tag{7.23}$$

*per $\theta_1 < \theta_2$, dove il test τ è, rispettivamente, quello dell'*equazione *(7.21) o quello dell'*equazione *(7.22).*

Il Teorema che segue è fondamentale considerando ipotesi composte.

TEOREMA 7.2. *Sia $\mathfrak{M}$ un modello statistico dominato con* RVM *crescente. Allora, fissato $\theta' \in \overset{\circ}{\Theta}$ e presi $r \in \mathbb{R}$ e $\gamma \in [0,1]$, il test*

$$\tau = \mathbb{I}_{\{T>r\}} + \gamma\, \mathbb{I}_{\{T=r\}} \tag{7.24}$$

è un test dell'ipotesi $\mathcal{H}_0 \colon \theta \leq \theta'$ contro l'alternativa $\mathcal{H}_1 \colon \theta > \theta'$ che risulta essere

1. *di ampiezza $\alpha = \mathbf{E}_{\theta'}(\tau)$;*
2. *ammissibile;*
3. UPP *nella classe $\mathfrak{T}_\alpha$.*

Inoltre, fissato $\alpha \in (0,1)$, si possono determinare r e γ in maniera tale che sia $\mathbf{E}_{\theta'}(\tau) = \alpha$.

Lo stesso vale usando il test

$$\tau = \mathbb{I}_{\{T<r\}} + \gamma\, \mathbb{I}_{\{T=r\}} \tag{7.25}$$

nel caso in cui il RVM *sia decrescente.*

*Volendo invece costruire un test dell'ipotesi $\mathcal{H}_0 \colon \theta \geq \theta'$ contro l'alternativa $\mathcal{H}_1 \colon \theta < \theta'$, $\theta' \in \overset{\circ}{\Theta}$, si userà il test dell'*equazione *(7.25) se il* RVM *è crescente e quello dell'*equazione *(7.24) se il* RVM *è decrescente.*

DIM. Dimostriamo separatamente le singole affermazioni.

1. In base alla *Proposizione (7.2)* sappiamo che la funzione $\theta \mapsto \mathbf{E}_\theta(\tau)$ è crescente; dunque l'ampiezza del test è pari a $\sup_{\theta \leq \theta'} \mathbf{E}_\theta(\tau) = \mathbf{E}_{\theta'}(\tau)$.

2. Si supponga che τ non sia ammissibile; esiste allora un altro test τ' migliore di τ. Valgono dunque le disuguaglianze

$$\text{per ogni } \theta \leq \theta' \quad \mathbf{E}_\theta(\tau') \leq \mathbf{E}_\theta(\tau)\,, \tag{7.26a}$$

$$\text{per ogni } \theta > \theta' \quad \mathbf{E}_\theta(\tau') \geq \mathbf{E}_\theta(\tau)\,, \tag{7.26b}$$

ove almeno una delle due è stretta per almeno un valore $\theta = \theta''$. Si osservi che, per quasi ogni $\omega \in \Omega$ e per ogni coppia $\theta_1, \theta_2 \in \Theta$, con $\theta_1 \leq \theta' < \theta_2$, vale la disuguaglianza

$$(\tau'(\omega) - \tau(\omega))\,(\mathbf{L}(\theta_2, \omega) - s\mathbf{L}(\theta_1, \omega)) \leq 0, \tag{7.27}$$

dove si è posto $s = \boldsymbol{\varrho}_{\theta_1,\theta_2}(r)$. Infatti, procedendo come nella *Proposizione (7.1)*, nulla vi è da dimostrare se $\mathbf{L}(\theta_2, \omega) = s\mathbf{L}(\theta_1, \omega)$. Se, invece, $\mathbf{L}(\theta_2, \omega) < s\mathbf{L}(\theta_1, \omega)$ allora $\omega \in \{T < r\}$ e dunque $\tau(\omega) = 0$, sicché il primo fattore nell'*equazione (7.27)* è positivo o nullo. Se, infine, $\mathbf{L}(\theta_2, \omega) > s\mathbf{L}(\theta_1, \omega)$ allora $\omega \in \{T > r\}$ e dunque $\tau(\omega) = 1$, sicché il primo fattore nell'*equazione (7.27)* è negativo o nullo. Integrando rispetto a μ si ottiene

$$\mathbf{E}_{\theta_2}(\tau') - \mathbf{E}_{\theta_2}(\tau) \leq s\,(\mathbf{E}_{\theta_1}(\tau') - \mathbf{E}_{\theta_1}(\tau))\,. \tag{7.28}$$

Di qui si procede esattamente come nel *Lemma di Neyman-Pearson*.

3. Sia τ' un test di livello $\alpha = \mathbf{E}_{\theta'}(\tau)$, sicché $\mathbf{E}_{\theta'}(\tau') \leq \mathbf{E}_{\theta'}(\tau)$. Basta ora applicare l'*equazione (7.28)* con $\theta_1 = \theta'$ e $\theta_2 = \theta > \theta'$ per trovare $\mathbf{E}_\theta(\tau') \leq \mathbf{E}_\theta(\tau)$.

Le affermazioni sono dunque dimostrate. □

OSSERVAZIONE 7.6.

Le conclusioni del *Teorema (7.2)* rimangono valide se si modifica l'ipotesi *composta* $\mathcal{H}_0\colon \theta \leq \theta'$ (o $\mathcal{H}_0\colon \theta \geq \theta'$) nell'ipotesi *semplice* $\mathcal{H}_0\colon \theta = \theta'$ — si veda l'*Osservazione (7.4)*.

ESEMPIO 7.9. ►

Si consideri il campione $\mathbf{X}^5_{\text{Be}(p)}$, dove $p \in \Theta = (0, 1)$. Dato $p' = \frac{1}{2}$ si formulino le ipotesi

$$\mathcal{H}_0\colon p \leq p' \quad \text{contro} \quad \mathcal{H}_1\colon p > p'.$$

A differenza di quanto visto nell'*Esempio (7.5)* sia $\mathcal{H}_0$ sia $\mathcal{H}_1$ sono ora ipotesi composte. Poiché il modello soggiacente è esponenziale (si veda l'*Esempio (1.5)*) è possibile applicare il *Teorema (7.2)*. Presi $0 < p_1 < p_2 < 1$ è facile verificare che il modello è a RVM crescente nella statistica $T = \sum_{i=1}^5 X_i$; dunque useremo il test $\tau = \mathbb{I}_{\{T>r\}} + \gamma\,\mathbb{I}_{\{T=r\}}$ dell'*equazione (7.24)*. Si tratta ora di determinare r e γ affinché sia $\mathbf{E}_{p'}(\tau) = \alpha$, cioè

$$\mathbf{P}_{p'}\{T > r\} + \gamma\,\mathbf{P}_{p'}\{T = r\} = \alpha.$$

Sia $\alpha = 0.1$. Per quanto già visto nell'*Esempio (7.5)*, preso $r = 4$ si ha

$$\mathbf{P}_{p'}\{T > 4\} + \gamma\,\mathbf{P}_{p'}\{T = 4\} = 0.1 \quad \text{ovvero} \quad \frac{1}{32} + \gamma\,\frac{5}{32} = \frac{1}{10},$$

e dunque $\gamma = \frac{22}{50} = 0.44$, un risultato che non sorprende in base all'*Osservazione (7.4)* e alla *Osservazione (7.6)*. Risultati analoghi a quelli dell'*Esempio (7.5)* si ottengono utilizzando livelli $\alpha' = 0.05$ e $\alpha'' = 0.01$; inoltre valgono le considerazioni già espresse in quella sede.

La potenza del test τ è data da, per $p > p'$,

$$\boldsymbol{\beta}_\tau(p) = \mathbf{E}_p(\tau) = \mathbf{P}_p\{T > 4\} + \gamma\,\mathbf{P}_p\{T = 4\} = (1 - 5\,\gamma)\,p^5 + 5\,\gamma\,p^4 \geq \mathbf{E}_{p'}(\tau) = \alpha,$$

che è una funzione crescente nella variabile p, come sancito dalla *Proposizione (7.2)*. Ancora una volta si ottengono risultati analoghi a quelli dell'*Esempio (7.5)* utilizzando livelli $\alpha' = 0.05$ e $\alpha'' = 0.01$, e inoltre valgono le considerazioni già espresse in quella sede. La *Figura (7.4)* mostra sia la funzione di potenza β_τ sia la probabilità di errore di TIPO II al variare del parametro $p \in (0, 1)$.

◄

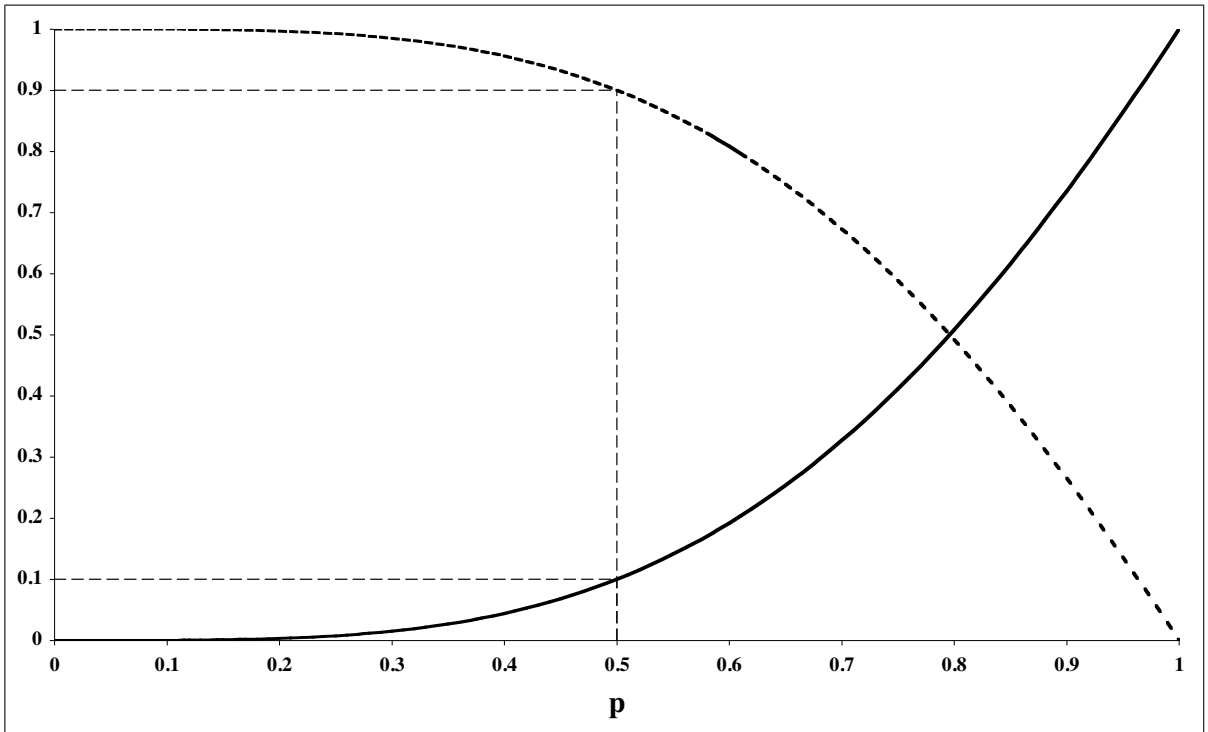

Figura 7.4: grafico della funzione di potenza β_τ (linea *continua*) e della probabilità di errore di TIPO II (linea *tratteggiata*) dell'*Esempio (7.9)*.

NOTA 7.8.

In tutti gli Esempi ove si è usato il campione $\mathbf{X}^5_{\mathrm{Be}(p)}$ (*Esempio (7.3)*, *Esempio (7.5)*, *Esempio (7.9)*) gli errori di TIPO II possono essere non trascurabili: essenzialmente ciò è dovuto all'ampiezza assai ridotta $(n = 5)$ del campione osservato.

ESEMPIO 7.10. ►

Si consideri il campione $\mathbf{X}^n_{\mathrm{e}(\theta)}$, $\theta > 0$. Per costruire un test dell'ipotesi $\mathcal{H}_0 \colon \mathbf{E}_\theta(X) \geq 1$ contro l'alternativa $\mathcal{H}_1 \colon \mathbf{E}_\theta(X) < 1$ (ovvero dell'ipotesi $\mathcal{H}_0 \colon \theta \leq 1$ contro l'alternativa $\mathcal{H}_1 \colon \theta > 1$) si può procedere nel seguente modo. La verosimiglianza del campione vale, per $\boldsymbol{x} \in \mathbb{R}^n_+$,

$$\mathbf{L}_\theta(\boldsymbol{x}) = \theta^n \, e^{-\theta \sum_{i=1}^n x_i},$$

sicché, per ogni coppia $0 < \theta_1 < \theta_2$, il rapporto di verosimiglianza $\boldsymbol{\varrho}_{\theta_1,\theta_2}$ vale

$$\boldsymbol{\varrho}_{\theta_1,\theta_2}(\boldsymbol{x}) = \left(\frac{\theta_2}{\theta_1}\right)^n e^{-(\theta_2-\theta_1)\sum_{i=1}^n x_i}.$$

Ponendo $T = \sum_{i=1}^n X_i$ si ha

$$\boldsymbol{\varrho}_{\theta_1,\theta_2}(t) = \left(\frac{\theta_2}{\theta_1}\right)^n e^{-(\theta_2-\theta_1)t},$$

e dunque si è in presenza di un modello a RVM decrescente. Perciò la regione critica di un "buon" test è $\{T < r\}$ (la legge Esponenziale è assolutamente continua), dove la costante r è da determinare in base al livello del test medesimo. Poiché il RVM è decrescente l'ampiezza di tale test vale

$$\sup_{\theta \leq 1} \mathbf{P}_\theta\{T < r\} = \mathbf{P}_{\theta=1}\{T < r\}\,.$$

Ora, quando $\theta = 1$ si ha $\sum_{i=1}^{n} X_i = T \sim \Gamma(n, 1)$ e dunque, presa $r > 0$,

$$\mathbf{P}_{\theta=1}\{T \leq r\} = \mathbf{P}_{\theta=1}\left\{\sum_{i=1}^{n} X_i \leq r\right\} = 1 - e^{-r} \sum_{i=0}^{n-1} \frac{r^i}{i!}.$$

Dato un livello $0 < \alpha < 1$, per calcolare la costante r occorre risolvere l'equazione

$$e^{-r} \sum_{i=0}^{n-1} \frac{r^i}{i!} = 1 - \alpha$$

nella variabile r; in generale ciò richiede l'applicazione di un algoritmo numerico. ◀

7.3 Rapporto di verosimiglianza generalizzato

Accanto ai test unilateri visti nella Sezione precedente, nelle applicazioni si presentano spesso situazioni in cui sia necessario verificare che il parametro appartenga o meno ad un intervallo limitato. A questo proposito si introducono i *test bilateri*.

DEFINIZIONE 7.18 (Test bilatero)**.** Sia Θ un intervallo di $\mathbb{R}$, limitato o no, e siano $\theta_1, \theta_2 \in \overset{\circ}{\Theta}$, $\theta_1 \leq \theta_2$. Si dice *bilatero* ogni test dell'ipotesi $\mathcal{H}_0 \colon \theta \in [\theta_1, \theta_2]$ contro l'alternativa $\mathcal{H}_1 \colon \theta \notin [\theta_1, \theta_2]$. In particolare, se $\theta_0 = \theta_1 = \theta_2$ allora si vaglia l'ipotesi (semplice) $\mathcal{H}_0 \colon \theta = \theta_0$ contro l'alternativa (composta) $\mathcal{H}_1 \colon \theta \neq \theta_0$. Si dicono *bilatere* le ipotesi formulate in un test bilatero.

In generale non esiste un test UPP per ipotesi bilatere; si rende allora opportuno introdurre ipotesi restrittive. Anzitutto ci si limita a considerare modelli esponenziali, con la verosimiglianza scritta nella forma canonica data dall'*equazione (1.4)*. Si suppone inoltre che sia $\mu(T = \boldsymbol{x}) = 0$, ove T è la statistica del test, $\boldsymbol{x} \in \mathbb{R}^n$ e μ è la misura dominante. Infine si considerano i soli *test corretti*.

DEFINIZIONE 7.19 (Test corretto)**.** Sia $\{\Theta_0, \Theta_1\}$ una partizione di Θ. Un test $\tau \in \mathfrak{T}_\alpha$ dell'ipotesi $\mathcal{H}_0 \colon \theta \in \Theta_0$ contro l'alternativa $\mathcal{H}_1 \colon \theta \in \Theta_1$ si dice *corretto* (o *non distorto*) se si ha, per ogni $\theta \in \Theta_1$,

$$\boldsymbol{\beta}_\tau(\theta) \geq \alpha. \tag{7.29}$$

Nel seguito indicheremo con $\mathfrak{T}_{\alpha,c}$ la classe dei test corretti di livello α.

OSSERVAZIONE 7.7.

Si noti che l'*equazione (7.29)* esprime il fatto che la potenza del test τ è superiore alla sua ampiezza: infatti la formulazione data è riscrivibile come

$$\boldsymbol{\beta}_\tau(\theta) \leq \alpha \quad \text{per ogni } \theta \in \Theta_0, \tag{7.30a}$$

$$\boldsymbol{\beta}_\tau(\theta) \geq \alpha \quad \text{per ogni } \theta \in \Theta_1. \tag{7.30b}$$

Una simile proprietà è certamente ragionevole per un "buon" test: in pratica si chiede che la probabilità di rifiutare un'ipotesi $\mathcal{H}_0$ falsa sia maggiore della probabilità di rifiutare $\mathcal{H}_0$ quando questa sia vera. Evidentemente, in base all'*Osservazione (7.3)* si ha che i test di Neyman-Pearson sono corretti. Ovviamente vale la relazione $\mathfrak{T}_{\alpha,c} \subseteq \mathfrak{T}_\alpha$. Si osservi inoltre che sia Θ sia la partizione $\{\Theta_0, \Theta_1\}$ possono essere arbitrari.

La verifica di ipotesi bilatere può essere affrontata ricorrendo ad una forma particolare del rapporto di verosimiglianza, di cui si sono già discussi alcuni aspetti nella *Sezione 7.2*. Come vedremo nel seguito, grazie al *rapporto di verosimiglianza generalizzato* è possibile costruire test che siano UPP o UPP non distorti, con proprietà interessanti almeno per grandi campioni.

DEFINIZIONE 7.20 (Rapporto di verosimiglianza generalizzato)**.** Sia $\{\Theta_0, \Theta_1\}$ una partizione di Θ. Si dice *rapporto di verosimiglianza generalizzato* (RVG) la quantità

$$\boldsymbol{\varrho}_*(\omega) = \frac{\sup_{\theta \in \Theta_0} \mathbf{L}_\theta(\omega)}{\sup_{\theta \in \Theta} \mathbf{L}_\theta(\omega)}, \tag{7.31}$$

dove $\omega \in \Omega$. L'insieme $\{\omega \in \Omega \colon \sup_{\theta \in \Theta} \mathbf{L}_\theta(\omega) \neq 0\}$ è detto *dominio* di $\boldsymbol{\varrho}_*$.

OSSERVAZIONE 7.8.

Evidentemente $\boldsymbol{\varrho}_*(\boldsymbol{X})$ è una variabile aleatoria e $0 \leq \boldsymbol{\varrho}_*(\boldsymbol{X}) \leq 1$. Il numeratore dell'*equazione (7.31)* è detto *la migliore spiegazione del campione $\boldsymbol{X}$ che l'ipotesi $\mathcal{H}_0$ possa fornire*, mentre il denominatore è detto *la migliore spiegazione del campione $\boldsymbol{X}$* (da intendersi nel senso della *massima verosimiglianza*).

Intuitivamente si rifiuta $\mathcal{H}_0$ se esiste una spiegazione del campione $\boldsymbol{X}$ che sia migliore della migliore spiegazione fornita da $\mathcal{H}_0$; in altri termini si rifiuta $\mathcal{H}_0$ se il rapporto che definisce $\boldsymbol{\varrho}_*$ risulta essere "troppo" piccolo. A tale fine si introducono i cosiddetti *test del rapporto di verosimiglianza generalizzato*.

DEFINIZIONE 7.21 (Test del rapporto di verosimiglianza generalizzato)**.** Sia $\{\Theta_0, \Theta_1\}$ una partizione di Θ. Un test τ dell'ipotesi $\mathcal{H}_0 \colon \theta \in \Theta_0$ contro l'alternativa $\mathcal{H}_1 \colon \theta \in \Theta_1$ si dice *test del rapporto di verosimiglianza generalizzato* se è della forma

$$\tau = \mathbb{I}_{\{\boldsymbol{\varrho}_* < r\}}, \tag{7.32}$$

dove la costante r è calcolata imponendo che sia

$$\sup_{\theta \in \Theta_0} \mathbf{P}_\theta\{\boldsymbol{\varrho}_* < r\} = \sup_{\theta \in \Theta_0} \mathbf{E}_\theta(\tau) = \alpha. \tag{7.33}$$

NOTA 7.9.

Se la legge di $\boldsymbol{\varrho}_*(\boldsymbol{X})$ è *assolutamente continua* allora qualsiasi scelta dell'ampiezza α è possibile; se invece tale legge è *discreta* può non esistere un test τ di ampiezza arbitraria. Si noti che sia Θ sia la partizione $\{\Theta_0, \Theta_1\}$ possono essere arbitrari.

Vale il seguente risultato.

PROPOSIZIONE 7.3. *Un test del rapporto di verosimiglianza generalizzato e un test di Neyman-Pearson non aleatorio sono equivalenti nel caso in cui sia $\mathcal{H}_0$ sia $\mathcal{H}_1$ siano ipotesi semplici.*

Gli esempi che seguono sono paradigmatici e mostrano l'uso del rapporto di verosimiglianza generalizzato; si noti che i procedimenti illustrati forniscono le soluzioni proposte, rispettivamente, nella riga (C) della *Tabella (7.1)*, nella riga (C) della *Tabella (7.2)* e nella riga (C) della *Tabella (7.3)*.

ESEMPIO 7.11 (Test sulla varianza di un campione Gaussiano)**.** ▶

Si consideri il campione $\mathbf{X}^n_{\mathbb{N}(\mu,\sigma^2)}$, dove i parametri $\mu \in \mathbb{R}$ e $\sigma^2 > 0$ sono incogniti, e sia $\sigma_0^2 > 0$ una costante nota. Si formulino le ipotesi (composte)

$$\mathcal{H}_0 \colon \sigma^2 = \sigma_0^2 \quad \text{contro} \quad \mathcal{H}_1 \colon \sigma^2 \neq \sigma_0^2.$$

Evidentemente $\Theta_0 = \left\{\left(\mu, \sigma_0^2\right) : \mu \in \mathbb{R}\right\}$ e $\Theta = \left\{\left(\mu, \sigma^2\right) : \mu \in \mathbb{R}, \sigma^2 > 0\right\}$. Ricordando l'espressione della verosimiglianza di un campione Gaussiano (si veda l'*Esempio (5.10)*) e prendendo $\boldsymbol{\theta} = \left(\mu, \sigma^2\right)$ si ha

$$\sup_{\boldsymbol{\theta} \in \Theta_0} \mathbf{L}_{\boldsymbol{\theta}}(\boldsymbol{x}) = \sup_{\mu \in \mathbb{R}} \frac{1}{\left(2\pi\sigma_0^2\right)^{n/2}} e^{-\frac{1}{2\sigma_0^2} \sum_{i=1}^n (x_i - \mu)^2} = \frac{1}{\left(2\pi\right)^{n/2} \left(\sigma_0^2\right)^{n/2}} e^{-\frac{1}{2\sigma_0^2} \sum_{i=1}^n (x_i - \overline{x}_n)^2},$$

essendo $\overline{x}_n$ lo SMV di μ nel caso in cui $\sigma^2 = \sigma_0^2$. Analogamente, lo SMV di $\boldsymbol{\theta}$ quando sia μ sia σ^2 sono incogniti vale $\hat{\boldsymbol{\theta}} = \left(\overline{x}_n, \overline{\sigma}_n^2\right)$. Dunque si ha

$$\sup_{\boldsymbol{\theta} \in \Theta} \mathbf{L}_{\boldsymbol{\theta}}(\boldsymbol{x}) = \sup_{\boldsymbol{\theta} \in \Theta} \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2} = \frac{1}{\left(2\pi\right)^{n/2} \left(\overline{\sigma}_n^2\right)^{n/2}} e^{-\frac{n}{2}}.$$

Ne segue che $\boldsymbol{\varrho}_*$ vale

$$\boldsymbol{\varrho}_*(\boldsymbol{x}) = \left(\frac{\overline{\sigma}_n^2}{\sigma_0^2}\right)^{n/2} e^{\frac{n}{2} - \frac{1}{2\sigma_0^2} \sum_{i=1}^n (x_i - \overline{x}_n)^2} = \left(\frac{1}{n} \frac{\sum_{i=1}^n \left(x_i - \overline{x}_n\right)^2}{\sigma_0^2}\right)^{n/2} e^{\frac{n}{2} - \frac{1}{2} \sum_{i=1}^n (x_i - \overline{x}_n)^2 / \sigma_0^2}.$$

Un test di livello $0 < \alpha < 1$ si ottiene risolvendo l'equazione $\sup_{\boldsymbol{\theta} \in \Theta_0} \mathbf{P}_{\boldsymbol{\theta}}\{\boldsymbol{\varrho}_*(\boldsymbol{x}) < r\} = \alpha$ nella variabile r. Ponendo

$$s = \frac{\sum_{i=1}^n \left(x_i - \overline{x}_n\right)^2}{\sigma_0^2}$$

è possibile riscrivere $\boldsymbol{\varrho}_*$ come

$$\boldsymbol{\varrho}_*(\boldsymbol{x}) = \left(\frac{e}{n}\right)^{n/2} s^{n/2} e^{-s/2},$$

cioè una funzione unimodale in s. Dunque si rifiuterà l'ipotesi $\mathcal{H}_0$ se $s < s_1$ o $s > s_2$, con $s_1 \leq s_2$. Ora, in base ai risultati della *Sezione 6.1* è noto che, sotto l'ipotesi $\mathcal{H}_0 \colon \sigma^2 = \sigma_0^2$, la statistica

$$S = S(\boldsymbol{X}) = (n-1)\,\frac{\overline{S}_n^2}{\sigma_0^2}$$

ha legge $\chi^2(n-1)$ — si veda la *Proposizione (6.6)*. Di conseguenza è sufficiente porre

$$s_1 = \chi^2_{n-1,\alpha/2} \quad \text{e} \quad s_2 = \chi^2_{n-1,1-\alpha/2},$$

cioè i quantili di ordine, rispettivamente, $\frac{\alpha}{2}$ e $1-\frac{\alpha}{2}$ della legge $\chi^2(n-1)$, per ottenere un test bilatero di livello α delle ipotesi date. ◀

ESEMPIO 7.12 (Test sulla speranza di un campione Gaussiano)**.** ▶ Si consideri il campione $\mathbf{X}^n_{\mathbb{N}(\mu,\sigma^2)}$, dove i parametri $\mu \in \mathbb{R}$ e $\sigma^2 > 0$ sono incogniti, e sia $\mu_0 \in \mathbb{R}$ una costante nota. Si formulino le ipotesi (composte)

$$\mathcal{H}_0 \colon \mu = \mu_0 \quad \text{contro} \quad \mathcal{H}_1 \colon \mu \neq \mu_0.$$

Evidentemente $\Theta_0 = \left\{\left(\mu_0, \sigma^2\right) : \sigma^2 > 0\right\}$ e $\Theta = \left\{\left(\mu, \sigma^2\right) : \mu \in \mathbb{R}, \sigma^2 > 0\right\}$. Ricordando l'espressione della verosimiglianza di un campione Gaussiano (si veda l'*Esempio (5.10)*) e prendendo $\boldsymbol{\theta} = \left(\mu, \sigma^2\right)$ si ha

$$\sup_{\boldsymbol{\theta}\in\Theta_0} \mathbf{L}_{\boldsymbol{\theta}}(\boldsymbol{x}) = \sup_{\sigma^2>0} \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu_0)^2} = \frac{1}{\left(2\pi\right)^{n/2}\left(\hat{\sigma}_0^2\right)^{n/2}} e^{-\frac{n}{2}},$$

essendo $\hat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^n \left(x_i-\mu_0\right)^2$ lo SMV di σ^2 nel caso in cui $\mu = \mu_0$ (che è diverso da $\overline{\sigma}_n^2$). Analogamente, lo SMV di $\boldsymbol{\theta}$ quando sia μ sia σ^2 sono incogniti vale $\hat{\boldsymbol{\theta}} = \left(\overline{x}_n, \overline{\sigma}_n^2\right)$. Dunque si ha

$$\sup_{\boldsymbol{\theta}\in\Theta} \mathbf{L}_{\boldsymbol{\theta}}(\boldsymbol{x}) = \sup_{\boldsymbol{\theta}\in\Theta} \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2} = \frac{1}{\left(2\pi\right)^{n/2}\left(\overline{\sigma}_n^2\right)^{n/2}} e^{-\frac{n}{2}}.$$

Ne segue che $\boldsymbol{\varrho}_*$ vale

$$\boldsymbol{\varrho}_*(\boldsymbol{x}) = \left(\frac{\overline{\sigma}_n^2}{\hat{\sigma}_0^2}\right)^{n/2} = \left(\frac{\sum_{i=1}^n \left(x_i-\overline{x}_n\right)^2}{\sum_{i=1}^n \left(x_i-\mu_0\right)^2}\right)^{n/2} = \left(1 + \frac{n\,\left(\overline{x}_n-\mu_0\right)^2}{\sum_{i=1}^n \left(x_i-\overline{x}_n\right)^2}\right)^{-n/2}.$$

Un test di livello $0 < \alpha < 1$ si ottiene risolvendo l'equazione $\sup_{\boldsymbol{\theta}\in\Theta_0} \mathbf{P}_{\boldsymbol{\theta}}\{\boldsymbol{\varrho}_*(\boldsymbol{x}) < r\} = \alpha$ nella variabile r. Ponendo

$$t = \frac{\sqrt{n}\,\left(\overline{x}_n-\mu_0\right)}{\sqrt{\sum_{i=1}^n \left(x_i-\overline{x}_n\right)^2}} = \frac{\sqrt{n}\,\left(\overline{x}_n-\mu_0\right)}{\sqrt{(n-1)\,\overline{s}_n^2}}$$

è possibile riscrivere $\boldsymbol{\varrho}_*$ come

$$\boldsymbol{\varrho}_*(\boldsymbol{x}) = \left(\frac{1}{1+t^2}\right)^{n/2},$$

cioè una funzione decrescente in t^2. Dunque si rifiuterà l'ipotesi $\mathcal{H}_0$ se $|t| > t_0$. Ora, in base ai risultati della *Sezione 6.2* è noto che, sotto l'ipotesi $\mathcal{H}_0 \colon \mu = \mu_0$, la statistica

$$T = T(\boldsymbol{X}) = \frac{\sqrt{n}\,\left(\overline{X}_n-\mu_0\right)}{\overline{S}_n}$$

ha legge $\mathtt{t}(n-1)$ — si veda la *Proposizione (6.8)*. Di conseguenza è sufficiente porre

$$t_0 = t_{n-1,1-\alpha/2},$$

cioè il quantile di ordine $1-\frac{\alpha}{2}$ della legge $\mathtt{t}(n-1)$, per ottenere un test bilatero di livello α delle ipotesi date. ◀

ESEMPIO 7.13 (Test sulle varianze di campioni Gaussiani)**.** ►
Siano $\mathbf{X}^n_{\mathtt{N}(\mu_1,\sigma_1^2)}$ e $\mathbf{Y}^m_{\mathtt{N}(\mu_2,\sigma_2^2)}$ campioni indipendenti, dove i parametri $\mu_1, \mu_2 \in \mathbb{R}$ e $\sigma_1^2, \sigma_2^2 > 0$ sono incogniti, e si formulino le ipotesi (composte)

$$\mathcal{H}_0 \colon \sigma_1^2 = \sigma_2^2 \quad \text{contro} \quad \mathcal{H}_1 \colon \sigma_1^2 \neq \sigma_2^2.$$

Evidentemente

$$\Theta_0 = \left\{ \left(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2\right) : \mu_i \in \mathbb{R}, \sigma_i^2 > 0, i = 1, 2, \sigma_1^2 = \sigma_2^2 \right\},$$
$$\Theta = \left\{ \left(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2\right) : \mu_i \in \mathbb{R}, \sigma_i^2 > 0, i = 1, 2 \right\}.$$

Sia $\boldsymbol{\theta} = \left(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2\right)$. La verosimiglianza congiunta vale

$$\mathbf{L}_{\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{(2\pi)^{(n+m)/2}\, \sigma_1^n\, \sigma_2^m}\, e^{-\frac{1}{2\sigma_1^2}\sum_{i=1}^n (x_i-\mu_1)^2 - \frac{1}{2\sigma_2^2}\sum_{i=1}^m (y_i-\mu_2)^2}.$$

Usando le tecniche illustrate nella *Sezione 5.2* si ricava che

$$\left(\hat{\mu}_1, \hat{\sigma}_1^2, \hat{\mu}_2, \hat{\sigma}_2^2\right) = \left(\overline{x}_n, \overline{\sigma}^2_{X,n}, \overline{y}_m, \overline{\sigma}^2_{Y,m}\right)$$

è lo SMV di $\boldsymbol{\theta}$ quando tutti i parametri siano incogniti. Se, invece, $\sigma_1^2 = \sigma_2^2 = \sigma^2$, allora lo SMV di σ^2 è dato da

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n \left(x_i - \overline{x}_n\right)^2 + \sum_{i=1}^m \left(y_i - \overline{y}_m\right)^2}{n+m}.$$

Dunque

$$\sup_{\boldsymbol{\theta}\in\Theta} \mathbf{L}_{\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{y}) = \frac{e^{-(n+m)/2}}{\left(\frac{2\pi}{n}\right)^{n/2} \left(\sum_{i=1}^n \left(x_i - \overline{x}_n\right)^2\right)^{n/2} \left(\frac{2\pi}{m}\right)^{m/2} \left(\sum_{i=1}^m \left(y_i - \overline{y}_m\right)^2\right)^{m/2}},$$
$$\sup_{\boldsymbol{\theta}\in\Theta_0} \mathbf{L}_{\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{y}) = \frac{e^{-(n+m)/2}}{\left(\frac{2\pi}{n+m}\right)^{(n+m)/2} \left(\sum_{i=1}^n \left(x_i - \overline{x}_n\right)^2 + \sum_{i=1}^m \left(y_i - \overline{y}_m\right)^2\right)^{(n+m)/2}}.$$

Ne segue che $\boldsymbol{\varrho}_*$ vale

$$\boldsymbol{\varrho}_*(\boldsymbol{x}, \boldsymbol{y}) = \left(\frac{n}{n+m}\right)^{n/2} \left(\frac{m}{m+n}\right)^{m/2} \frac{\left(\sum_{i=1}^n \left(x_i - \overline{x}_n\right)^2\right)^{n/2} \left(\sum_{i=1}^m \left(y_i - \overline{y}_m\right)^2\right)^{m/2}}{\left(\sum_{i=1}^n \left(x_i - \overline{x}_n\right)^2 + \sum_{i=1}^m \left(y_i - \overline{y}_m\right)^2\right)^{(n+m)/2}}.$$

Un test di livello $0 < \alpha < 1$ si ottiene risolvendo l'equazione $\sup_{\boldsymbol{\theta}\in\Theta_0} \mathbf{P}_{\boldsymbol{\theta}}\{\boldsymbol{\varrho}_*(\boldsymbol{x}, \boldsymbol{y}) < r\} = \alpha$ nella variabile r. Essendo

$$\frac{\left(\sum_{i=1}^n \left(x_i - \overline{x}_n\right)^2\right)^{\frac{n}{2}} \left(\sum_{i=1}^m \left(y_i - \overline{y}_m\right)^2\right)^{\frac{m}{2}}}{\left(\sum_{i=1}^n \left(x_i - \overline{x}_n\right)^2 + \sum_{i=1}^m \left(y_i - \overline{y}_m\right)^2\right)^{\frac{n+m}{2}}} = \frac{1}{\left(1 + \frac{\sum_{i=1}^n (x_i-\overline{x}_n)^2}{\sum_{i=1}^m (y_i-\overline{y}_m)^2}\right)^{\frac{n}{2}} \left(1 + \frac{\sum_{i=1}^m (y_i-\overline{y}_m)^2}{\sum_{i=1}^n (x_i-\overline{x}_n)^2}\right)^{\frac{m}{2}}}$$

si ha che, ponendo

$$f = \frac{m-1}{n-1} \frac{\sum_{i=1}^n \left(x_i - \overline{x}_n\right)^2}{\sum_{i=1}^m \left(y_i - \overline{y}_m\right)^2},$$

è possibile riscrivere $\boldsymbol{\varrho}_*$ come

$$\boldsymbol{\varrho}_*(\boldsymbol{x}, \boldsymbol{y}) = \left(\frac{n}{n+m}\right)^{n/2} \left(1 + \frac{m-1}{n-1}\frac{1}{f}\right)^{-n/2} \left(\frac{m}{m+n}\right)^{m/2} \left(1 + \frac{n-1}{m-1} f\right)^{-m/2}.$$

Lo studio della funzione $\boldsymbol{\varrho}_*$ mostra che la disuguaglianza $\boldsymbol{\varrho}_*(\boldsymbol{x}, \boldsymbol{y}) < r$ è equivalente alle disuguaglianze $f < r_1$ o $f > r_2$, con $r_1 \leq r_2$, che forniscono i criteri di rifiuto dell'ipotesi $\mathcal{H}_0$. Ora, in base ai risultati della *Sezione 6.3* è noto che, sotto l'ipotesi $\mathcal{H}_0 \colon \sigma_1^2 = \sigma_2^2$, la statistica

$$F = F(\boldsymbol{X}, \boldsymbol{Y}) = \frac{m-1}{n-1} \frac{\sum_{i=1}^n \left(X_i - \overline{X}_n\right)^2}{\sum_{i=1}^m \left(Y_i - \overline{Y}_m\right)^2} = \frac{\overline{S}^2_{n,X}}{\overline{S}^2_{m,Y}}$$

ha legge $\mathtt{F}(n-1, m-1)$ — si veda la *Proposizione (6.10)*. Di conseguenza è sufficiente porre

$$r_1 = f_{n-1,m-1,\alpha/2} \quad \text{e} \quad r_2 = f_{n-1,m-1,1-\alpha/2},$$

cioè i quantili di ordine, rispettivamente, $\frac{\alpha}{2}$ e $1-\frac{\alpha}{2}$ della legge $\mathtt{F}(n-1, m-1)$, per ottenere un test bilatero di livello α delle ipotesi date. ◀

7.4 Verifica di ipotesi per campioni Gaussiani

Come già visto nel *Capitolo 6*, i campioni Gaussiani giocano un ruolo di primo piano nell'ambito della Statistica Matematica, in particolare per quanto concerne la verifica di ipotesi. Tale verifica si basa sulle tecniche mostrate nelle Sezioni precedenti e sarà illustrata fornendo le regioni critiche dei test proposti. In particolare, i test del rapporto di verosimiglianza generalizzato saranno fondamentali e risulteranno essere UPP o UPP non distorti.

7.4.1 Test del Chi-quadro (Varianza)

I test del *Chi-quadro* riguardano essenzialmente la verifica di ipotesi sulla varianza di un campione Gaussiano $\mathbf{X}^n_{\mathtt{N}(\mu,\sigma^2)}$. Sia $\sigma_0 \in \mathbb{R}_+$ nota. Nella *Tabella (7.1)* si mostrano le regioni di rifiuto di diverse ipotesi $\mathcal{H}_0$ per test di livello $0 < \alpha < 1$; qui $\chi^2_{n,p}$ indica il quantile di ordine p della legge $\chi^2(n)$.

	Ipotesi		**Test / Regione di rifiuto**	
	$\mathcal{H}_0$	$\mathcal{H}_1$	**μ nota**	**μ non nota**
(A)	$\sigma \le \sigma_0$	$\sigma > \sigma_0$	$\sum_{i=1}^n (X_i-\mu)^2 \ge \sigma_0^2\, \chi^2_{n,1-\alpha}$	$\overline{S}^2_n \ge \frac{\sigma_0^2}{n-1}\, \chi^2_{n-1,1-\alpha}$
(B)	$\sigma \ge \sigma_0$	$\sigma < \sigma_0$	$\sum_{i=1}^n (X_i-\mu)^2 \le \sigma_0^2\, \chi^2_{n,\alpha}$	$\overline{S}^2_n \le \frac{\sigma_0^2}{n-1}\, \chi^2_{n-1,\alpha}$
(C)	$\sigma = \sigma_0$	$\sigma \ne \sigma_0$	$\sum_{i=1}^n (X_i-\mu)^2 \le \sigma_0^2\, \chi^2_{n,\alpha/2}$ oppure $\sum_{i=1}^n (X_i-\mu)^2 \ge \sigma_0^2\, \chi^2_{n,1-\alpha/2}$	$\overline{S}^2_n \le \frac{\sigma_0^2}{n-1}\, \chi^2_{n-1,\alpha/2}$ oppure $\overline{S}^2_n \ge \frac{\sigma_0^2}{n-1}\, \chi^2_{n-1,1-\alpha/2}$

Tabella 7.1: regioni di rifiuto dell'ipotesi $\mathcal{H}_0$ specificata per test del *Chi-quadro*.

Se μ è nota i test (A) e (B) sono UPP. Se μ non è nota i test (A) e (B) sono UPP non distorti. Nei test (C) i valori s_1, s_2 sono tali che ciascuna delle due regioni critiche abbia probabilità $\frac{\alpha}{2}$, come è tradizione in Letteratura — si veda, ad esempio, [24] — anche se ciò invalida la proprietà di non distorsione (almeno per piccoli campioni).

7.4.2 Test t-Student (Speranza)

I test *t-Student* riguardano essenzialmente la verifica di ipotesi sulla speranza di un campione Gaussiano $\mathbf{X}^n_{\mathtt{N}(\mu,\sigma^2)}$. Sia $\mu_0 \in \mathbb{R}$ nota. Nella *Tabella (7.2)* si mostrano le regioni di rifiuto di diverse ipotesi $\mathcal{H}_0$ per test di livello $0 < \alpha < 1$; qui z_p indica il quantile di ordine p della legge $\mathtt{N}(0,1)$ e $t_{n,p}$ il quantile di ordine p della legge $\mathtt{t}(n)$.

Se σ è nota i test (A) e (B) sono UPP e i test (C) sono UPP non distorti. Se σ non è nota tutti i test sono UPP non distorti. Nei test (C) il valore t_0 è tale che ciascuna delle due regioni critiche abbia probabilità $\frac{\alpha}{2}$. Se n è sufficientemente grande i quantili $t_{n,p}$ della legge $\mathtt{t}(n)$ possono essere sostituiti dai quantili z_p della legge $\mathtt{N}(0,1)$ — tradizionalmente in Letteratura si assume che ciò sia fattibile per $n > 30$ (si veda l'*Osservazione (6.2)*). Infine, si notino le proprietà di *simmetria* dei quantili utilizzati.

7.4.3 Test di Fisher-Snedecor (Confronto Varianze)

I test di *Fisher-Snedecor* riguardano essenzialmente la verifica di ipotesi sulle varianze di campioni Gaussiani $\mathbf{X}^n_{\mathtt{N}(\mu_1,\sigma_1^2)}$ e $\mathbf{Y}^m_{\mathtt{N}(\mu_2,\sigma_2^2)}$ *indipendenti*. Nella *Tabella (7.3)* si mostrano le regioni di rifiuto di diverse ipotesi $\mathcal{H}_0$ per

	Ipotesi		**Test / Regione di rifiuto**	
	$\mathcal{H}_0$	$\mathcal{H}_1$	**σ nota**	**σ non nota**
(A)	$\mu \leq \mu_0$	$\mu > \mu_0$	$\overline{X}_n \geq \mu_0 + \frac{\sigma}{\sqrt{n}}\, z_{1-\alpha}$	$\overline{X}_n \geq \mu_0 + \frac{\overline{S}_n}{\sqrt{n}}\, t_{n-1,1-\alpha}$
(B)	$\mu \geq \mu_0$	$\mu < \mu_0$	$\overline{X}_n \leq \mu_0 + \frac{\sigma}{\sqrt{n}}\, z_{\alpha}$	$\overline{X}_n \leq \mu_0 + \frac{\overline{S}_n}{\sqrt{n}}\, t_{n-1,\alpha}$
(C)	$\mu = \mu_0$	$\mu \neq \mu_0$	$\left\lvert \overline{X}_n - \mu_0 \right\rvert \geq \frac{\sigma}{\sqrt{n}}\, z_{1-\alpha/2}$	$\left\lvert \overline{X}_n - \mu_0 \right\rvert \geq \frac{\overline{S}_n}{\sqrt{n}}\, t_{n-1,1-\alpha/2}$

Tabella 7.2: regioni di rifiuto dell'ipotesi $\mathcal{H}_0$ specificata per test *t-Student*.

test di livello $0 < \alpha < 1$; qui $f_{n,m,p}$ indica il quantile di ordine p della legge $\mathtt{F}(n, m)$. In base alla *Proposizione (6.6)* si ha che $\frac{n-1}{\sigma_1^2}\, \overline{S}^2_{n,X} \sim \chi^2(n-1)$ e $\frac{m-1}{\sigma_2^2}\, \overline{S}^2_{m,Y} \sim \chi^2(m-1)$, e le due v.a. sono indipendenti; dunque, come illustrato nella *Sezione 6.3*, la v.a. $R_{X,Y} = \left(\sigma_2^2\, \overline{S}^2_{n,X}\right) / \left(\sigma_1^2\, \overline{S}^2_{m,Y}\right)$ ha legge $\mathtt{F}(n-1, m-1)$. In particolare, la v.a. $R_{X,Y} = \overline{S}^2_{n,X} / \overline{S}^2_{m,Y}$ ha legge $\mathtt{F}(n-1, m-1)$ nel caso in cui $\sigma_1^2 = \sigma_2^2$. Tale statistica tende ad essere "grande" quando $\sigma_1^2 > \sigma_2^2$ e "piccola" quando $\sigma_1^2 < \sigma_2^2$; sfruttando questo differente comportamento è possibile formulare i test di ipotesi sulle varianze di campioni Gaussiani indipendenti mostrati in *Tabella (7.3)*.

	Ipotesi		**Test / Regione di rifiuto**	
	$\mathcal{H}_0$	$\mathcal{H}_1$	**μ_1, μ_2 note**	**μ_1, μ_2 non note**
(A)	$\sigma_1 \leq \sigma_2$	$\sigma_1 > \sigma_2$	$\frac{\sum_{i=1}^n (X_i - \mu_1)^2}{\sum_{i=1}^m (Y_i - \mu_2)^2} \geq \frac{n}{m}\, f_{n,m,1-\alpha}$	$\frac{\overline{S}^2_{n,X}}{\overline{S}^2_{m,Y}} \geq f_{n-1,m-1,1-\alpha}$
(B)	$\sigma_1 \geq \sigma_2$	$\sigma_1 < \sigma_2$	$\frac{\sum_{i=1}^m (Y_i - \mu_2)^2}{\sum_{i=1}^n (X_i - \mu_1)^2} \geq \frac{m}{n}\, f_{m,n,1-\alpha}$	$\frac{\overline{S}^2_{m,Y}}{\overline{S}^2_{n,X}} \geq f_{m-1,n-1,1-\alpha}$
(C)	$\sigma_1 = \sigma_2$	$\sigma_1 \neq \sigma_2$	$\frac{\sum_{i=1}^n (X_i - \mu_1)^2}{\sum_{i=1}^m (Y_i - \mu_2)^2} \geq \frac{n}{m}\, f_{n,m,1-\alpha/2}$ oppure $\frac{\sum_{i=1}^m (Y_i - \mu_2)^2}{\sum_{i=1}^n (X_i - \mu_1)^2} \geq \frac{m}{n}\, f_{m,n,1-\alpha/2}$	$\frac{\overline{S}^2_{n,X}}{\overline{S}^2_{m,Y}} \geq f_{n-1,m-1,1-\alpha/2}$, se $\overline{S}^2_{n,X} \geq \overline{S}^2_{m,Y}$ oppure $\frac{\overline{S}^2_{m,Y}}{\overline{S}^2_{n,X}} \geq f_{m-1,n-1,1-\alpha/2}$, se $\overline{S}^2_{n,X} < \overline{S}^2_{m,Y}$

Tabella 7.3: regioni di rifiuto dell'ipotesi $\mathcal{H}_0$ specificata per test di *Fisher-Snedecor*.

Se μ_1, μ_2 sono note i test (A) e (B) sono UPP. Se μ_1, μ_2 non sono note i test (A) e (B) sono UPP non distorti. Nei test (C) i valori r_1, r_2 sono tali che ciascuna delle due regioni critiche abbia probabilità $\frac{\alpha}{2}$, anche se ciò invalida la proprietà di non distorsione. Infine, si osservi come i test illustrati utilizzino i quantili della coda destra della legge di Fisher-Snedecor; ciò è possibile ricordando il risultato dell'*Esempio (6.4)*: $f_{n,m,p} = f^{-1}_{m,n,1-p}$.

Capitolo 8

STIMA PER INTERVALLI

Nelle applicazioni è spesso necessario saper individuare una regione dello spazio dei parametri che contenga i valori incogniti degli stessi (o di loro funzioni) con una probabilità prefissata. Ad esempio, potrebbe essere utile stabilire un limite inferiore o superiore al valore del parametro incognito di una certa legge con probabilità prefissata. A questo scopo si introduce la nozione di *regione di confidenza*.

DEFINIZIONE 8.1 (Regione di confidenza)**.** Sia $\mathfrak{M}$ un modello statistico dominato. Si dice *regione di confidenza* (o *di fiducia)* di livello $0 < \alpha < 1$ per la funzione $\xi\colon \Theta \to \Theta$ un'applicazione $\omega \mapsto I_\omega \subseteq \Theta$ tale che

$$\text{per ogni } \theta \in \Theta \text{ è } \{\omega \in \Omega\colon I_\omega \ni \xi(\theta)\} \in \mathfrak{F}, \tag{8.1a}$$

$$\text{per ogni } \theta \in \Theta \text{ vale } \mathbf{P}_\theta\{I_\omega \ni \xi(\theta)\} \geq 1 - \alpha. \tag{8.1b}$$

Se I_ω è un intervallo si parla di *intervallo di confidenza* (o *di fiducia)* (IC) di *livello di confidenza* $1 - \alpha$.

OSSERVAZIONE 8.1.

In termini pratici una regione di confidenza è una funzione che associa al caso (ovvero l'osservazione ω) un certo sottoinsieme I_ω di Θ che si presume possa contenere $\xi(\theta)$; la regione $I_\omega \subseteq \Theta$ dipende da ω ma non da θ.

Si noti che nell'*equazione (8.1b)* non è corretto affermare che "*la probabilità che $\xi(\theta)$ appartenga a I_ω è almeno pari a $1 - \alpha$*": infatti il valore $\xi(\theta)$, per quanto incognito, è *deterministico*, mentre è la regione I_ω ad essere dipendente dal caso (cioè dall'osservazione casuale $\omega \in \Omega$). Piuttosto l'*equazione (8.1b)* dovrebbe essere letta come "*la probabilità che I_ω contenga $\xi(\theta)$ è almeno pari a $1 - \alpha$*". In altri termini, il soggetto è I_ω, su cui effettivamente "transita" l'intenzione, mentre l'oggetto è $\xi(\theta)$, che di per sé non è "proteso" a stare in I_ω.

All'*equazione (8.1b)* è possibile dare la seguente interpretazione *frequentista*: si supponga di eseguire un certo numero N di esperimenti identici e indipendenti, in ciascuno dei quali I_ω può o meno contenere il valore $\xi(\theta)$ (infatti I_ω può variare da esperimento a esperimento); al crescere di N la proporzione di esperimenti in cui I_ω contiene il valore $\xi(\theta)$ convergerà ad un valore non inferiore a $(1 - \alpha)$. In generale, essendo $\xi(\theta)$ comunque incognito, è solamente possibile formulare un'affermazione statistica (del tipo "con probabilità almeno pari a $(1 - \alpha)$") circa il fatto che in ciascun esperimento I_ω contenga o meno il valore $\xi(\theta)$.

Attraverso gli intervalli di confidenza è possibile quantificare oggettivamente quanto uno stimatore di ξ sia effettivamente "vicino" a ξ. Nel seguito ci limiteremo al caso $\Theta \subseteq \mathbb{R}$ e $\xi(\theta) = \theta$.

DEFINIZIONE 8.2 (Intervalli aleatori)**.** Una famiglia I_ω di intervalli di $\Theta \subseteq \mathbb{R}$, dove I_ω dipende dall'osservazione $\omega \in \Omega$ ma non da $\theta \in \Theta$, è detta *famiglia di intervalli aleatori*. In particolare, se $I_\omega = (b(\omega), t(\omega))$ allora $b(\omega)$ è detto *limite inferiore* e $t(\omega)$ è detto *limite superiore*, dove le statistiche b, t dipendono dall'osservazione $\omega \in \Omega$ ma non da $\theta \in \Theta$. Si noti che può essere $b(\omega) = -\infty$ o $t(\omega) = +\infty$.

DEFINIZIONE 8.3 (Limiti di confidenza)**.** Sia I_ω un IC di livello $0 < \alpha < 1$ per la funzione $\xi(\theta)$. Se

$$I_\omega = \{\theta \in \Theta\colon b(\omega) \leq \xi(\theta) < +\infty\} \tag{8.2}$$

allora b è detto *limite inferiore di confidenza* di livello α per la funzione ξ. Se

$$I_\omega = \{\theta \in \Theta\colon -\infty < \xi(\theta) \leq t(\omega)\} \tag{8.3}$$

allora t è detto *limite superiore di confidenza* di livello α per la funzione ξ. Gli IC delle *equazioni (8.2)–(8.3)* sono detti *unilateri*. Se, invece,

$$I_\omega = \{\theta \in \Theta\colon\ -\infty < b(\omega) \le \xi(\theta) \le t(\omega) < +\infty\} \tag{8.4}$$

allora I_ω è un IC *bilatero* di livello α.

★ **Intervalli di confidenza e verifica di ipotesi.** Il risultato che segue è fondamentale, in quanto sancisce la sostanziale equivalenza tra la teoria degli intervalli di confidenza e quella della verifica di ipotesi.

TEOREMA 8.1. *Per ogni $\theta' \in \Theta$ sia $A_{\theta'}$ la regione di accettazione di un test τ di livello $0 < \alpha < 1$ dell'ipotesi $\mathcal{H}_0\colon \theta = \theta'$ contro l'alternativa $\mathcal{H}_1\colon \theta \neq \theta'$. Per ogni $\boldsymbol{x} \in \mathbb{R}^n$ sia $I_{\boldsymbol{x}} \subseteq \Theta$ tale che*

$$I_{\boldsymbol{x}} = \{\theta' \in \Theta\colon \boldsymbol{x} \in A_{\theta'}\}\,. \tag{8.5}$$

Allora, dato il campione $\boldsymbol{X}$, l'insieme $I_{\boldsymbol{X}}$ è una regione di confidenza per θ di livello α.

Viceversa, dato il campione $\boldsymbol{X}$, sia $I_{\boldsymbol{X}}$ una regione di confidenza per θ di livello $0 < \alpha < 1$. Per ogni $\theta' \in \Theta$ sia

$$A_{\theta'} = \{\boldsymbol{x} \in \mathbb{R}^n\colon I_{\boldsymbol{x}} \ni \theta'\}\,. \tag{8.6}$$

Allora l'insieme $A_{\theta'}$ è la regione di accettazione di un test τ di livello α dell'ipotesi $\mathcal{H}_0\colon \theta = \theta'$ contro l'alternativa $\mathcal{H}_1\colon \theta \neq \theta'$.

DIM. Per definizione di regione di accettazione si ha $\mathbf{P}_{\theta'}\{\boldsymbol{X} \notin A_{\theta'}\} \le \alpha$, ovvero $\mathbf{P}_{\theta'}\{\boldsymbol{X} \in A_{\theta'}\} \ge 1 - \alpha$. Stante l'arbitrarietà di θ', l'ultima disuguaglianza è valida per ogni $\theta \in \Theta$. Perciò, grazie all'*equazione (8.5)* si ha, per ogni $\theta \in \Theta$,

$$\mathbf{P}_\theta\{I_{\boldsymbol{X}} \ni \theta\} = \mathbf{P}_\theta\{\boldsymbol{X} \in A_\theta\} \ge 1 - \alpha,$$

cioè $I_{\boldsymbol{X}}$ è una regione di confidenza per θ di livello α.

Viceversa, l'errore di TIPO I di un test τ dell'ipotesi $\mathcal{H}_0\colon \theta = \theta'$ contro l'alternativa $\mathcal{H}_1\colon \theta \neq \theta'$ avente regione di accettazione $A_{\theta'}$ è

$$\mathbf{P}_{\theta'}\{\boldsymbol{X} \notin A_{\theta'}\} = \mathbf{P}_{\theta'}\{I_{\boldsymbol{X}} \not\ni \theta'\} \le \alpha.$$

Dunque τ è un test di livello α delle ipotesi considerate. □

L'Esempio che segue illustra l'applicazione del *Teorema (8.1)*.

ESEMPIO 8.1. ►

Si consideri un campione $\mathbf{X}^n_{\mathbb{N}(\mu,1)}$, dove $\mu \in \mathbb{R}$ non è nota. La costruzione di un IC bilatero di livello $0 < \alpha < 1$ per μ è riconducibile alla verifica delle ipotesi

$$\mathcal{H}_0\colon \mu = \mu_0 \quad \text{contro} \quad \mathcal{H}_1\colon \mu \neq \mu_0,$$

dove $\mu_0 \in \mathbb{R}$ è una costante nota ma arbitraria, un problema già affrontato nell'*Esempio (7.12)*. La riga (C) della *Tabella (7.2)* mostra che l'ipotesi $\mathcal{H}_0$ può essere accettata, al livello α, se $|\overline{x}_n - \mu_0| \le z_{1-\alpha/2}/\sqrt{n}$ o, equivalentemente, se

$$\overline{x}_n - \frac{z_{1-\alpha/2}}{\sqrt{n}} \le \mu_0 \le \overline{x}_n + \frac{z_{1-\alpha/2}}{\sqrt{n}}.$$

Dunque, essendo $\mathbf{P}_{\mathcal{H}_0}\{\text{Acc. } \mathcal{H}_0\} = 1 - \mathbf{P}_{\mathcal{H}_0}\{\text{Rif. } \mathcal{H}_0\} = 1 - \alpha$, si ha

$$\mathbf{P}_{\mathcal{H}_0}\left\{\overline{X}_n - \frac{z_{1-\alpha/2}}{\sqrt{n}} \le \mu_0 \le \overline{X}_n + \frac{z_{1-\alpha/2}}{\sqrt{n}}\right\} = 1 - \alpha.$$

Tuttavia, data l'arbitrarietà di μ_0, tale relazione è vera quale che sia $\mu_0 \in \mathbb{R}$, sicché l'uguaglianza

$$\mathbf{P}_\mu\left\{\overline{X}_n - \frac{z_{1-\alpha/2}}{\sqrt{n}} \le \mu \le \overline{X}_n + \frac{z_{1-\alpha/2}}{\sqrt{n}}\right\} = 1 - \alpha$$

vale per ogni $\mu \in \Theta = \mathbb{R}$. Ne segue che

$$I_{\boldsymbol{X}} = \left(\overline{X}_n - \frac{z_{1-\alpha/2}}{\sqrt{n}}, \overline{X}_n + \frac{z_{1-\alpha/2}}{\sqrt{n}}\right)$$

è un IC bilatero di livello α (o di livello di confidenza $1-\alpha$) per μ.

Scegliendo, ad esempio, $\alpha = 0.05$ l'intervallo cercato è (approssimativamente)

$$I_{\boldsymbol{X}} \approx \left(\overline{X}_n - \frac{1.96}{\sqrt{n}}, \overline{X}_n + \frac{1.96}{\sqrt{n}}\right),$$

che ha lunghezza pari a circa $\frac{3.92}{\sqrt{n}}$. Se, invece, $\alpha = 0.1$ allora l'intervallo cercato è (approssimativamente)

$$I_{\boldsymbol{X}} \approx \left(\overline{X}_n - \frac{1.645}{\sqrt{n}}, \overline{X}_n + \frac{1.645}{\sqrt{n}}\right),$$

che ha lunghezza pari a circa $\frac{3.29}{\sqrt{n}}$, inferiore a quella trovata in precedenza. In effetti, aumentare α equivale a diminuire $1-\alpha$, cioè la probabilità con cui si chiede che $I_{\boldsymbol{X}}$ contenga μ. Si noti che in entrambi i casi la lunghezza di $I_{\boldsymbol{X}}$ tende a zero al crescere di n (ossia dell'"informazione" disponibile). ◀

OSSERVAZIONE 8.2.

Sia l'*Esempio (8.1)* sia il *Teorema (8.1)* chiariscono le connessioni tra la teoria degli intervalli di confidenza e quella della verifica di ipotesi; in particolare, l'Esempio mostra un procedimento tipico di "inversione" di un test al fine di ottenere un intervallo di confidenza. Non si deve tuttavia dimenticare che mentre nella verifica di ipotesi si cerca di individuare quali siano i valori campionari, ovvero la regione di accettazione, consistenti con le ipotesi sui parametri (che risultano così in qualche modo fissati), nella stima per intervalli è l'osservazione del campione ad essere fissata e ci si chiede quali siano i valori dei parametri consistenti con i dati disponibili.

L'Esempio che segue illustra una soluzione alternativa a quella proposta nell'*Esempio (8.1)*.

ESEMPIO 8.2. ►

Si consideri un campione $\mathbf{X}^n_{\mathtt{N}(\mu,1)}$, dove $\mu \in \mathbb{R}$ non è nota. Per costruire un IC bilatero di livello $0 < \alpha < 1$ per μ ricorriamo alla media campionaria $\overline{X}_n \sim \mathtt{N}\big(\mu, \frac{1}{n}\big)$. La *disuguaglianza di Čebyšev* fornisce, per $\epsilon > 0$,

$$\mathbf{P}_\mu\left\{\left|\overline{X}_n - \mu\right| \leq \frac{\epsilon}{\sqrt{n}}\right\} \geq 1 - \frac{1}{\epsilon^2}.$$

Fissato $0 < \alpha < 1$ e preso $\epsilon = \alpha^{-1/2}$ si ottiene l'IC bilatero di livello α

$$I_{\boldsymbol{X}} = \left(\overline{X}_n - \frac{1}{\sqrt{\alpha}}\,\frac{1}{\sqrt{n}}, \overline{X}_n + \frac{1}{\sqrt{\alpha}}\,\frac{1}{\sqrt{n}}\right),$$

per il quale vale $\mathbf{P}_\mu\{I_{\boldsymbol{X}} \ni \mu\} \geq 1-\alpha$.

Se, ad esempio, $\alpha = 0.05$ allora $\epsilon \approx 4.47$ e dunque l'intervallo cercato è (approssimativamente)

$$I_{\boldsymbol{X}} = \left(\overline{X}_n - 4.47\,\frac{1}{\sqrt{n}}, \overline{X}_n + 4.47\,\frac{1}{\sqrt{n}}\right),$$

che ha lunghezza pari a circa $\frac{8.94}{\sqrt{n}}$. Se, invece, $\alpha = 0.1$ allora $\epsilon \approx 3.16$ e dunque l'intervallo cercato è (approssimativamente)

$$I_{\boldsymbol{X}} = \left(\overline{X}_n - 3.16\,\frac{1}{\sqrt{n}}, \overline{X}_n + 3.16\,\frac{1}{\sqrt{n}}\right),$$

che ha lunghezza pari a circa $\frac{6.32}{\sqrt{n}}$, inferiore a quella trovata in precedenza. Anche in questo caso valgono le osservazioni circa gli effetti dell'aumento del livello α svolte nell'*Esempio (8.1)*; si noti inoltre come gli intervalli trovati ora siano più lunghi di quelli calcolati in precedenza: essenzialmente ciò è dovuto al fatto che spesso la *disuguaglianza di Čebyšev* fornisce stime troppo grossolane. ◀

ESEMPIO 8.3. ►

Si consideri un campione $\mathbf{X}^n_{\mathtt{U}_{\mathtt{c}}(0,\theta)}$, dove $\theta > 0$ non è noto. Per costruire un IC bilatero di livello $0 < \alpha < 1$ per θ ricorriamo alla statistica d'ordine $X_{(n)}$. Per quanto visto nel *Capitolo 2* e nell'*Esempio (4.6)* tale v.a. ha densità data da

$$f_\theta(x) = \frac{n}{\theta^n}\, x^{n-1}\, \mathbb{I}_{(0,\theta)}(x),$$

e dunque

$$\mathbf{E}_\theta\big(X_{(n)}\big) = \frac{n}{n+1}\,\theta \quad \text{e} \quad \mathbf{E}_\theta\Big(\big|X_{(n)} - \theta\big|^2\Big) = \frac{2}{(n+1)(n+2)}\,\theta^2.$$

La *disuguaglianza di Čebyšev* fornisce, per $\epsilon > 0$,

$$\mathbf{P}_\theta\left\{\frac{\big|X_{(n)} - \theta\big|}{\theta}\sqrt{\frac{(n+1)(n+2)}{2}} \le \epsilon\right\} \ge 1 - \frac{1}{\epsilon^2}.$$

Poiché[1] $X_{(n)} \xrightarrow{\mathbf{P}} \theta$, per n sufficientemente grande è possibile approssimare tale espressione con

$$\mathbf{P}_\theta\left\{\frac{\big|X_{(n)} - \theta\big|}{X_{(n)}}\sqrt{\frac{(n+1)(n+2)}{2}} \le \epsilon\right\} \ge 1 - \frac{1}{\epsilon^2}.$$

Fissato $0 < \alpha < 1$ e preso $\epsilon = \alpha^{-1/2}$ si ottiene l'IC bilatero (approssimato) di livello α

$$I_{\boldsymbol{X}} \approx \left(X_{(n)}\left(1 - \sqrt{\frac{2}{\alpha(n+1)(n+2)}}\right), X_{(n)}\left(1 + \sqrt{\frac{2}{\alpha(n+1)(n+2)}}\right)\right)$$

o, ancora, essendo $\sqrt{\frac{1}{(n+1)(n+2)}} \approx \frac{1}{n}$ per $n \gg 1$, è possibile ottenere

$$I_{\boldsymbol{X}} \approx \left(X_{(n)}\left(1 - \frac{1}{n}\sqrt{\frac{2}{\alpha}}\right), X_{(n)}\left(1 + \frac{1}{n}\sqrt{\frac{2}{\alpha}}\right)\right),$$

di lunghezza (aleatoria) pari a $\frac{2}{n}\sqrt{\frac{2}{\alpha}}\,X_{(n)}$, che tende a zero per $n \to \infty$.

In alternativa, si consideri la v.a. $T = \frac{X_{(n)}}{\theta}$ che ha densità data da

$$f(t) = n\,t^{n-1}\,\mathbb{I}_{(0,1)}(t),$$

che *non* dipende da θ. Si tratta dunque di trovare dei valori $0 \le t_1 < t_2 \le 1$ tali che sia, per ogni $\theta > 0$,

$$\mathbf{P}\{t_1 < T < t_2\} = \mathbf{P}\left\{\frac{X_{(n)}}{t_2} < \theta < \frac{X_{(n)}}{t_1}\right\} = 1 - \alpha,$$

dove $0 < \alpha < 1$. Risolvendo rispetto alle variabili t_1, t_2 l'equazione

$$\mathbf{P}\{t_1 < T < t_2\} = \int_{t_1}^{t_2} f(t)\;\mathrm{d}t = \int_{t_1}^{t_2} n\,t^{n-1}\;\mathrm{d}t = 1 - \alpha$$

si ottiene $t_2^n - t_1^n = 1 - \alpha$, che ha infinite soluzioni. Ponendo $t_2 = 1$ si ottiene $t_1 = \alpha^{1/n}$ e dunque un IC bilatero di livello α è dato da

$$I'_{\boldsymbol{X}} = \left(X_{(n)}, \alpha^{-1/n}\,X_{(n)}\right),$$

di lunghezza (aleatoria) pari a $(\alpha^{-1/n} - 1)\,X_{(n)}$, che tende a zero per $n \to \infty$.

La *Figura (8.1)* mostra il confronto tra le lunghezze degli IC $I_{\boldsymbol{X}}$ e $I'_{\boldsymbol{X}}$ al crescere dell'ampiezza n del campione per diversi valori del livello α: evidentemente, per campioni di ampiezza sufficiente, gli intervalli $I'_{\boldsymbol{X}}$ risultano essere più corti dei corrispondenti intervalli $I_{\boldsymbol{X}}$, ed entrambi si accorciano al crescere di n; inoltre, a parità di ampiezza del campione, gli intervalli si allungano al diminuire di α.

◀

Negli esempi appena visti, fissato $0 < \alpha < 1$, esiste una grande varietà di IC di livello α. Ovviamente, quanto maggiore è la lunghezza dell'intervallo I_ω tanto maggiore è la probabilità che esso contenga $\xi(\theta)$. In effetti l'intervallo $I_\infty = (-\infty, +\infty)$, che ignora completamente il campione $\boldsymbol{X}$, contiene $\xi(\theta)$ con probabilità 1. Tuttavia, quanto più ampio è I_ω tanto meno *significativo* ("informativo") esso risulta. Perciò nella pratica, fissato il livello di confidenza desiderato, si cerca di costruire un IC di lunghezza quanto più piccola possibile; purtroppo in generale non esiste un IC che sia di lunghezza *uniformemente minima* tra gli IC di livello α.

[1] Conoscendo la densità di $X_{(n)}$ è facile dimostrare la *convergenza in probabilità* $X_{(n)} \xrightarrow{\mathbf{P}} \theta$ di $X_{(n)}$ a θ, cioè che per ogni $\epsilon > 0$ si ha $\mathbf{P}\{|X_{(n)} - \theta| \le \epsilon\} \to 1$ quando $n \to \infty$.

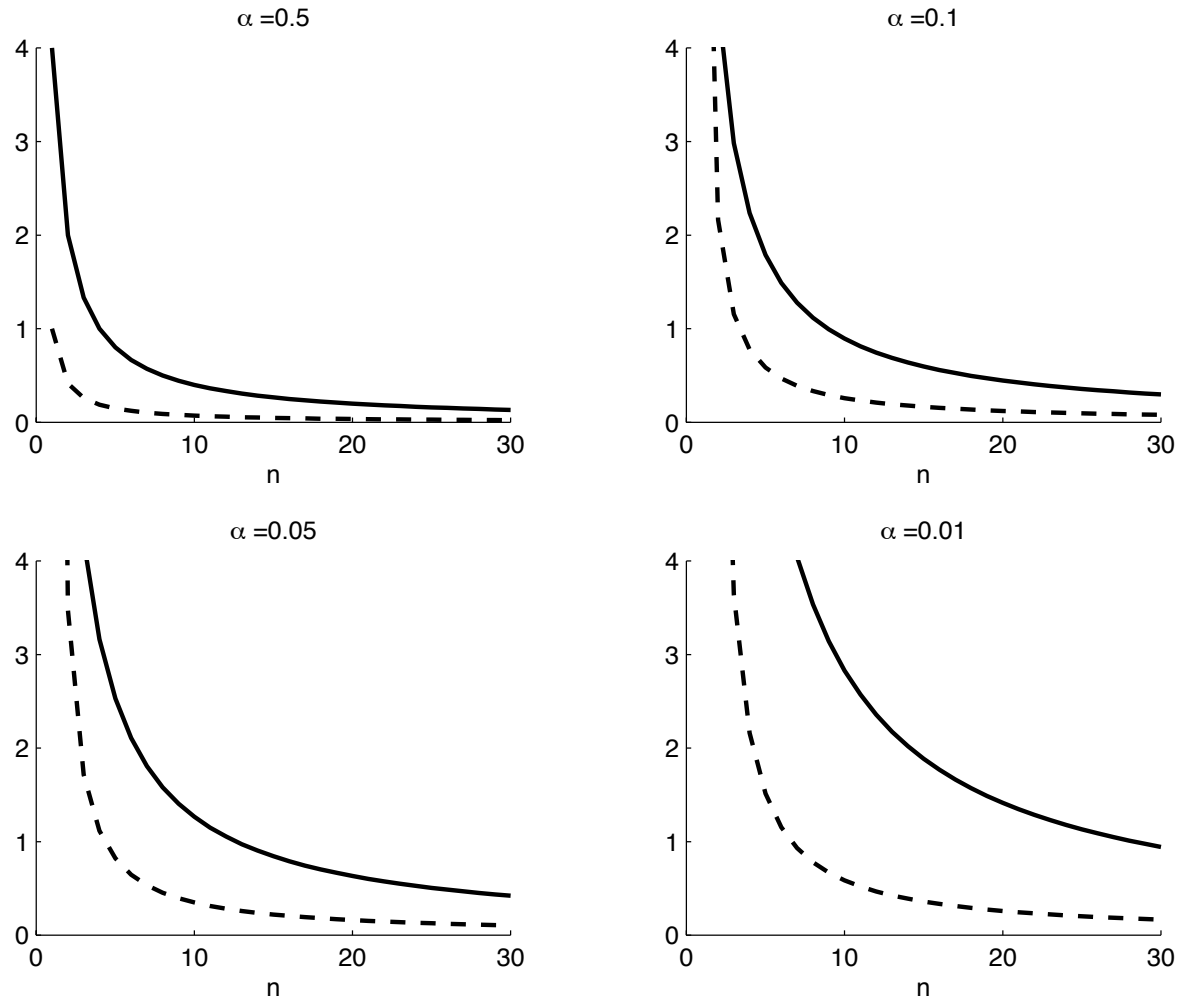

Figura 8.1: lunghezza degli IC $I_{\boldsymbol{X}}$ (linea *continua*) e $I'_{\boldsymbol{X}}$ (linea *tratteggiata*) dell'*Esempio (8.3)* al crescere dell'ampiezza n del campione per diversi valori del livello α; il fattore comune $X_{(n)}$ è stato trascurato.

8.1 Metodo del pivot

In questa Sezione generalizzeremo la procedura di costruzione di IC già adottata nella soluzione alternativa dell'*Esempio (8.3).* In generale, sia $\boldsymbol{X}$ un campione e sia $T(\boldsymbol{X}, \theta)$ una v.a. la cui legge sia *indipendente* da θ. Supponiamo che esistano funzioni $t_1(\alpha), t_2(\alpha)$, con $0 < \alpha < 1$, tali per cui valga, per ogni $\theta \in \Theta$,

$$\mathbf{P}\{t_1(\alpha) < T(\boldsymbol{X}, \theta) < t_2(\alpha)\} = 1 - \alpha,$$

dove $\mathbf{P}$ è la probabilità indotta dalla legge di T (che non dipende da θ), e supponiamo che tale equazione possa essere riscritta (trasformata) come

$$\mathbf{P}\{b(\boldsymbol{X}) < \theta < t(\boldsymbol{X})\} = 1 - \alpha,$$

dove b, t sono due statistiche le cui leggi non dipendono da θ. Evidentemente $(b(\boldsymbol{X}), t(\boldsymbol{X}))$ rappresenta un IC di livello α per θ. Come già visto, le funzioni $t_1(\alpha), t_2(\alpha)$ possono essere scelte in diversi modi; in particolare si dovrebbe cercare di fare in modo che la differenza $t(\boldsymbol{X}) - b(\boldsymbol{X})$ sia minima: in tal caso l'intervallo trovato, basato su T, sarebbe un IC di livello α per θ di lunghezza minima, e quindi massimamente "informativo". Si noti tuttavia che intervalli basati su statistiche diverse da T potrebbero avere lunghezza inferiore.

DEFINIZIONE 8.4 (Statistica pivotale)**.** Una v.a. $T(\boldsymbol{X}, \theta)$, funzione del campione $\boldsymbol{X}$ e del parametro θ, la cui legge sia *indipendente* da θ è detta *statistica*[2] *pivotale* (o *statistica centrale*).

Le statistiche pivotali costituiscono dunque un valido strumento nella costruzione di IC. Fissato $0 < \alpha < 1$, resta il problema della scelta delle costanti $t_1(\alpha), t_2(\alpha)$ così da minimizzare la lunghezza dell'IC ottenuto.

ESEMPIO 8.4. ►
Si riconsideri l'*Esempio (8.3)*. Evidentemente la v.a. $T = T(\boldsymbol{X}, \theta) = \frac{X_{(n)}}{\theta}$ è una statistica pivotale, con densità f data da

$$f(t) = n\, t^{n-1}\, \mathbb{I}_{(0,1)}(t),$$

che *non* dipende da θ. Per costruire un IC bilatero di livello $0 < \alpha < 1$ per θ possiamo procedere come segue. Anzitutto si tratta di trovare dei valori $0 \le t_1 < t_2 \le 1$ tali per cui sia, per ogni θ,

$$1 - \alpha = \mathbf{P}\{t_1 < T < t_2\} = \mathbf{P}\left\{ t_1 < \frac{X_{(n)}}{\theta} < t_2 \right\} = \mathbf{P}\left\{ \frac{X_{(n)}}{t_2} < \theta < \frac{X_{(n)}}{t_1} \right\}.$$

Tale IC ha dunque lunghezza $L = X_{(n)} \left(\frac{1}{t_1} - \frac{1}{t_2} \right)$, da rendere minima sotto il vincolo

$$\mathbf{P}\{t_1 < T < t_2\} = \int_{t_1}^{t_2} f(t)\, \mathrm{d}t = \int_{t_1}^{t_2} n\, t^{n-1}\, \mathrm{d}t = t_2^n - t_1^n = 1 - \alpha.$$

Essendo $(1-\alpha)^{1/n} < t_2 \le 1$ si ha

$$\frac{\mathrm{d}L}{\mathrm{d}t_2} = X_{(n)} \left(-\frac{1}{t_1^2} \frac{\mathrm{d}t_1}{\mathrm{d}t_2} + \frac{1}{t_2^2} \right) = X_{(n)}\, \frac{t_1^{n+1} - t_2^{n+1}}{t_2^2\, t_1^{n+1}} < 0,$$

ovvero L è una funzione decrescente nella variabile t_2, sicché il minimo si ottiene per $t_2 = 1$. Dunque l'IC di lunghezza minima (basato sulla statistica pivotale T) è quello già calcolato nell'*Esempio (8.3)*, cioè

$$I_{\boldsymbol{X}} = \left(X_{(n)}, \alpha^{-1/n}\, X_{(n)} \right),$$

di lunghezza (aleatoria) pari a $(\alpha^{-1/n} - 1)\, X_{(n)}$, che tende a zero per $n \to \infty$. ◄

8.2 IC per campioni Gaussiani

Usando i risultati del *Capitolo 6* si dimostra l'esistenza di statistiche pivotali per campioni Gaussiani.

PROPOSIZIONE 8.1. *Si consideri un campione Gaussiano $\mathbf{X}^n_{\mathtt{N}(\mu,\sigma^2)}$, con $\mu \in \mathbb{R}$ e $\sigma > 0$. Le v.a. seguenti sono candidate ad essere usate quali statistiche pivotali.*

1. *Se σ^2 è nota si usi per μ la v.a.*

$$T = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathtt{N}(0,1)\,. \tag{8.7}$$

2. *Se σ^2 non è nota si usi per μ la v.a.*

$$T = \frac{\overline{X}_n - \mu}{\overline{S}_n/\sqrt{n}} \sim \mathtt{t}(n-1)\,. \tag{8.8}$$

3. *Se μ è nota si usi per σ^2 la v.a.*

$$T = \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} \sim \chi^2(n)\,. \tag{8.9}$$

[2] In base alla *Definizione (1.5)* il termine "statistica" è improprio: infatti T dipende da θ, sebbene la sua legge ne sia indipendente.

4. Se μ non è nota si usi per σ^2 la v.a.

$$T = (n-1)\,\frac{\overline{S}_n^2}{\sigma^2} \sim \chi^2(n-1)\,. \tag{8.10}$$

Ovviamente nessuna delle leggi di tali v.a. dipende da μ e/o σ^2.

Sfruttando quanto già visto nella *Sezione 7.4* e i risultati della *Proposizione (8.1)* è possibile costruire IC per i due parametri (μ, σ^2) della legge $\mathtt{N}\big(\mu, \sigma^2\big)$. Alla luce del *Teorema (8.1)* valgono le considerazioni già espresse in precedenza circa la verifica di ipotesi per campioni Gaussiani.

Nella *Tabella (8.1)* si illustrano diversi IC di livello $0 < \alpha < 1$ per la speranza di un campione Gaussiano $\mathbf{X}^n_{\mathtt{N}(\mu,\sigma^2)}$; qui z_p indica il quantile di ordine p della legge $\mathtt{N}(0,1)$ e $t_{n,p}$ il quantile di ordine p della legge $\mathtt{t}(n)$, $p \in (0,1)$. Gli IC (A) e (B) sono unilateri, gli IC (C) sono bilateri e di lunghezza minima tra quelli basati sulle rispettive statistiche pivotali.

	σ nota	**σ non nota**
(A)	$\left(\overline{X}_n - \frac{\sigma}{\sqrt{n}}\, z_{1-\alpha}, \infty\right)$	$\left(\overline{X}_n - \frac{\overline{S}_n}{\sqrt{n}}\, t_{n-1,1-\alpha}, \infty\right)$
(B)	$\left(-\infty, \overline{X}_n + \frac{\sigma}{\sqrt{n}}\, z_{1-\alpha}\right)$	$\left(-\infty, \overline{X}_n + \frac{\overline{S}_n}{\sqrt{n}}\, t_{n-1,1-\alpha}\right)$
(C)	$\left(\overline{X}_n - \frac{\sigma}{\sqrt{n}}\, z_{1-\alpha/2}, \overline{X}_n + \frac{\sigma}{\sqrt{n}}\, z_{1-\alpha/2}\right)$	$\left(\overline{X}_n - \frac{\overline{S}_n}{\sqrt{n}}\, t_{n-1,1-\alpha/2}, \overline{X}_n + \frac{\overline{S}_n}{\sqrt{n}}\, t_{n-1,1-\alpha/2}\right)$

Tabella 8.1: IC di livello $0 < \alpha < 1$ per la media di un campione Gaussiano.

Nella *Tabella (8.2)* si illustrano diversi IC di livello $0 < \alpha < 1$ per la varianza di un campione Gaussiano $\mathbf{X}^n_{\mathtt{N}(\mu,\sigma^2)}$; qui $\chi^2_{n,p}$ indica il quantile di ordine p della legge $\chi^2(n)$, $p \in (0,1)$. Gli IC (A) e (B) sono unilateri, gli IC (C) sono bilateri *a code uguali* ma non di lunghezza minima tra quelli basati sulle rispettive statistiche pivotali.

	μ nota	**μ non nota**
(A)	$\left(\frac{\sum_{i=1}^n (X_i-\mu)^2}{\chi^2_{n,1-\alpha}}, \infty\right)$	$\left(\frac{(n-1)\,\overline{S}_n^2}{\chi^2_{n-1,1-\alpha}}, \infty\right)$
(B)	$\left(-\infty, \frac{\sum_{i=1}^n (X_i-\mu)^2}{\chi^2_{n,\alpha}}\right)$	$\left(-\infty, \frac{(n-1)\,\overline{S}_n^2}{\chi^2_{n-1,\alpha}}\right)$
(C)	$\left(\frac{\sum_{i=1}^n (X_i-\mu)^2}{\chi^2_{n,1-\alpha/2}}, \frac{\sum_{i=1}^n (X_i-\mu)^2}{\chi^2_{n,\alpha/2}}\right)$	$\left(\frac{(n-1)\,\overline{S}_n^2}{\chi^2_{n-1,1-\alpha/2}}, \frac{(n-1)\,\overline{S}_n^2}{\chi^2_{n-1,\alpha/2}}\right)$

Tabella 8.2: IC di livello $0 < \alpha < 1$ per la varianza di un campione Gaussiano.

Capitolo 9

STATISTICA NON PARAMETRICA

Si parla di *statistica non parametrica* (o, usando la terminologia anglosassone, di *distribution-free statistics*) quando, nell'affrontare problemi di natura statistica, non si formulino ipotesi circa la legge di probabilità in gioco. Ad esempio, ciò accade quando si approssimi la distribuzione incognita F del campione in esame usando la f.r. empirica, oppure quando si stimino i quantili della popolazione tramite le statistiche d'ordine.

9.1 Stima non parametrica corretta

In questa Sezione ci occuperemo di *stima non parametrica corretta*; come vedremo, alcune generalizzazioni di concetti già noti porteranno ad un analogo del *Teorema di Lehman–Scheffé* (per maggiori dettagli si veda [24]).

Sia $\boldsymbol{X} = (X_1, \ldots, X_n)$ un campione di v.a. i.i.d.; nel seguito indicheremo con $\mathfrak{F}$ la classe di tutte le possibili distribuzioni F_X di X che siano o assolutamente continue o puramente discrete, o sotto-classi di queste.

DEFINIZIONE 9.1 (Statistica sufficiente per $\mathfrak{F}$)**.** Una statistica S si dice *sufficiente* per la famiglia di f.r. $\mathfrak{F}$ se la distribuzione condizionale del campione dato l'evento $\{S = s\}$ è la stessa qualunque sia la f.r. $F \in \mathfrak{F}$ del campione.

ESEMPIO 9.1. ►
Sia $\mathbf{X}_F^n$ un campione, con F assolutamente continua, e sia $\boldsymbol{S} = (X_{(1)}, \ldots, X_{(n)})$ l'insieme delle statistiche d'ordine. Poiché la densità condizionale del campione $f(\boldsymbol{x} \mid \boldsymbol{S} = \boldsymbol{s})$ vale sempre $1/n!$, la statistica S è sufficiente per la sotto-classe delle funzioni di ripartizione assolutamente continue su $\mathbb{R}$. ◄

DEFINIZIONE 9.2 (Famiglia completa)**.** Una famiglia di f.r. $\mathfrak{F}$ si dice *completa* se l'unico stimatore non distorto di 0 è la statistica identicamente nulla, ovvero se, data la funzione sommabile g,

$$\mathbf{E}_F(g) = 0 \quad \text{per ogni } F \in \mathfrak{F} \tag{9.1}$$

implica che $g \equiv 0$ (tranne, eventualmente, per un insieme di misura nulla rispetto ad ogni $F \in \mathfrak{F}$).

DEFINIZIONE 9.3 (Statistica completa per $\mathfrak{F}$)**.** Una statistica S si dice *completa* per la famiglia di f.r. $\mathfrak{F}$ se la classe delle distribuzioni indotte da S è completa.

ESEMPIO 9.2. ►
Si consideri il campione $\mathbf{X}_{\mathtt{Be}(\theta)}^n$, $\theta \in (0,1)$, di ampiezza $n > 0$ fissata. La statistica $S = \sum_{i=1}^n X_i$ è completa per famiglia di leggi Bernoulliane $\{\mathtt{Be}(\theta)\}_{\theta \in (0,1)}$, essendo completa la famiglia di leggi Binomiali $\{\mathtt{Bi}(n, \theta)\}_{\theta \in (0,1)}$ indotta da S. Infatti, se la funzione g è sommabile, si ha

$$\mathbf{E}_\theta(g(S)) = \sum_{s=0}^{n} g(s) \binom{n}{s} \theta^s (1-\theta)^{n-s} = 0 \quad \text{per ogni } \theta \in (0,1)$$

se, e solo se,

$$(1-\theta)^n \sum_{s=0}^{n} g(s) \binom{n}{s} \left(\frac{\theta}{1-\theta}\right)^s = 0 \quad \text{per ogni } \theta \in (0,1).$$

Ricordando che un polinomio è identicamente nullo se, e solo se, sono nulli tutti i suoi coefficienti si ricava che $g(s) = 0$ per $s = 0, \ldots, n$, ovvero $g \equiv 0$. ◄

Il risultato che segue è importante.

PROPOSIZIONE 9.1. *L'insieme delle statistiche d'ordine $(X_{(1)}, \ldots, X_{(n)})$ è una statistica sufficiente e completa, purché la legge comune del campione sia o assolutamente continua o discreta.*

Nel seguito ci concentreremo sulla stima di parametri.

DEFINIZIONE 9.4 (Parametro stimabile)**.** Un parametro reale ξ_F si dice *stimabile* se ne esiste uno stimatore non distorto, ovvero se esiste una statistica T tale che

$$\mathbf{E}_F(T) = \xi_F \tag{9.2}$$

per ogni $F \in \mathfrak{F}$.

Alcuni esempi di parametri stimabili sono i seguenti; qui è sottointeso che $F \in \mathfrak{F}$.

1. Nella famiglia di distribuzioni $\mathfrak{F}$ si ha che $\pi_c = \mathbf{P}_F\{X > c\}$, dove $c \in \mathbb{R}$ è una costante nota, è un parametro stimabile: basta considerare la statistica $T = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{(c,\infty)}(X_i)$.
2. Nella famiglia di distribuzioni $\mathfrak{F}$ dotate di momento del primo ordine si ha che μ_F, la speranza della popolazione F, è un parametro stimabile: basta considerare la media campionaria $\overline{X}_n$.
3. Nella famiglia di distribuzioni $\mathfrak{F}$ dotate di momento del secondo ordine si ha che σ_F^2, la varianza della popolazione F, è un parametro stimabile: basta considerare la varianza campionaria (corretta) $\overline{S}_n^2$.

DEFINIZIONE 9.5 (Grado di un parametro stimabile)**.** Il *grado* $\nu \geq 1$ di un parametro stimabile ξ_F è la minima ampiezza del campione per la quale ξ_F sia stimabile, ovvero è il più piccolo valore n tale che esista una statistica $T(X_1, \ldots, X_n)$ tale che $\mathbf{E}_F(T) = \xi_F$ per ogni $F \in \mathfrak{F}$.

Considerando gli esempi appena visti si ha che:

1. il grado del parametro stimabile π_c è $\nu = 1$;
2. il grado del parametro stimabile μ_F è $\nu = 1$ (si assume che esista almeno una distribuzione $F \in \mathfrak{F}$ tale che $\mu_F \neq 0$);
3. il grado del parametro stimabile σ_F^2 è $\nu = 2$ (si dimostra, infatti, che non è possibile stimare correttamente la varianza utilizzando una sola osservazione).

DEFINIZIONE 9.6 (Nucleo)**.** Uno stimatore corretto di un parametro stimabile basato su di un campione di ampiezza minima (pari al grado ν) è detto *nucleo* (o *kernel*).

Per quanto già visto, considerando un campione $X_1, \ldots, X_n$, ciascuna singola v.a. X_i, $i = 1, \ldots, n$, costituisce un nucleo (di grado $\nu = 1$) per μ_F. Similmente, ricordando l'espressione della varianza $\mathbf{V}(X) = \mathbf{E}\big(X^2\big) - \mathbf{E}(X)^2$, la statistica

$$T(X_i, X_j) = X_i^2 - X_i X_j, \quad i, j = 1, \ldots, n \quad (i \neq j),$$

costituisce un nucleo (di grado $\nu = 2$) per σ_F^2. Il risultato che segue è importante e sancisce l'esistenza di un nucleo invariante rispetto alle permutazioni degli indici.

PROPOSIZIONE 9.2. *Esiste un nucleo simmetrico per ogni parametro stimabile.*

DIM. Se $T(X_1, \ldots, X_\nu)$ costituisce un nucleo per ξ_F, allora tale è anche la statistica

$$T_s(X_1, \ldots, X_\nu) = \frac{1}{\nu!} \sum_{\mathfrak{P}} T(X_{i_1}, \ldots, X_{i_\nu}), \tag{9.3}$$

dove la somma è sull'insieme $\mathfrak{P}$ di tutte le $\nu!$ permutazioni di $\{1, \ldots, \nu\}$. □

Per quanto già visto, considerando un campione $X_1, \ldots, X_n$, un nucleo simmetrico per σ_F^2 è dato da

$$T_s(X_i, X_j) = \frac{1}{2}\,(T(X_i, X_j) + T(X_j, X_i)) = \frac{1}{2}\,(X_i - X_j)^2, \quad i, j = 1, \ldots, n \quad (i \neq j).$$

DEFINIZIONE 9.7 (*U*-statistica)**.** Sia ξ_F un parametro stimabile di grado ν e sia $X_1, \ldots, X_n$ un campione di ampiezza $n \geq \nu$. Si dice *U-statistica* per il campione la v.a. data da

$$U(X_1, \ldots, X_n) = \frac{1}{\binom{n}{\nu}} \sum_{\mathfrak{C}} T_s(X_{i_1}, \ldots, X_{i_\nu}), \tag{9.4}$$

dove la somma è sull'insieme $\mathfrak{C}$ di tutte le $\binom{n}{\nu}$ combinazioni di ν indici scelti nell'insieme $\{1, \ldots, n\}$.

OSSERVAZIONE 9.1.

Evidentemente la U-statistica della *Definizione (9.7)* è simmetrica nelle v.a. $X_1, \ldots, X_n$; inoltre $\mathbf{E}_F(U) = \xi_F$ per ogni $F \in \mathfrak{F}$. Si noti che una U-statistica potrebbe essere scritta quale funzione delle statistiche d'ordine, in quanto è possibile dimostrare che una v.a. è simmetrica nelle osservazioni se, e solo se, può essere calcolata come funzione delle statistiche d'ordine.

Considerando un campione $X_1, \ldots, X_n$, in base alla *Definizione (9.7)* si ha che:

1. una U-statistica per la stima di μ_F (quando quest'ultima esiste) è

$$U(\boldsymbol{X}) = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}_n;$$

2. una U-statistica per la stima di σ^2_F (quando quest'ultima esiste) è

$$U(\boldsymbol{X}) = \frac{1}{\binom{n}{2}} \sum_{i_1 < i_2} \frac{1}{2} \left(X_{i_1} - X_{i_2}\right)^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}_n\right)^2 = \overline{S}^2_n.$$

Il risultato che segue mostra l'importanza delle U-statistiche.

TEOREMA 9.1 (Stimatore UMVU)**.** *Sia $\mathfrak{F}$ la famiglia di tutte le f.r. su $\mathbb{R}$ che siano o assolutamente continue o puramente discrete. Allora per il parametro stimabile ξ_F, $F \in \mathfrak{F}$, esiste un unico stimatore non distorto e simmetrico nelle osservazioni di varianza uniformemente minima nella classe degli stimatori non distorti di quadrato sommabile di ξ_F.*

Quale importante conseguenza si ha il seguente risultato.

COROLLARIO 9.1. *Se T è uno stimatore non distorto di ξ_F, $F \in \mathfrak{F}$, la corrispondente U-statistica è uno stimatore (essenzialmente unico) di varianza uniformemente minima nella classe degli stimatori non distorti di quadrato sommabile di ξ_F.*

Dunque, nella ricerca di stimatori UMVU, è sufficiente considerare stimatori che siano simmetrici nelle osservazioni e quindi renderli non distorti. Ad esempio, per quanto già visto, $\overline{X}_n$ e $\overline{S}^2_n$ sono gli unici stimatori UMVU, rispettivamente, della media e della varianza di una popolazione.

ESEMPIO 9.3. ▶

Sia $\mathfrak{F}$ la famiglia di tutte le f.r. su $\mathbb{R}$ che siano assolutamente continue e sia $X_1, \ldots, X_n$ un campione di ampiezza n. Per stimare $\pi_c = \mathbf{P}_F\{X > c\}$, dove $c \in \mathbb{R}$ è una costante nota, si definiscano le v.a. (nuclei)

$$Y_i = \mathbb{I}_{(c,\infty)}(X_i), \quad i = 1, \ldots, n,$$

e si consideri la statistica $T = \sum_{i=1}^{n} a_i\, Y_i$ quale stimatore di π_c, dove le a_i sono costanti reali. Per trovare uno stimatore UMVU è necessario rendere T simmetrico nella v.a. Y_i; ciò è possibile prendendo $a_i = a$, $i = 1, \ldots, n$, e $T = a \sum_{i=1}^{n} Y_i$. Affinché T sia corretto deve essere

$$\mathbf{E}_F(T) = a \sum_{i=1}^{n} \mathbf{E}_F(Y_i) = a\, n\, \pi_c = \pi_c,$$

così che $a = \frac{1}{n}$. Dunque $T = \frac{1}{n} \sum_{i=1}^{n} Y_i = \overline{Y}_n$ è lo stimatore UMVU di π_c. Si noti che $Y_i \sim \mathtt{Be}(\pi_c)$. ◀

9.2 I test del Chi-quadro

Sia $\boldsymbol{X} = (X_1, \ldots, X_n)$ un campione costituito da v.a. i.i.d. discrete che assumono valori nell'insieme finito $\{x_1, \ldots, x_k\}$: si pensi, ad esempio, all'esito del lancio di una moneta o di un dado. La legge di ogni v.a. X_i del campione è nota quando si conosca il vettore di probabilità $\boldsymbol{q} = \{q_1, \ldots, q_k\}$, dove $0 \leq q_j = \mathbf{P}\{X_i = x_j\} \leq 1$, $j = 1, \ldots, k$, e naturalmente $\sum_{j=1}^{k} q_j = 1$. Sia ora assegnato un vettore di probabilità $\boldsymbol{p} = \{p_1, \ldots, p_k\}$ di riferimento, nel quale sia $0 < p_j < 1$ per ogni indice j. Il test del Chi-quadro fu introdotto per saggiare se il vettore $\boldsymbol{q}$ relativo al campione coincida (o, meglio, sia statisticamente compatibile) oppure no con il vettore $\boldsymbol{p}$ di riferimento; esso si basa sulla statistica

$$T_n = \sum_{j=1}^{k} \frac{\left(N_j^{(n)} - np_j\right)^2}{np_j}, \tag{9.5a}$$

dove

$$N_j^{(n)} = \sum_{i=1}^{n} \mathbb{I}_{\{X_i = x_j\}} \tag{9.5b}$$

rappresenta il numero di osservazioni nel campione $\boldsymbol{X}$ uguali a x_j. Il risultato che segue è fondamentale.

TEOREMA 9.2. *Sia $\{X_n\}$, $n \in \mathbb{N}$, una successione di v.a. i.i.d. discrete di legge $\boldsymbol{q} = \{q_1, \ldots, q_k\}$, e sia $\boldsymbol{p} = \{p_1, \ldots, p_k\}$ un vettore di probabilità assegnato. Allora valgono i seguenti risultati per $n \to \infty$:*

1. *se $\boldsymbol{q} \neq \boldsymbol{p}$ si ha $T_n \xrightarrow{q.c.} +\infty$;*
2. *se $\boldsymbol{q} = \boldsymbol{p}$ si ha $T_n \xrightarrow{\mathcal{L}} \chi^2(k-1)$.*

Siamo ora in grado di formulare il *test del Chi-quadro*. Assegnato il livello $\alpha \in (0, 1)$, sia n l'ampiezza del campione considerato, siano $\{x_1, \ldots, x_k\}$ i valori che le v.a. del campione possono assumere e siano $\hat{\boldsymbol{q}} = \{\hat{q}_1, \ldots, \hat{q}_k\}$ le frequenze osservate (cioè, in pratica, le stime dei valori q_j per ogni indice j), ovvero

$$\hat{q}_j = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{\{X_i = x_j\}} = \frac{N_j^{(n)}}{n}, \quad j = 1, \ldots, k.$$

Sia infine t_n il valore assunto dalla statistica T_n dell'*equazione (9.5a)*, ossia

$$t_n = n \sum_{j=1}^{k} \frac{(\hat{q}_j - p_j)^2}{p_j}.$$

Se $t_n \geq \chi^2_{k-1,1-\alpha}$ allora si rifiuterà, al livello α, l'ipotesi $\mathcal{H}_0$ che sia $\boldsymbol{q} = \boldsymbol{p}$, altrimenti la si accetterà.

NOTA 9.1.

Nella pratica (si vedano, ad esempio, [10, 2, 17]) in genere si richiede che l'ampiezza del campione sia almeno $n \geq 50$, e inoltre che sia $N_j^{(n)} \geq 5$ per ogni indice $j = 1, \ldots, k$ (ovvero, che ci siano almeno cinque osservazioni per ciascuno dei valori x_j ammissibili).

Gli Esempi che seguono illustrano alcune applicazioni del test del Chi-quadro. Tale test è, in generale, di facile utilizzo e lo si può applicare in tutte le situazioni nelle quali il risultato di un esperimento statistico appartenga solamente ad uno di k eventi disgiunti $A_1, \ldots, A_k$. I quantili di interesse si possono trovare su apposite Tavole (si veda, ad esempio, [25]) oppure calcolare tramite specifici programmi.

ESEMPIO 9.4 (Buffon (1707–1788))**.** ▶

Lo scienziato Georges Louis Leclerc Comte de *Buffon*, noto per i suoi studi ante litteram nel campo della Teoria della Probabilità, eseguì un esperimento lanciando $n = 4040$ volte un moneta, ottenendo testa $n_1 = 2048$ volte e croce $n_2 = 1992$ volte. Si vuole stabilire se questi dati confermino o no l'ipotesi $\mathcal{H}_0$ che la moneta lanciata sia "equilibrata"; in queso caso la legge di riferimento è $\boldsymbol{p} = \left(\frac{1}{2}, \frac{1}{2}\right)$. Il calcolo di T_n fornisce

$$t_n = \frac{(2048 - 2020)^2}{2020} + \frac{(1992 - 2020)^2}{2020} \approx 0.78.$$

Al livello $\alpha = 0.05$ si ha, dalle Tavole, $\chi^2_{1,0.95} \approx 3.84$, e dunque l'ipotesi $\mathcal{H}_0$ è da accettare. ◀

ESEMPIO 9.5 (Mendel (1823–1884)). ►
Nel corso dei suoi famosi esperimenti, Gregor *Mendel* (precursore dei moderni genetisti) osservò le frequenze dei vari semi prodotti da ibridi di piselli gialli lisci e piselli verdi rugosi, al fine di verificare le previsioni formulate dalla nascente Teoria dell'Ereditarietà. Nella *Tabella (9.1)* sono raccolte le osservazioni ottenute. Si vuole stabilire se questi dati confermino o no l'ipotesi $\mathcal{H}_0$ che i risultati delle $n = 556$ osservazioni siano in accordo con la teoria; in queso caso la legge di riferimento è $\boldsymbol{p} = \left(\frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16}\right)$.

(a)

	Colore	
Buccia	**Giallo**	**Verde**
Liscia	9/16	3/16
Rugosa	3/16	1/16

(b)

Tipo di semi	**Osservazioni**
Lisci e gialli	315
Rugosi e gialli	101
Lisci e verdi	108
Rugosi e verdi	32

Tabella 9.1: (*a*) frequenze previste; (*b*) osservazioni.

Il calcolo di T_n fornisce

$$t_n = \frac{(315 - 312.75)^2}{312.75} + \frac{(101 - 104.25)^2}{104.25} + \frac{(108 - 104.25)^2}{104.25} + \frac{(32 - 34.75)^2}{34.75} \approx 0.47.$$

Ai livelli $\alpha = 0.05, 0.01$ si ha, dalle Tavole, $\chi^2_{3,0.95} \approx 7.82$ e $\chi^2_{3,0.99} \approx 11.3$. Dunque l'ipotesi $\mathcal{H}_0$ è da accettare ad entrambi i livelli considerati. ◄

L'*Esempio (9.5)* mostra che il test del Chi-quadro si può applicare anche in situazioni nelle quali le osservazioni non siano di tipo numerico ma, invece, di natura qualitativa.

ESEMPIO 9.6. ►
In un ospedale, per $n = 2880$ nascite è stato registrato l'orario dei parti naturali. La *Tabella (9.2)* mostra i dati raccolti, suddivisi in 24 classi corrispondenti all'ora della nascita.

Ora	**# nascite**	**Ora**	**# nascite**	**Ora**	**# nascite**	**Ora**	**# nascite**
0	127	6	129	12	130	18	99
1	139	7	113	13	125	19	97
2	143	8	126	14	112	20	100
3	138	9	122	15	97	21	119
4	134	10	121	16	115	22	127
5	115	11	119	17	94	23	139

Tabella 9.2: i dati dell'*Esempio (9.6)*.

Si vuole stabilire se questi dati confermino o no l'ipotesi $\mathcal{H}_0$ che le nascite siano uniformemente distribuite nell'arco della giornata; in questo caso la legge di riferimento è $\boldsymbol{p} = \left(\frac{1}{24}, \ldots, \frac{1}{24}\right)$ — si veda la *Figura (9.1)*.

Essendo $np_j = 120$, $j = 0, \ldots, 23$, il calcolo di T_n fornisce

$$t_n = \sum_{j=0}^{23} \frac{\left(N_j^{(2880)} - 120\right)^2}{120} = \frac{4956}{120} \approx 41.3.$$

Ai livelli $\alpha = 0.025, 0.01$ si ha, dalle Tavole, $\chi^2_{23,0.975} \approx 38.1$ e $\chi^2_{23,0.99} \approx 41.6$. Dunque l'ipotesi $\mathcal{H}_0$ è da rifiutare al livello $\alpha = 0.025$ e da accettare al livello $\alpha = 0.01$. ◄

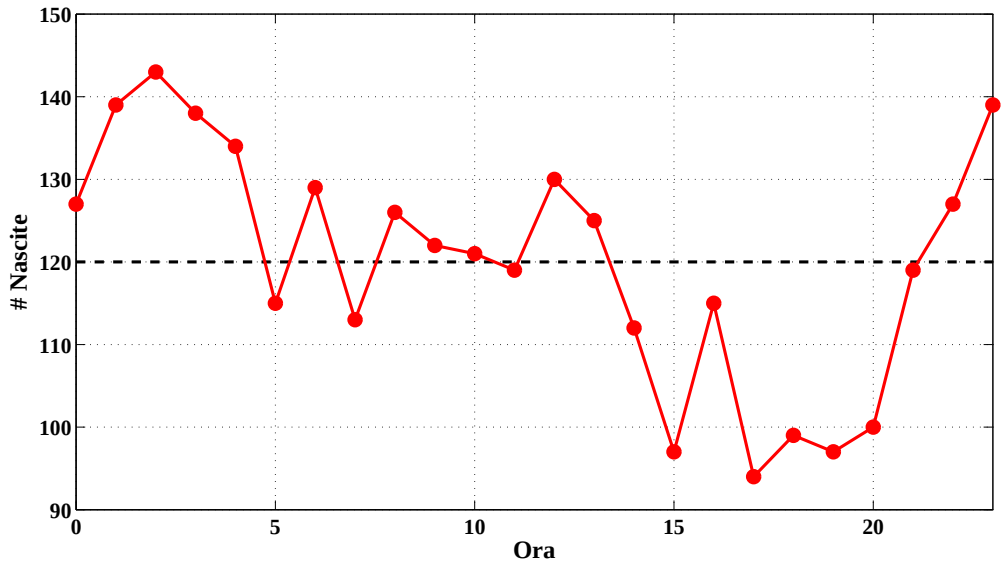

Figura 9.1: i dati dell'*Esempio (9.6)*. La *linea tratteggiata* indica il numero di nascite attese nell'ipotesi di distribuzione Uniforme, la *linea continua* quelle effettivamente osservate.

OSSERVAZIONE 9.2 (Discretizzazione)**.**

Nella pratica il test del Chi-quadro si può usare anche quando la legge di riferimento sia assolutamente continua; in tal caso si partiziona il supporto della densità $f\colon \mathbb{R} \to \mathbb{R}_+$ in k intervalli disgiunti $I_1, \ldots, I_k$ tali che sia $\text{supp}(f) = \bigcup_{j=1}^{k} I_j$, e si pone

$$p_j = \int_{I_j} f(x)\ \mathrm{d}x, \quad j = 1, \ldots, k,$$

usando poi la legge $\boldsymbol{p} = (p_1, \ldots, p_k)$ così ottenuta quale legge di riferimento. Evidentemente vi è un margine di ambiguità legato alla scelta del numero k di intervalli e, una volta che si sia scelto k, alla determinazione degli intervalli stessi: ad esempio, nelle applicazioni si sceglie spesso una partizione ad intervalli *equiprobabili*.

Considerando l'*Esempio (9.6)*, sarebbe stato naturale effettuare una verifica di ipotesi usando quale modello per l'orario delle nascite la legge (assolutamente continua) Uniforme sulle 24 ore. Tuttavia, forse per esigenze di interpretazione, i dati sono stati arbitrariamente aggregati su scala oraria, ciò che corrisponde ad una operazione di discretizzazione.

Se, invece, la v.a. X è solo continua, ma non assolutamente continua, si può procedere adattando il procedimento appena delineato, partendo dalla funzione di ripartizione continua F di X. Infatti, si scelgano $k-1$ punti di $\mathbb{R}$, ossia $-\infty < a_1 < \cdots < a_{k-1} < +\infty$, si consideri la partizione di $\mathbb{R}$ nei k intervalli disgiunti

$$(-\infty, a_1), (a_1, a_2), \ldots, (a_{k-2}, a_{k-1}), (a_{k-1}, +\infty),$$

e si ponga $p_1 = F(a_1)$, $p_j = F(a_j) - F(a_{j-1})$ per $j = 2, \ldots, k-1$, e $p_k = 1 - F(a_{k-1})$. Il vettore $\boldsymbol{p}$ così ottenuto può essere assunto quale vettore di probabilità di riferimento. Naturalmente, anche in questo caso sussistono tutte le ambiguità legate alla scelta di k e dei k intervalli.

Le discretizzazioni appena mostrate sono ovviamente estendibili a f.r. multidimensionali, ciò che rende il test del Chi-quadro di grande utilità ed interesse nelle applicazioni. Si noti che tale test può anche essere utilizzato in fase di stima dei parametri della distribuzione in gioco.

9.2.1 Test del Chi-quadro per l'indipendenza

Sia dato un campione di ampiezza n di vettori bidimensionali indipendenti $((X_1, Y_1), \ldots, (X_n, Y_n))$ tutti di legge F, una f.r. bidimensionale. Qui il problema è quello di vagliare l'ipotesi $\mathcal{H}_0$ che il vettore (X, Y) abbia *componenti indipendenti*, ovvero che sia, per ogni $(x, y) \in \mathbb{R}^2$,

$$F(x, y) = F_X(x)F_Y(y), \tag{9.6}$$

dove F_X e F_Y sono f.r. unidimensionali, o, equivalentemente, che sia $F = F_X \otimes F_Y$. Nel vagliare l'ipotesi di indipendenza tra X e Y è rilevante solo il fatto che F si fattorizzi nel prodotto di due f.r. unidimensionali, ovvero che valga l'*equazione (9.6)*; non si richiede invece che essa si fattorizzi nel prodotto di due *particolari* f.r. F_X e F_Y. Di conseguenza, queste ultime sono da considerarsi incognite, e per superare tale difficoltà esse sono usualmente stimate con il metodo della Massima Verosimiglianza, come illustrato nel seguito.

Si supponga che le componenti X e Y siano entrambe v.a. discrete che assumono un numero finito di valori: $x_1, \ldots, x_r$ per la v.a. X e $y_1, \ldots, y_c$ per la v.a. Y. Se il modello originale ha una struttura differente si penserà di raggruppare i valori di X e/o di Y nel modo appena indicato, eventualmente discretizzando come illustrato nell'*Osservazione (9.2)*. In definitiva, il vettore (X, Y) assumerà gli rc valori (x_i, y_j), con $i = 1, \ldots, r$ e $j = 1, \ldots, c$. I risultati delle osservazioni possono essere convenientemente rappresentati mediante una cosiddetta *Tabella di contingenza*:

$\boldsymbol{X, Y}$	$\boldsymbol{y_1}$	$\boldsymbol{y_2}$	$\cdots$	$\boldsymbol{y_c}$	
$\boldsymbol{x_1}$	n_{11}	n_{12}	$\cdots$	n_{1c}	$n_{1\bullet}$
$\boldsymbol{x_2}$	n_{21}	n_{22}	$\cdots$	n_{2c}	$n_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\boldsymbol{x_r}$	n_{r1}	n_{r2}	$\cdots$	n_{rc}	$n_{r\bullet}$
	$n_{\bullet 1}$	$n_{\bullet 2}$	$\cdots$	$n_{\bullet c}$	n

nella quale

- gli elementi n_{ij}, con $i = 1, \ldots, r$ e $j = 1, \ldots, c$, dati da

$$n_{ij} = \sum_{k=1}^{n} \mathbb{I}_{\{X_k = x_i\} \cap \{Y_k = y_j\}} \tag{9.7}$$

  contano il numero di volte nelle quali, nelle n osservazioni, si riscontra la coppia (x_i, y_j);

- gli elementi $n_{i\bullet}$, $i = 1, \ldots, r$, dati da

$$n_{i\bullet} = \sum_{j=1}^{c} n_{ij} \tag{9.8}$$

  rappresentano la somma degli elementi della i-esima riga, e contano il numero di volte nelle quali, nelle n osservazioni, si riscontra il valore x_i;

- gli elementi $n_{\bullet j}$, $j = 1, \ldots, c$, dati da

$$n_{\bullet j} = \sum_{i=1}^{r} n_{ij} \tag{9.9}$$

  rappresentano la somma degli elementi della j-esima colonna, e contano il numero di volte nelle quali, nelle n osservazioni, si riscontra il valore y_j.

Ovviamente vale

$$\sum_{i=1}^{r} \sum_{j=1}^{c} n_{ij} = \sum_{i=1}^{r} n_{i\bullet} = \sum_{j=1}^{c} n_{\bullet j} = n. \tag{9.10}$$

NOTA 9.2 (Notazione)**.**

La notazione "$\bullet$" indica che si è operata una somma sull'indice soppresso.

Ora, se $\mathbf{P}$ è la legge su $\big(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2)\big)$ indotta dalla f.r. F, si ponga

$$p_{ij} = \mathbf{P}\{X = x_i, Y = y_j\}, \qquad i = 1, \ldots, r \quad \text{e} \quad j = 1, \ldots, c.$$

L'ipotesi $\mathcal{H}_0$ di indipendenza significa che esistono $r + c$ costanti positive $p_{i\bullet}$ e $p_{\bullet j}$, con $i = 1, \ldots, r$ e $j = 1, \ldots, c$, tali che sia

$$\sum_{i=1}^{r} p_{i\bullet} = \sum_{j=1}^{c} p_{\bullet j} = 1 \tag{9.11a}$$

e

$$p_{ij} = p_{i\bullet} p_{\bullet j}, \qquad i = 1, \ldots, r \quad \text{e} \quad j = 1, \ldots, c. \tag{9.11b}$$

In pratica le probabilità p_{ij} non sono note; tuttavia, grazie all'*equazione (9.11b)*, è possibile rimpiazzarle con i valori $p_{i\bullet} p_{\bullet j}$, che possono essere stimati. In particolare, la stima dei parametri incogniti $p_{i\bullet}$ e $p_{\bullet j}$ si può effettuare calcolando i rispettivi SMV che, se è vera l'ipotesi $\mathcal{H}_0$, risultano essere (utilizzando, ad esempio, il metodo dei *moltiplicatori di Lagrange*)

$$\begin{cases} \hat{p}_{i\bullet} = n_{i\bullet}/n, & i = 1, \ldots, r \\ \hat{p}_{\bullet j} = n_{\bullet j}/n, & j = 1, \ldots, c \end{cases}. \tag{9.12}$$

In sostanza, il test si esegue confrontando i valori osservati n_{ij} mostrati nella Tabella di contingenza con quelli "attesi" raccolti nella seguente tabella:

$\boldsymbol{X, Y}$	$\boldsymbol{y_1}$	$\boldsymbol{y_2}$	$\cdots$	$\boldsymbol{y_c}$
$\boldsymbol{x_1}$	$n\hat{p}_{1\bullet}\hat{p}_{\bullet 1}$	$n\hat{p}_{1\bullet}\hat{p}_{\bullet 2}$	$\ldots$	$n\hat{p}_{1\bullet}\hat{p}_{\bullet c}$
$\boldsymbol{x_2}$	$n\hat{p}_{2\bullet}\hat{p}_{\bullet 1}$	$n\hat{p}_{2\bullet}\hat{p}_{\bullet 2}$	$\ldots$	$n\hat{p}_{2\bullet}\hat{p}_{\bullet c}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\boldsymbol{x_r}$	$n\hat{p}_{r\bullet}\hat{p}_{\bullet 1}$	$n\hat{p}_{r\bullet}\hat{p}_{\bullet 2}$	$\ldots$	$n\hat{p}_{r\bullet}\hat{p}_{\bullet c}$

A causa dei vincoli espressi dall'*equazione (9.11a)*, in realtà solo $(r-1) + (c-1) = r + c - 2$ parametri sono effettivamente stimati. Per verificare l'ipotesi $\mathcal{H}_0$ si adotta la statistica

$$T_n = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(n_{ij} - n\,\hat{p}_{i\bullet}\,\hat{p}_{\bullet j})^2}{n\,\hat{p}_{i\bullet}\,\hat{p}_{\bullet j}} \tag{9.13}$$

che, grazie al *Teorema (9.2)*, asintoticamente ha legge $\chi^2((r-1)(c-1))$, essendo il numero di parametri liberi effettivi pari a $rc - (r + c - 2) - 1 = (r-1)(c-1)$. In definitiva, l'ipotesi $\mathcal{H}_0$ di indipendenza è da rigettarsi, al livello $\alpha \in (0, 1)$, se $t_n \geq \chi^2_{(r-1)(c-1),1-\alpha}$. Nelle applicazioni, una regola empirica afferma che il test sia (ragionevolmente) applicabile quando $\frac{n}{n_{i\bullet}} \approx r$ e $\frac{n}{n_{\bullet j}} \approx c$ per ogni $i = 1, \ldots, r$ e $j = 1, \ldots, c$, e inoltre $\frac{n}{r} \geq 5$ e $\frac{n}{c} \geq 5$.

ESEMPIO 9.7. ►

Il seguente studio è stato commissionato per saggiare se la durata dell'impiego D presso la stessa ditta dipenda o meno dal salario S percepito mensilmente. La *Tabella (9.3)* mostra i dati raccolti durante l'indagine su $n = 400$ impiegati, dove D è espressa in anni e S in Dollari; tra parentesi sono indicate le quantità "attese" $n\hat{p}_{i\bullet}\hat{p}_{\bullet j}$ che ci si aspetterebbe di trovare se l'ipotesi $\mathcal{H}_0$ fosse vera: tali valori sono ottenuti semplicemente moltiplicando tra loro i totali della riga/colonna corrispondente e dividendo per l'ampiezza n del campione.

Il calcolo di T_n fornisce

$$t_n = \frac{12.5^2}{37.5} + 0 + \frac{12.5^2}{37.5} + 0 + 0 + 0 + \frac{12.5^2}{37.5} + 0 + \frac{12.5^2}{37.5} \approx 16.7.$$

Il numero di gradi di libertà è $(3-1)(3-1) = 4$. Al livello $\alpha = 0.05$ si ha, dalle Tavole, $\chi^2_{4,0.95} \approx 9.5$ e dunque l'ipotesi $\mathcal{H}_0$ di indipendenza tra D e S è da rifiutare. Si noti come nemmeno al livello $\alpha = 0.005$ l'ipotesi $\mathcal{H}_0$ si possa accettare, essendo $\chi^2_{4,0.995} \approx 14.9$. ◄

D, S	$< 8000\$$	$8000\$–15000\$$	$> 15000\$$	**Totale**
< 5 anni	50 (37.5)	75 (75)	25 (37.5)	150
5–10 anni	25 (25)	50 (50)	25 (25)	100
> 10 anni	25 (37.5)	75 (75)	50 (37.5)	150
Totale	100	200	100	400

Tabella 9.3: i dati dell'*Esempio (9.7)*.

9.2.2 Test del Chi-quadro per l'omogeneità

Siano dati k campioni indipendenti

$$\left(X_1^{(1)}, \ldots, X_{n^{(1)}}^{(1)}\right), \ldots, \left(X_1^{(k)}, \ldots, X_{n^{(k)}}^{(k)}\right)$$

di ampiezza variabile $n^{(1)}, \ldots, n^{(k)}$. Se i diversi campioni descrivono lo stesso fenomeno, osservato in condizioni o in tempi differenti, si presenta frequentemente il problema di stabilire se questi abbiano o no la stessa legge. A tale fine si può usare il *test del Chi-quadro per l'omogeneità*, assumendo che i dati siano discreti (e in numero finito) o siano stati preliminarmente discretizzati, come già accennato nell'*Osservazione (9.2)*. Un vantaggio dell'uso di tale test è che esso consente di considerare contemporaneamente un numero arbitrario (finito) di campioni di ampiezza variabile.

Si supponga che le componenti X possano assumere uno dei valori $x_1, \ldots, x_s$, e si ponga uguale a $n_j^{(i)}$ il numero di osservazioni del valore x_j nel campione i-esimo; ovviamente vale $\sum_{j=1}^{s} n_j^{(i)} = n^{(i)}$. Sia ora $\boldsymbol{p} = (p_1, \ldots, p_s)$ un vettore di probabilità non noto; l'ipotesi $\mathcal{H}_0$ di omogeneità corrisponde a supporre che sia, usando una notazione ovvia,

$$\left(p_1^{(i)}, \ldots, p_s^{(i)}\right) = \boldsymbol{p}, \quad i = 1, \ldots, k. \tag{9.14}$$

Per verificare l'ipotesi $\mathcal{H}_0$ si adotta la statistica

$$T_n = \sum_{i=1}^{k} \sum_{j=1}^{s} \frac{\left(n_j^{(i)} - n^{(i)}\, p_j\right)^2}{n^{(i)}\, p_j}.$$

Grazie al *Teorema (9.2)*, T_n risulta essere la somma di k v.a. indipendenti, ciascuna (asintoticamente) di legge $\chi^2(s-1)$, e dunque, approssimativamente, $T_n \sim \chi^2(k(s-1))$. Tuttavia, poiché il vettore $\boldsymbol{p}$ non è noto occorre valutare le sue componenti; come già mostrato nella *Sezione 9.2.1*, esse sono usualmente stimate tramite il metodo della Massima Verosimiglianza. In particolare, la stima dei parametri incogniti p_j si può effettuare calcolando i rispettivi SMV che, sotto l'ipotesi $\mathcal{H}_0$, risultano essere

$$\hat{p}_j = \frac{1}{n} \sum_{i=1}^{k} n_j^{(i)}, \quad j = 1, \ldots, s, \tag{9.15}$$

dove si è posto $n = \sum_{i=1}^{k} n^{(i)}$. Dunque, la statistica che si userà è

$$T_n = \sum_{i=1}^{k} \sum_{j=1}^{s} \frac{\left(n_j^{(i)} - n^{(i)}\, \hat{p}_j\right)^2}{n^{(i)}\, \hat{p}_j} \tag{9.16}$$

che, asintoticamente, ha legge $\chi^2((k-1)(s-1))$, essendo il numero di parametri liberi effettivi pari a $k(s-1) - (s-1) = (k-1)(s-1)$. In definitiva, l'ipotesi $\mathcal{H}_0$ di omogeneità è da rigettarsi, al livello $\alpha \in (0,1)$, se $t_n \geq \chi^2_{(k-1)(s-1),1-\alpha}$. Nelle applicazioni, una regola empirica afferma che il test sia (ragionevolmente) applicabile quando $\sum_{i=1}^{k} n_j^{(i)} \geq 5$ per ogni $j = 1, \ldots, s$.

$\boldsymbol{X^{(1)}}$			$\boldsymbol{X^{(2)}}$			$\boldsymbol{X^{(3)}}$		
0.0835	0.5827	0.7639	0.4623	0.1524	0.7204	0.4641	0.0464	0.7573
0.1958	0.6854	0.6655	0.0483	0.2033	0.0913	0.0713	0.9519	0.1148
0.9776	0.9671	0.8933	0.4604	0.8193	0.7333	0.5812	0.1690	0.5752
0.3662	0.5503	0.9125	0.8000	0.0584	0.5068	0.5660	0.8267	0.4734
0.1394	0.7772	0.1022	0.2894	0.5385	0.6223	0.2553	0.6114	0.4081
0.0148	0.6104	0.2047	0.6951	0.1902	0.8841	0.2385	0.8473	0.6832
0.6406	0.9886	0.0268	0.2593	0.5995	0.9898	0.0160	0.1141	0.1957
0.7377	0.0483	0.9854	0.7132	0.2923	0.6156	0.3847	0.6492	0.1333
$n^{(1)} = 24$			$n^{(2)} = 24$			$n^{(3)} = 24$		

Tabella 9.4: i campioni dell'*Esempio (9.8)*.

ESEMPIO 9.8. ►
Un software per la generazione di numeri casuali ha fornito i tre campioni estratti da una distribuzione $\mathtt{U_c}(0,1)$ mostrati nella *Tabella (9.4)* e nella *Figura (9.2)* — si veda anche l'*Esempio (9.9)*.

Si vuole verificare l'ipotesi $\mathcal{H}_0$ che tutti i campioni abbiano effettivamente la stessa legge. A tale fine si suddivide l'intervallo $[0,1]$ in quattro sotto-intervalli disgiunti di uguale ampiezza

$$I_1 = [0, 0.25), \quad I_2 = [0.25, 0.5), \quad I_3 = [0.5, 0.75), \quad I_4 = [0.75, 1],$$

e si calcola, per ciascun campione, il numero di osservazioni che cadono nei diversi sotto-intervalli. La tabella che segue mostra i risultati dell'analisi, una volta effettuata la discretizzazione dei dati; tra parentesi sono indicate le quantità "attese" che ci si aspetterebbe di trovare se l'ipotesi $\mathcal{H}_0$ fosse vera.

	$\boldsymbol{X^{(1)}}$	$\boldsymbol{X^{(2)}}$	$\boldsymbol{X^{(3)}}$
$\boldsymbol{I_1}$	8 (6)	6 (6)	9 (6)
$\boldsymbol{I_2}$	1 (6)	5 (6)	5 (6)
$\boldsymbol{I_3}$	7 (6)	9 (6)	6 (6)
$\boldsymbol{I_4}$	8 (6)	4 (6)	4 (6)

Il calcolo di T_n fornisce $t_n \approx 10.3$. Il numero di gradi di libertà è $(3-1)(4-1) = 6$. Al livello $\alpha = 0.05$ si ha, dalle Tavole, $\chi^2_{6,0.95} \approx 12.6$ e dunque l'ipotesi $\mathcal{H}_0$ è da accettare. Si noti come anche al livello $\alpha = 0.1$ l'ipotesi $\mathcal{H}_0$ si possa accettare, essendo $\chi^2_{6,0.9} \approx 10.6$. ◄

9.3 I test di Kolmogorov–Smirnov

I test del Chi-quadro si applicano quando la f.r. di riferimento sia discreta, o sia stata opportunamente discretizzata. Quando, invece, questa sia continua (ma non necessariamente assolutamente continua) si può ricorrere ai *test di Kolmogorov–Smirnov*. Prima di procedere è utile richiamare la definizione di *funzione di ripartizione empirica* (o *campionaria*), che rappresenta uno stimatore non parametrico della f.r. vera ma incognita.

DEFINIZIONE 9.8 (Funzione di ripartizione empirica)**.** Sia $\mathbf{X}_F^n$ un campione. La funzione

$$F_n^*(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{(-\infty,x]}(X_i) \tag{9.17}$$

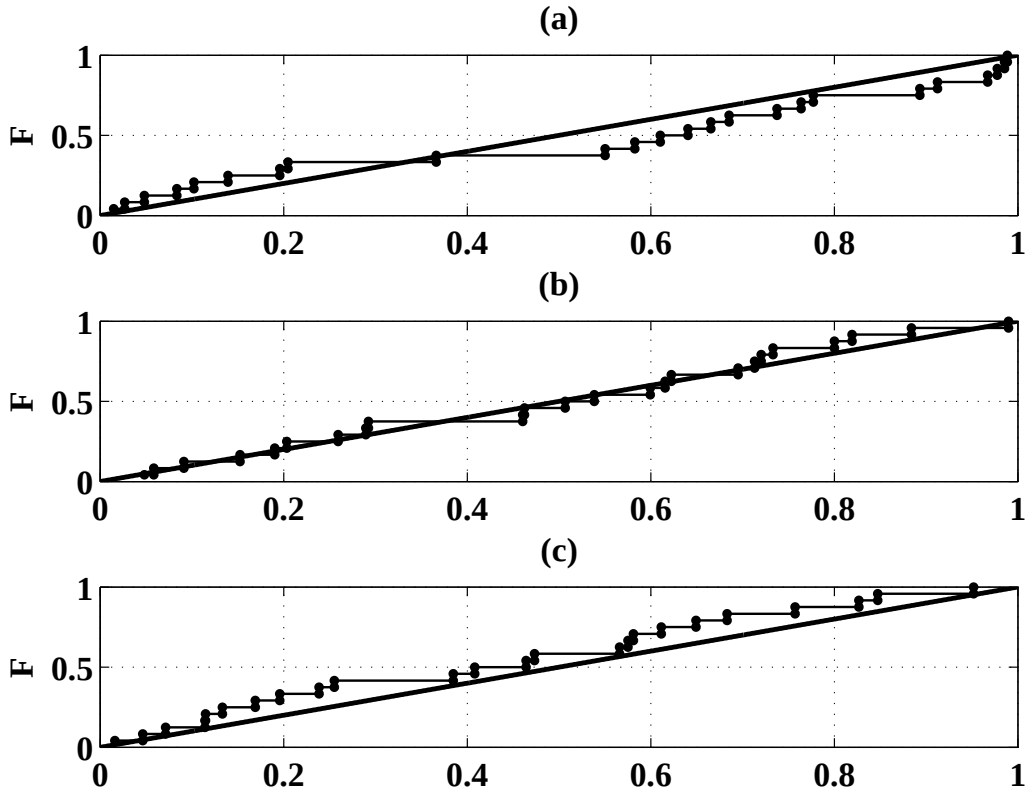

Figura 9.2: i campioni dell'*Esempio (9.8).*

è detta *funzione di ripartizione empirica* associata al campione. Se $X_{(1)}, \ldots, X_{(n)}$ denotano le statistiche d'ordine del campione, allora

$$F_n^*(x) = \begin{cases} 0, & \text{se } x < X_{(1)} \\ \frac{k}{n}, & \text{se } X_{(k)} \leq x < X_{(k+1)}, k = 1, \ldots, n-1 \\ 1, & \text{se } x \geq X_{(n)} \end{cases} . \tag{9.18}$$

NOTA 9.3.

La funzione di ripartizione empirica rappresenta un contatore (riscalato). Si osservi che, per $x \in \mathbb{R}$ fissato, $F_n^*(x)$ è una variabile aleatoria, in quanto somma (riscalata) di v.a. indipendenti di legge $\texttt{Be}(F(x))$.

Il Teorema che segue è fondamentale.

TEOREMA 9.3 (Glivenko–Cantelli)**.** *La f.r. empirica F_n^* converge uniformemente a F, cioè, per ogni $\epsilon > 0$,*

$$\lim_{n \to \infty} \mathbf{P}\left\{ \sup_{x \in \mathbb{R}} |F_n^*(x) - F(x)| > \epsilon \right\} = 0, \tag{9.19}$$

dove $\mathbf{P}$ è la legge di probabilità indotta da F su $(\mathbb{R}, \mathcal{B})$.

9.3.1 Il test di adattamento di Kolmogorov–Smirnov

Sia $\{X_n\}$, $n \in \mathbb{N}$, una successione di v.a. i.i.d.. Si voglia inferire se la legge comune F delle v.a. della successione coincida oppure no con una f.r. di riferimento *continua* F_0, e ciò lo si voglia effettuare mediante l'osservazione

della realizzazione del campione $(X_1, \ldots, X_n)$. Nel seguito si suppone che la f.r. F sia continua. Qui l'ipotesi nulla è $\mathcal{H}_0 \colon X \sim F_0$ contro $\mathcal{H}_1 \colon X \sim F$ e $F(x) \neq F_0(x)$ per qualche $x \in \mathbb{R}$. Se si vogliono paragonare le f.r. F e F_0 (in inglese, *Goodness-of-Fit test*, ovvero "test di adattamento") è necessario introdurre alcune statistiche opportune. Nel seguito la sigla "KS" farà da abbreviazione a "Kolmogorov–Smirnov".

DEFINIZIONE 9.9 (Statistiche KS)**.** Sia $\mathbf{X}_F^n$ un campione e sia F_n^* la corrispondente f.r. empirica. La v.a.

$$D_n = \sup_{x \in \mathbb{R}} |F_n^*(x) - F(x)| \tag{9.20}$$

è detta *statistica* KS *a due code*. Analogamente si definiscono le v.a.

$$D_n^+ = \sup_{x \in \mathbb{R}} (F_n^*(x) - F(x)) \tag{9.21a}$$

e

$$D_n^- = \sup_{x \in \mathbb{R}} (F(x) - F_n^*(x)) , \tag{9.21b}$$

dette *statistiche* KS *a una coda*.

Evidentemente $D_n = \max \{D_n^-, D_n^+\}$.

NOTA 9.4.

Quale conseguenza del *Teorema (9.3)* si ha che, per $n \to \infty$, $D_n \xrightarrow{\text{q.c.}} 0$.

Il risultato che segue è importante e mette in luce una proprietà fondamentale delle statistiche KS.

PROPOSIZIONE 9.3. *La legge delle statistiche* KS D_n, D_n^+ *e* D_n^- *non dipende da* F *(purché questa sia continua).*

Le espressioni analitiche delle leggi delle statistiche KS sono estremamente complesse; esistono, tuttavia, alcuni risultati asintotici, di grande utilità nelle applicazioni.

TEOREMA 9.4 (Kolmogorov)**.** *Sia* F *una f.r. continua. Allora vale, per ogni* $x > 0$*,*

$$\lim_{n \to \infty} \mathbf{P}\{\sqrt{n}\, D_n \leq x\} = 1 - 2 \sum_{i=1}^{\infty} (-1)^{i-1} e^{-2i^2 x^2}. \tag{9.22}$$

Il *Teorema (9.4)* può essere usato per calcolare, almeno in via approssimata, la legge di D_n, e quindi i quantili $d_{n,\alpha}$ per $\alpha \in (0, 1)$; la formula fornisce buoni risultati già per $n > 20$ (alcuni valori sono comunque riportati sulle Tavole per combinazioni selezionate dei parametri n, α).

TEOREMA 9.5 (Smirnov)**.** *Sia* F *una f.r. continua. Allora vale, per ogni* $x > 0$*,*

$$\lim_{n \to \infty} \mathbf{P}\{\sqrt{n}\, D_n^+ \leq x\} = 1 - e^{-2x^2}. \tag{9.23}$$

OSSERVAZIONE 9.3.

Per ragioni di simmetria, il risultato del *Teorema (9.5)* vale sia per D_n^+ sia per D_n^-. Inoltre si ha che, asintoticamente, la v.a.

$$V_n = 4\, n \, \left(D_n^+\right)^2 \tag{9.24}$$

ha legge $\chi^2(2)$: infatti, scende dall'*equazione (9.23)* che $\lim_{n \to \infty} \mathbf{P}\{V_n \leq 4x^2\} = 1 - e^{-2x^2}$, $x > 0$, e dunque $\lim_{n \to \infty} \mathbf{P}\{V_n \leq x\} = 1 - e^{-\frac{x}{2}}$, che corrisponde alla f.r. di una v.a. di legge $\chi^2(2)$.

Anche il *Teorema (9.5)* può essere usato per calcolare, almeno in via approssimata, la legge di D_n^+ e D_n^-, e quindi i quantili $d_{n,\alpha}^+$ e $d_{n,\alpha}^-$, $\alpha \in (0, 1)$; la formula fornisce buoni risultati già per $n > 20$ (alcuni valori sono comunque riportati sulle Tavole per combinazioni selezionate dei parametri n, α).

L'uso del *test di adattamento di Kolmogorov–Smirnov*, al livello $\alpha \in (0, 1)$, è come segue. Si rifiuta l'ipotesi

$$\mathcal{H}_0 \colon F = F_0 \quad \text{se} \quad d_n > d_{n,1-\alpha}.$$

Analogamente, si rifiuta l'ipotesi

$$\mathcal{H}_0 \colon F \geq F_0 \quad \text{o} \quad \mathcal{H}_0 \colon F \leq F_0 \quad \text{se, rispettivamente,} \quad d_n^- > d_{n,1-\alpha}^+ \quad \text{o} \quad d_n^+ > d_{n,1-\alpha}^+.$$

Si noti che, sotto opportune condizioni, esistono versioni multivariate del *test di Kolmogorov–Smirnov* [27].

ESEMPIO 9.9. ►

Un software per la generazione di numeri casuali ha fornito un campione $\boldsymbol{X}$ di ampiezza $n = 20$. Si vuole verificare l'ipotesi $\mathcal{H}_0$ che il campione abbia legge $F \sim \mathtt{N}(0,1)$. A tale fine si calcola la corrispondente f.r. empirica F^*_{20} e la si confronta con la f.r. di riferimento $F_0 \sim \mathtt{N}(0,1)$. La *Tabella (9.5)* mostra le osservazioni $(x_1, \ldots, x_{20})$ ordinate in senso crescente — cioè le statistiche d'ordine $x_{(1)}, \ldots, x_{(20)}$, i corrispondenti valori delle funzioni F^*_{20} e F_0, e i moduli delle differenze tra queste due.

i	X	F^*_{20}	F_0	$\lvert\Delta\rvert$	i	X	F^*_{20}	F_0	$\lvert\Delta\rvert$
1	-2.4701	0.05	0.0068	0.0432	11	0.1543	0.55	0.5613	0.0113
2	-2.3598	0.10	0.0091	0.0909	12	0.2505	0.60	0.5989	0.0011
3	-2.0682	0.15	0.0193	0.1307	13	0.4838	0.65	0.6857	0.0357
4	-1.4057	0.20	0.0799	0.1201	14	0.6232	0.70	0.7334	0.0334
5	-1.3495	0.25	0.0886	**0.1614**	15	0.6303	0.75	0.7358	0.0142
6	-0.9729	0.30	0.1653	0.1347	16	0.9019	0.80	0.8164	0.0164
7	-0.7970	0.35	0.2127	0.1373	17	1.0495	0.85	0.8530	0.0030
8	-0.5943	0.40	0.2762	0.1238	18	1.2486	0.90	0.8941	0.0059
9	-0.5229	0.45	0.3005	0.1495	19	1.5913	0.95	0.9442	0.0058
10	-0.3132	0.50	0.3771	0.1229	20	1.8409	1.00	0.9672	0.0328

Tabella 9.5: i dati dell'*Esempio (9.9)*.

Il calcolo di $D_n = \sup_x |\Delta(x)|$ fornisce $d_n \approx 0.1614$. Al livello $\alpha = 0.05$ si ha, dalle Tavole, $d_{20,0.95} \approx 0.294$ e dunque l'ipotesi $\mathcal{H}_0$ è da accettare. La *Figura (9.3)* mostra sia il grafico di F_0 sia quello di F^*_{20}; si noti il tipico andamento "a scala" di quest'ultima: gli scarti in orizzontale valgono $x_{(i+1)} - x_{(i)}$, mentre quelli in verticale sono tutti pari a $\frac{1}{n} = 0.05$.

A scopo comparativo, si può verificare l'ipotesi $\mathcal{H}_0$ utilizzando il test del Chi-quadro illustrato nella *Sezione 9.2*. Poiché le f.r. coinvolte sono (assolutamente) continue, è necessario anzitutto procedere ad una discretizzazione. A tale fine si considerino gli intervalli disgiunti

$$I_1 = (-\infty, -0.5], \quad I_2 = (-0.5, 0.5], \quad I_3 = [0.5, \infty),$$

e si calcoli il numero di osservazioni che cadono nei diversi intervalli. La tabella che segue mostra i risultati dell'analisi; tra parentesi sono indicate le quantità "attese" che ci si aspetterebbe di trovare se l'ipotesi $\mathcal{H}_0$ fosse vera: tali valori sono ottenuti moltiplicando l'ampiezza $n = 20$ del campione per le probabilità (p_1, p_2, p_3) che ciascuna singola osservazione appartenga all'intervallo corrispondente.

I_1	p_1	I_2	p_2	I_3	p_3
9 (6.17)	0.3085	4 (7.66)	0.3830	7 (6.17)	0.3085

Il calcolo di T_n fornisce

$$t_n = \frac{(9-6.17)^2}{6.17} + \frac{(4-7.66)^2}{7.66} + \frac{(7-6.17)^2}{6.17} \approx 3.16.$$

Il numero di gradi di libertà è $3-1=2$. Al livello $\alpha = 0.05$ si ha, dalle Tavole, $\chi^2_{2,0.95} \approx 5.99$ e dunque l'ipotesi $\mathcal{H}_0$ è da accettare. ◄

L'*Esempio (9.9)* spinge ad alcune considerazioni circa le differenze tra il test del Chi-quadro e quello di adattamento di Kolmogorov–Smirnov nel caso continuo. Quest'ultimo tratta direttamente le singole osservazioni, mentre il primo deve fare ricorso ad una discretizzazione (che, oltre ad essere arbitraria, può portare a perdita di "informazione"). Inoltre, il test KS è applicabile anche per campioni di ampiezza ridotta, mentre il test del

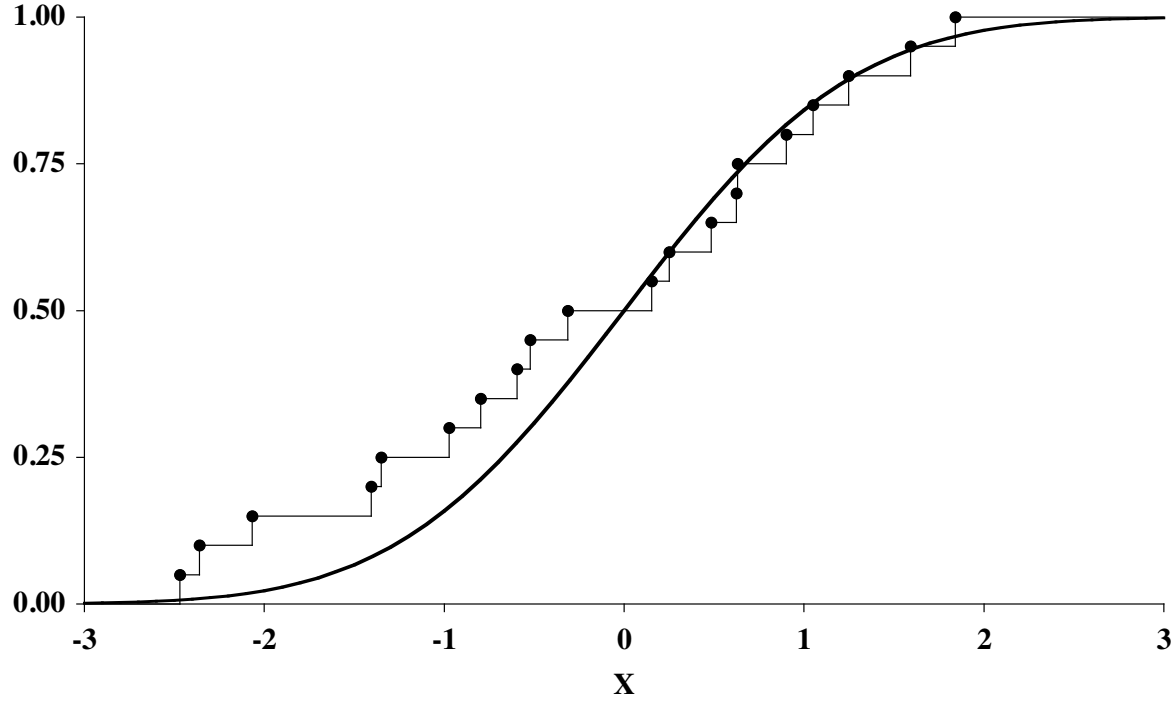

Figura 9.3: grafico delle funzioni F_0 (*linea spessa*) e F^*_{20} (*linea sottile*) dell'*Esempio (9.9).*

Chi-quadro è pensato essenzialmente per campioni di ampiezza considerevole. Tuttavia, quest'ultimo può essere opportunamente modificato così da servire in fase di stima dei parametri di una distribuzione, una flessibilità che il test KS non possiede. Infine, il test del Chi-quadro è applicabile a f.r. sia discrete sia (assolutamente) continue, mentre il test KS assume che la f.r. sia (almeno) continua.

ESEMPIO 9.10. ►

Negli *Esempi (12.1)–(12.3)* del *Capitolo 12*, riguardante la generazione di v.a., sono stati simulati alcuni campioni di ampiezza $n = 30$ estratti, rispettivamente, dalla distribuzione Uniforme (generalizzata), Esponenziale e Pareto Generalizzata. Si vuole verificare l'ipotesi $\mathcal{H}_0$ che tali campioni abbiano effettivamente le leggi desiderate. A tale fine si calcolano le corrispondenti f.r. empiriche e le si confrontano con le f.r. di riferimento come nell'*Esempio (9.9).* La *Tabella (9.6)* mostra i risultati di tale analisi.

Esempio (12.1)	$\mathtt{U_c}(0,1)$	$\mathtt{U_c}(-1,1)$	$\mathtt{U_c}(-1,0)$
	0.192	0.115	0.169
Esempio (12.2)	$\mathtt{e}(1)$	$\mathtt{e}(3)$	$\mathtt{e}(0.1)$
	0.088	0.123	0.147
Esempio (12.3)	$\mathtt{PG}(3,1,-0.3)$	$\mathtt{PG}(0,2,-0.6)$	$\mathtt{PG}(-10,7,-1.5)$
	0.139	0.087	0.150

Tabella 9.6: valori della statistica D_{30} per le simulazioni degli *Esempi (12.1)–(12.3).*

Ai livelli $\alpha = 0.01, 0.05, 0.1$ si ha, dalle Tavole, $d_{30,0.99} \approx 0.290$, $d_{30,0.95} \approx 0.242$ e $d_{30,0.90} \approx 0.218$. Dunque l'ipotesi $\mathcal{H}_0$ è da accettare in ogni caso. ◄

9.3.2 Il test di omogeneità di Kolmogorov–Smirnov

Siano $\mathbf{X}_F^n$ e $\mathbf{Y}_G^m$ due campioni indipendenti. Come già accennato nella *Sezione 9.2.2*, se i due campioni descrivono lo stesso fenomeno, osservato in condizioni o in tempi differenti, si può presentare il problema di stabilire se questi abbiano o no la stessa legge. A tale fine si può usare il *test di omogeneità di Kolmogorov–Smirnov*, assumendo che le leggi F, G in gioco siano entrambe (assolutamente) continue. Anche in questo caso i campioni possono avere ampiezza diversa. L'ipotesi nulla è $\mathcal{H}_0 \colon F = G$. Se si vogliono paragonare le f.r. F e G, ancora una volta si usano le rispettive f.r. empiriche F_n^* e G_m^*, e si introducono alcune statistiche opportune.

DEFINIZIONE 9.10 (Statistiche KS)**.** Siano $\mathbf{X}_F^n$ e $\mathbf{Y}_G^m$ due campioni indipendenti, e siano F_n^* e G_m^* le rispettive f.r. empiriche. La v.a.

$$D_{n,m} = \sup_{t \in \mathbb{R}} |F_n^*(t) - G_m^*(t)| \tag{9.25}$$

è detta *statistica* KS *a due code*. Analogamente si definiscono le v.a.

$$D_{n,m}^+ = \sup_{t \in \mathbb{R}} (F_n^*(t) - G_m^*(t)) \tag{9.26a}$$

e

$$D_{n,m}^- = \sup_{t \in \mathbb{R}} (G_m^*(t) - F_n^*(t)) \tag{9.26b}$$

dette *statistiche* KS *a una coda*.

L'uso del *test di omogeneità di Kolmogorov–Smirnov*, al livello $\alpha \in (0, 1)$, è come segue. Si rifiuta l'ipotesi

$$\mathcal{H}_0 \colon F = G$$

contro l'alternativa $\mathcal{H}_1 \colon F(t) \neq G(t)$ per qualche $t \in \mathbb{R}$ se $d_{n,m} \geq d_{n,m,1-\alpha}$. Analogamente, si rifiuta l'ipotesi $\mathcal{H}_0$ contro l'alternativa

$$\mathcal{H}_1 \colon G(t) \leq F(t) \quad \text{per ogni } t \in \mathbb{R} \text{ e} \quad G(t) < F(t) \quad \text{per qualche } t \in \mathbb{R}$$

se $d_{n,m}^+ \geq d_{n,m,1-\alpha}^+$, e si rifiuta l'ipotesi $\mathcal{H}_0$ contro l'alternativa

$$\mathcal{H}_1 \colon F(t) \leq G(t) \quad \text{per ogni } t \in \mathbb{R} \text{ e} \quad F(t) < G(t) \quad \text{per qualche } t \in \mathbb{R}$$

se $d_{n,m}^- \geq d_{n,m,1-\alpha}^-$.

Le espressioni analitiche delle leggi delle statistiche KS sono estremamente complesse; esistono, tuttavia, alcuni risultati asintotici, di grande utilità nelle applicazioni.

TEOREMA 9.6 (Smirnov)**.** *Nelle ipotesi sin qui specificate valgono i seguenti risultati. Sia* $c = \frac{nm}{n+m}$*; allora, per ogni* $t > 0$*,*

$$\lim_{n,m \to \infty} \mathbf{P}\big\{\sqrt{c}\, D_{n,m} \leq t\big\} = \sum_{i=-\infty}^{\infty} (-1)^i\, e^{-2i^2 t^2}, \tag{9.27a}$$

$$\lim_{n,m \to \infty} \mathbf{P}\big\{\sqrt{c}\, D_{n,m}^+ \leq t\big\} = 1 - e^{-2t^2}. \tag{9.27b}$$

Il *Teorema (9.6)* può essere usato per calcolare, almeno in via approssimata, le leggi e i quantili delle statistiche KS di interesse; alcuni valori sono comunque riportati sulle Tavole per combinazioni selezionate dei parametri n, m, α. Si noti come nel test di omogeneità KS si confrontino solo due campioni alla volta.

9.4 I test di Kendall e Spearman

Sia dato un campione di ampiezza n di vettori bidimensionali indipendenti $((X_1, Y_1), \ldots, (X_n, Y_n))$ tutti di legge F, una f.r. bidimensionale. Come già visto nella *Sezione 9.2.1*, qui il problema è quello di vagliare l'ipotesi $\mathcal{H}_0$ che il vettore (X, Y) abbia *componenti indipendenti*. Nel seguito assumeremo che F sia (assolutamente) continua, così come le due f.r. unidimensionali marginali F_X e F_Y. *En passant*, si noti come le statistiche introdotte nel seguito (cioè la τ *di Kendall* e la ρ_S *di Spearman*) risultino assai utili nella Teoria delle Copule — si veda la *Sezione 12.4*.

9.4.1 Il test di indipendenza di Kendall

Prima di procedere è opportuno introdurre alcune definizioni.

DEFINIZIONE 9.11 (Perfetta concordanza/discordanza)**.** Tra la coppia di vettori (X_i, Y_i) e (X_j, Y_j), $i \neq j$, sussiste una relazione di *perfetta concordanza* quando

$$Y_i < Y_j \Leftrightarrow X_i < X_j \quad \text{oppure} \quad Y_i > Y_j \Leftrightarrow X_i > X_j, \tag{9.28a}$$

e una relazione di *perfetta discordanza* quando

$$Y_i < Y_j \Leftrightarrow X_i > X_j \quad \text{oppure} \quad Y_i > Y_j \Leftrightarrow X_i < X_j. \tag{9.28b}$$

Se π_c e π_d indicano, rispettivamente, le probabilità di perfetta concordanza e discordanza, si ha

$$\pi_c = \mathbf{P}\{(X_j - X_i)(Y_j - Y_i) > 0\} \tag{9.29a}$$

e

$$\pi_d = \mathbf{P}\{(X_j - X_i)(Y_j - Y_i) < 0\}\,. \tag{9.29b}$$

Dunque, essendo le leggi marginali continue,

$$\begin{aligned} \pi_c &= (\mathbf{P}\{Y_i < Y_j\} - \mathbf{P}\{X_i > X_j \quad \text{e} \quad Y_i < Y_j\}) \\ &\quad + (\mathbf{P}\{Y_i > Y_j\} - \mathbf{P}\{X_i < X_j \quad \text{e} \quad Y_i > Y_j\}) \\ &= 1 - \pi_d, \end{aligned} \tag{9.30}$$

come si evince dalla analisi della tabella che segue.

$\mathbf{Y/X}$	$\mathbf{X_i < X_j}$	$\mathbf{X_i > X_j}$
$\mathbf{Y_i < Y_j}$	Concordanza	Discordanza
$\mathbf{Y_i > Y_j}$	Discordanza	Concordanza

Al fine di poter valutare gli aspetti statistici del comportamento congiunto delle componenti X e Y si introduce la seguente *misura di associazione*.

DEFINIZIONE 9.12 ($\boldsymbol{\tau}$ di Kendall)**.** La misura di associazione tra le v.a. X e Y definita come

$$\tau = \pi_c - \pi_d \tag{9.31}$$

è detta τ *di Kendall*.

In pratica τ misura l'eventuale "eccesso" di concordanza/discordanza presente nel campione. Le *Figure (12.7)–(12.10)* illustrano in maniera chiara il concetto.

NOTA 9.5.

In base all'*equazione (9.30)* possiamo riscrivere τ come

$$\tau = 1 - 2\pi_d = 2\pi_c - 1. \tag{9.32}$$

Ovviamente è $-1 \leq \tau \leq 1$, essendo τ una differenza di probabilità. Si osservi inoltre come l'esistenza di τ sia garantita per tutti i vettori (X, Y).

Essendo $\mathbf{P}\{X_i < X_j\} = \mathbf{P}\{X_i > X_j\} = \frac{1}{2}$ (la v.a. $X_i - X_j$ è simmetrica), se X e Y sono indipendenti e continue si ha

$$\begin{aligned} \pi_c &= \mathbf{P}\{X_i < X_j\}\,\mathbf{P}\{Y_i < Y_j\} + \mathbf{P}\{X_i > X_j\}\,\mathbf{P}\{Y_i > Y_j\} \\ &= \mathbf{P}\{X_i > X_j\}\,\mathbf{P}\{Y_i < Y_j\} + \mathbf{P}\{X_i < X_j\}\,\mathbf{P}\{Y_i > Y_j\} \\ &= \pi_d, \end{aligned}$$

da cui scende che $\tau = 0$ per v.a. indipendenti. In generale il viceversa non vale (tranne nel caso di leggi Normali bivariate). Ora, volendo costruire un test di indipendenza basato sulla τ di Kendall è necessario, anzitutto, costruirne uno stimatore. Sia

$$A_{ij} = \mathrm{Sgn}(X_j - X_i)\ \mathrm{Sgn}(Y_j - Y_i)\,, \quad i, j = 1, \ldots, n,$$

che assume i valori

$$a_{ij} = \begin{cases} +1, & \text{se } (X_i, Y_i) \text{ e } (X_j, Y_j) \text{ sono concordanti} \\ 0, & \text{se } (X_i, Y_i) \text{ e } (X_j, Y_j) \text{ non sono né concordanti né discordanti} \\ -1, & \text{se } (X_i, Y_i) \text{ e } (X_j, Y_j) \text{ sono discordanti} \end{cases}.$$

Chiaramente vale

$$\mathbf{P}\{A_{ij} = a_{ij}\} = \begin{cases} \pi_c, & \text{se } a_{ij} = +1 \\ 1 - \pi_c - \pi_d, & \text{se } a_{ij} = 0 \\ \pi_d, & \text{se } a_{ij} = -1 \end{cases},$$

e $\mathbf{E}(A_{ij}) = \pi_c - \pi_d = \tau$. Dunque la v.a. A_{ij} è uno stimatore corretto di τ. Essendo $a_{ij} = a_{ji}$ e $a_{ii} = 0$, è possibile introdurre quale stimatore della τ di Kendall la statistica

$$T_n = \frac{1}{\binom{n}{2}} \sum_{i<j} A_{ij}, \tag{9.33}$$

che gode della seguente proprietà.

PROPOSIZIONE 9.4. *La statistica T_n data dall'equazione (9.33) è uno stimatore non distorto della τ di Kendall.*

NOTA 9.6.

Un modo alternativo per calcolare T_n, spesso usato nelle applicazioni, è il seguente. Se

$$P = \#(A_{ij} > 0) \quad \text{e} \quad N = \#(A_{ij} < 0)\,,$$

dove $1 \le i < j \le n$, allora

$$T_n = \frac{1}{\binom{n}{2}} (P - N). \tag{9.34}$$

Se non vi sono valori ripetuti tra le componenti X o Y delle osservazioni (detti "Ties" in Statistica), allora $A_{ij} \neq 0$ per $i \neq j$, e $P + N = \binom{n}{2}$. In tal caso si ha

$$T_n = \frac{1}{\binom{n}{2}} 2P - 1 = 1 - 2N \frac{1}{\binom{n}{2}}. \tag{9.35}$$

Intuitivamente si tende a rifiutare l'ipotesi di indipendenza se $|T_n|$ è "grande". Nell'ipotesi $\mathcal{H}_0$ si ha $\tau = 0$, e la corrispondente legge di T_n è simmetrica. L'uso del *test di indipendenza di Kendall*, al livello $\alpha \in (0, 1)$, è come segue. Si rifiuta l'ipotesi $\mathcal{H}_0 \colon X \perp\!\!\!\perp Y$ se $|t_n| \ge \tau_{n,1-\alpha/2}$. Per valori "piccoli" della ampiezza n la legge e i quantili di T_n possono essere calcolati direttamente; alcuni valori sono comunque riportati sulle Tavole per combinazioni selezionate dei parametri n, α. Si può dimostrare che, asintoticamente (in realtà, già per $n > 7$), la statistica T_n ha legge Normale, con media nulla e varianza pari a $\frac{2(2n+5)}{9n(n-1)}$; dunque, per campioni di "grande" ampiezza, la statistica

$$\tilde{T}_n = \frac{3\sqrt{n(n-1)}}{\sqrt{2(2n+5)}}\, T_n$$

può essere trattata approssimativamente come una v.a. di legge $\mathtt{N}(0, 1)$. Versioni multivariate della τ di Kendall possono trovarsi in [14, 20].

ESEMPIO 9.11. ▶

Durante un concorso due commissari, X e Y, giudicano quattro candidati A, B, C e D. Si vuole verificare l'ipotesi $\mathcal{H}_0$ che i giudizi siano indipendenti. La *Tabella (9.7)* mostra gli esiti del concorso, ove i candidati sono

	A	*B*	*C*	*D*
X	3	4	2	1
Y	3	1	4	2

	D	*C*	*A*	*B*
X	1	2	3	4
Y	2	4	3	1

Tabella 9.7: i dati dell'*Esempio (9.11)*.

classificati con punteggi da 1 a 4; onde facilitare i calcoli, la stessa tabella è ripetuta a lato ordinata secondo il punteggio crescente dato dal commissario X.

Per verificare l'ipotesi $\mathcal{H}_0$ calcoliamo anzitutto i valori di P e N; in realtà, è sufficiente calcolare P. Si vede subito che i vettori concordanti sono (usando una notazione ovvia)

$$\begin{pmatrix} X_D \\ Y_D \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \text{ e } \begin{pmatrix} X_C \\ Y_C \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \end{pmatrix}, \qquad \begin{pmatrix} X_D \\ Y_D \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \text{ e } \begin{pmatrix} X_A \\ Y_A \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix}.$$

Dunque $P = 2$ e $t_4 = -\frac{1}{3}$. Poiché la legge di T_4 è discreta, non è possibile trovare un valore unico del quantile $\tau_{4,1-\alpha/2}$ per qualsiasi livello $\alpha \in (0, 1)$. Tuttavia possiamo ragionare come segue. Nell'ipotesi $\mathcal{H}_0$ la statistica T_4 ha legge simmetrica, come illustrato in *Figura (9.4)*: i valori mostrati sono ricavati dalle Tavole.

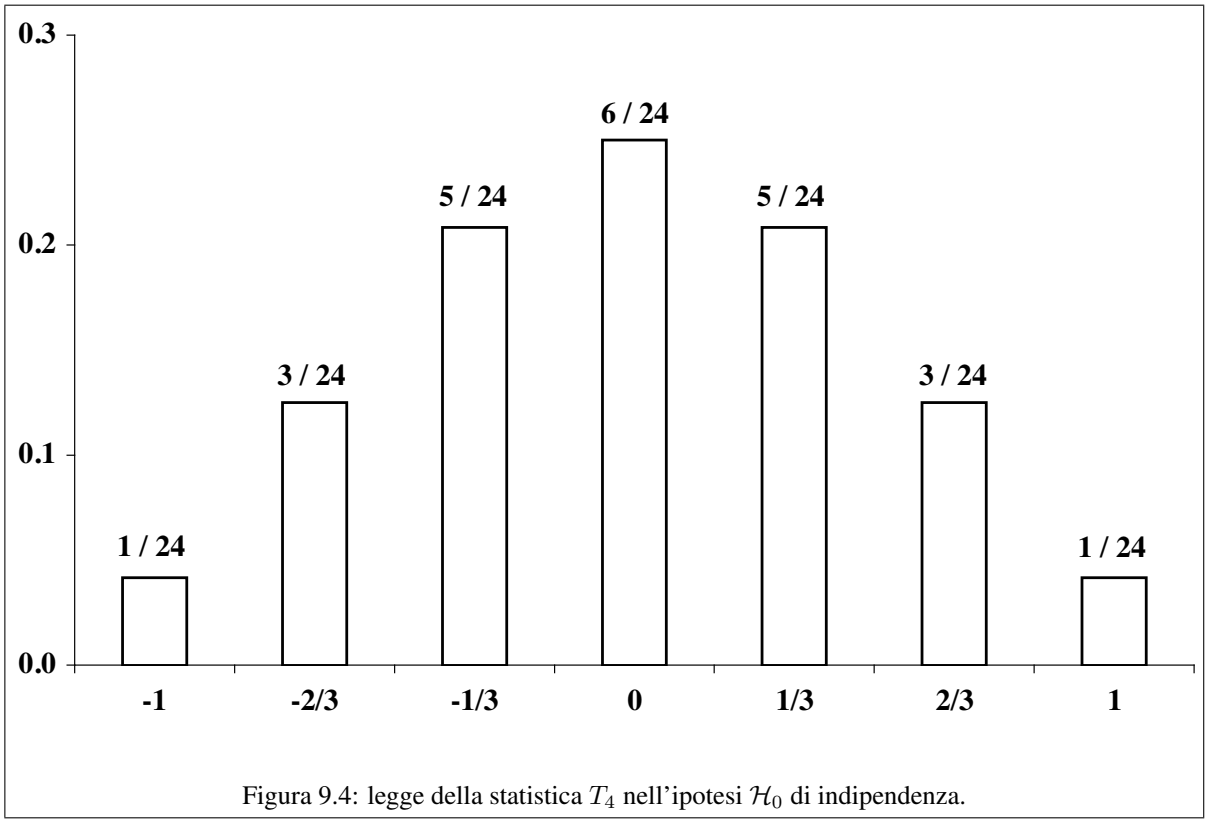

Figura 9.4: legge della statistica T_4 nell'ipotesi $\mathcal{H}_0$ di indipendenza.

Dunque, $\mathbf{P}_{\mathcal{H}_0}\left\{|T_4| \geq \frac{1}{3}\right\} = \frac{18}{24}$; ne segue che il valore $\frac{1}{3}$ viene eguagliato o superato con il 75% di probabilità (ovvero, in pratica, con una frequenza alquanto elevata), e quindi esso non rappresenta un valore "critico" di $|T_4|$. Dunque, il fatto di aver osservato l'evento $\left\{T_4 = -\frac{1}{3}\right\}$ porta ad accettare l'ipotesi $\mathcal{H}_0$, almeno ai livelli usualmente adottati nella pratica statistica. Si noti come il test di indipendenza di Kendall si possa usare anche in presenza di dati non di natura numerica. ◀

9.4.2 Il test di indipendenza di Spearman

Prima di procedere, è opportuno introdurre la nozione di *rango*.

DEFINIZIONE 9.13 (Rango)**.** Siano $(X_1, \ldots, X_n)$ un campione, X_i un generico elemento del campione e $X_{(j_i)}$ la statistica d'ordine corrispondente, con $i, j_i \in \{1, \ldots, n\}$. L'indice j_i è detto *rango* di X_i, e si scrive

$$\mathbf{r}(X_i) = j_i. \tag{9.36}$$

In pratica il rango di un elemento del campione corrisponde alla posizione dello stesso all'interno del campione ordinato in senso crescente, e quindi all'indice della statistica d'ordine corrispondente. Poiché si considerano solamente leggi continue, il rango è univocamente definito. Evidentemente il rango non dipende dalla distribuzione del campione in esame.

Sia ora $((X_1, Y_1), \ldots, (X_n, Y_n))$ un campione di ampiezza n di vettori bidimensionali indipendenti. Se le v.a. in questione sono di quadrato sommabile, il *coefficiente di correlazione lineare di Pearson*

$$\boldsymbol{\rho}_{X,Y} = \frac{\mathbf{C}(X, Y)}{\sqrt{\mathbf{V}(X)\,\mathbf{V}(Y)}}$$

può fornire informazioni circa il legame eventualmente esistente tra X e Y: in particolare, se $\boldsymbol{\rho}_{X,Y} \neq 0$ allora le v.a. non sono indipendenti (si ricordi, tuttavia, che $\boldsymbol{\rho}_{X,Y} = 0$ non implica che X e Y siano indipendenti).

NOTA 9.7.

Si osservi che l'esistenza di $\boldsymbol{\rho}_{X,Y}$ è legata a quella dei momenti del secondo ordine di X e Y e quindi non è garantita per tutti i vettori (X, Y).

Nella pratica, la statistica

$$R_P = \frac{\sum_{i=1}^n \left(X_i - \overline{X}_n\right)\left(Y_i - \overline{Y}_n\right)}{\sqrt{\sum_{i=1}^n \left(X_i - \overline{X}_n\right)^2}\sqrt{\sum_{i=1}^n \left(X_i - \overline{X}_n\right)^2}}$$

è utilizzata quale stimatore di $\boldsymbol{\rho}_{X,Y}$; tuttavia, anche nei casi più semplici, è difficile determinarne la legge. Pertanto, può essere talora conveniente sostituire R con un'altra statistica basata sui ranghi delle v.a. del campione.

Si considerino le statistiche

$$R_i = \mathbf{r}(X_i) \quad \text{e} \quad S_i = \mathbf{r}(Y_i)\,, \quad i = 1, \ldots, n.$$

Ovviamente $R_i, S_i \in \{1, \ldots, n\}$ e, inoltre,

$$\sum_{i=1}^n R_i = \sum_{i=1}^n S_i = \frac{n(n+1)}{2}, \tag{9.37a}$$

$$\overline{R}_n = \frac{1}{n}\sum_{i=1}^n R_i = \frac{n+1}{2}, \quad \overline{S}_n = \frac{1}{n}\sum_{i=1}^n S_i = \frac{n+1}{2}, \tag{9.37b}$$

$$\sum_{i=1}^n (R_i - \overline{R}_n)^2 = \sum_{i=1}^n (S_i - \overline{S}_n)^2 = \frac{n(n^2-1)}{12}. \tag{9.37c}$$

Operando una sostituzione formale nella definizione di R_P è possibile definire una nuova statistica ρ_S:

$$\begin{aligned} \rho_S &= \frac{12\sum_{i=1}^n \left(R_i - \overline{R}_n\right)\left(S_i - \overline{S}_n\right)}{n^3 - n} \\ &= \frac{12\sum_{i=1}^n R_i S_i}{n(n^2-1)} - \frac{3(n+1)}{n-1}. \end{aligned} \tag{9.38}$$

Ponendo $D_i = R_i - S_i = (R_i - \overline{R}_n) - (S_i - \overline{S}_n)$ si ottiene

$$\begin{aligned} \sum_{i=1}^n D_i^2 &= \sum_{i=1}^n (R_i - \overline{R}_n)^2 + \sum_{i=1}^n (S_i - \overline{S}_n)^2 - 2\sum_{i=1}^n (R_i - \overline{R}_n)(S_i - \overline{S}_n) \\ &= \frac{1}{6}\,n(n^2-1) - 2\sum_{i=1}^n (R_i - \overline{R}_n)(S_i - \overline{S}_n). \end{aligned}$$

Possiamo ora dare un nome alla statistica ρ_S.

DEFINIZIONE 9.14 (Coefficiente di correlazione dei ranghi di Spearman)**.** La statistica ρ_S definita come

$$\rho_S = 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^{n} D_i^2 \tag{9.39}$$

è detta *coefficiente di correlazione dei ranghi di Spearman*.

NOTA 9.8.

Grazie alla Teoria delle Copule [26] è possibile mostrare che ρ_S "misura" la *distanza media* (in senso integrale) tra la copula $\mathbf{C}_{XY}$ del vettore (X, Y) e quella dell'indipendenza $\mathbf{\Pi}_2$, essendo

$$\rho_S \propto \iint_{\mathbb{I}^2} \mathbf{C}_{XY}(u, v) - \mathbf{\Pi}_2(u, v)\, \mathrm{d}u\, \mathrm{d}v.$$

Si osservi inoltre come l'esistenza di ρ_S sia garantita per tutti i vettori (X, Y), essendo il calcolo di ρ_S basato esclusivamente sui ranghi.

Dopo alcune (tediose) manipolazioni algebriche è possibile mostrare che

$$\mathbf{E}(\rho_S) = \frac{12}{n^2-1}\, \mathbf{E}(RS) - \frac{3(n+1)}{n-1}.$$

Nell'ipotesi $\mathcal{H}_0$ le v.a. X e Y sono indipendenti, e dunque lo sono anche i rispettivi ranghi R e S (in quanto funzioni misurabili dei rispettivi campioni). Ne scende che

$$\mathbf{E}_{\mathcal{H}_0}(RS) = \mathbf{E}(R)\, \mathbf{E}(S) = \left(\frac{n+1}{2}\right)^2,$$

e inoltre

$$\mathbf{E}_{\mathcal{H}_0}(\rho_S) = 0. \tag{9.40}$$

Dunque, intuitivamente, si rifiuta l'ipotesi $\mathcal{H}_0$ di indipendenza se $|\rho_S|$ è "troppo" grande, ovvero, al livello $\alpha \in (0, 1)$, se $|r_S| > r_{1-\alpha/2}$, con ovvio significato della notazione. Il calcolo dei quantili r_α richiede la conoscenza della legge di ρ_S nell'ipotesi $\mathcal{H}_0$; in realtà, è sufficiente calcolare solamente una coda della distribuzione, essendo questa simmetrica. Per campioni di ampiezza $n \geq 10$ è possibile mostrare che la statistica

$$\tilde{\rho}_S = \sqrt{n-1}\, \rho_S$$

ha, approssimativamente, una legge $\mathtt{N}(0, 1)$. Si noti che $-1 \leq \rho_S \leq 1$, dove i valori estremali $+1$ e -1 si manifestano quando, rispettivamente, $R_i = S_i$ oppure $R_i = n + 1 - S_i$ per tutti gli indici. Versioni multivariate del coefficiente di correlazione dei ranghi di Spearman possono trovarsi in [29].

ESEMPIO 9.12. ▶

Si riconsideri l'*Esempio (9.11)*, dove la *Tabella (9.7)* mostra i ranghi di interesse. Il calcolo di ρ_S in base ai risultati del concorso fornisce

$$r_S = \frac{12 \cdot 23}{4 \cdot 15} - \frac{3 \cdot 5}{3} = -0.4.$$

Poiché, in base alle Tavole, si ha che $\mathbf{P}_{\mathcal{H}_0}\{|\rho_S| \geq 0.4\} = 0.75$, invocando un ragionamento già svolto si accetta l'ipotesi $\mathcal{H}_0$ ai livelli usualmente adottati nella pratica statistica. ◀

ESEMPIO 9.13. ▶

Si misurano le altezze di $n = 25$ coppie di sposi, e il calcolo di ρ_S fornisce $r_S = 0.3$. Si vuole vagliare, al livello $\alpha = 0.05$, l'ipotesi $\mathcal{H}_0$ che le stature di moglie e marito siano indipendenti. L'ampiezza n del campione è sufficiente per poter invocare l'approssimazione con la legge Normale, la quale sancisce che, approssimativamente, $\sqrt{24}\, \rho_S \sim \mathtt{N}(0, 1)$. Dunque la regione critica è del tipo $\left\{\left|\sqrt{24}\, \rho_S\right| > z\right\}$, dove il quantile di interesse z può essere calcolato con l'ausilio delle Tavole della legge Normale, dovendo essere

$$\mathbf{P}\left\{\left|\sqrt{24}\, \rho_S\right| > z\right\} = 0.05.$$

Ne scende che $z \approx 1.96$, e dunque la regione critica è $\{|\rho_S| > 0.400\}$, che impone di accettare l'ipotesi $\mathcal{H}_0$. Si noti come l'ipotesi sarebbe da accettare anche al livello $\alpha = 0.1$, essendo in tal caso $z \approx 1.645$, con regione critica $\{|\rho_S| > 0.336\}$. In altre parole, la statistica pare confermare che "l'amore è cieco". ◀

Capitolo 10

ANALISI DELLA VARIANZA

L'*analisi della varianza*, abbreviata dall'acronimo anglosassone ANOVA ("ANalysis Of VAriance"), rappresenta una delle tecniche statistiche più utilizzate nella pratica. L'idea che sta alla base di tale metodologia è quella, come vedremo meglio nel seguito, di partire le fluttuazioni: infatti, a dispetto dell'acronimo che la definisce, l'ANOVA si concentra non tanto sulla analisi delle varianze quanto su quella delle variazioni nelle medie. Uno dei problemi a cui l'ANOVA cerca di rispondere è quello di capire come ricavare quante più informazioni possibili sul maggior numero di popolazioni usando campioni di ampiezza minima. Storicamente le tecniche ANOVA sono state sviluppate a seguito di ricerche sul confronto della produttività agricola; in pratica, tali metodologie possono essere usate nello studio di un gran numero di problemi nei più svariati campi. Esempi ne sono:

- la produzione agricola in corrispondenza di diversi fertilizzanti e/o diversi terreni;
- lo studio degli effetti di un nuovo medicinale del quale si vogliono paragonare diversi dosaggi;
- il confronto del rendimento di diversi tipi di benzina sullo stesso modello di auto e/o su modelli differenti;
- il confronto dell'usura di materiali differenti;
- il confronto del rendimento scolastico per scuole ed insegnanti differenti.

Gli esempi si potrebbero facilmente moltiplicare. Nel seguito studieremo alcune tra le tecniche ANOVA più comunemente utilizzate nella pratica.

10.1 Analisi della varianza ad una via

Nella tecnica ANOVA *ad una via* (dal termine anglosassone *one-way* — si vedano, ad esempio, [24, 6]) si assume che i dati disponibili $\boldsymbol{Y}$ siano osservati secondo il modello

$$Y_{ij} = \theta_i + \epsilon_{ij}, \quad i = 1, \ldots, k, \quad j = 1, \ldots n_i,$$

dove le quantità θ_i sono parametri reali incogniti e le variabili ϵ_{ij} sono v.a. che descrivono le possibili fluttuazioni (errori). Tipicamente l'indice i denota il *trattamento* (o il *livello*) e il generico parametro θ_i assume il significato di *media* del trattamento i-esimo: tale terminologia deriva dal fatto che le tecniche ANOVA sono nate nell'ambito degli studi sulla efficacia di diversi fertilizzanti, e serve appunto a identificare gruppi di osservazioni raccolte in condizioni differenti per quanto riguarda, ad esempio, il tipo di concime o il dosaggio di un farmaco.

Al fine di poter effettuare stime statistiche e verifiche di ipotesi sul vettore dei parametri incogniti $\boldsymbol{\theta}$ è necessario chiarire quale sia il contesto probabilistico nel quale il modello dato si colloca.

DEFINIZIONE 10.1 (ANOVA ad una via)**.** Le v.a. Y_{ij} sono osservate in accordo al modello

$$Y_{ij} = \theta_i + \epsilon_{ij}, \quad i = 1, \ldots, k, \quad j = 1, \ldots n_i, \tag{10.1}$$

dove:

1. $\mathbf{E}(\epsilon_{ij}) = 0$ e $\mathbf{V}(\epsilon_{ij}) = \sigma_i^2 < \infty$ per tutti gli indici i, j; inoltre $\mathbf{C}(\epsilon_{ij}, \epsilon_{lm}) = 0$ se $(i, j) \neq (l, m)$;

2. le fluttuazioni (errori) sono v.a. Normali e indipendenti;
3. $\sigma_i^2 = \sigma^2$ per ogni indice i.

OSSERVAZIONE 10.1.

Le assunzioni alla base del modello dato richiedono alcuni commenti.

1. L'ipotesi che $\mathbf{E}(\epsilon_{ij}) = 0$ per tutti gli indici i, j non lede la generalità del modello; quale conseguenza si ha, per ogni $i = 1, \ldots, k$,
$$\mathbf{E}(Y_{ij}) = \theta_i, \quad j = 1, \ldots n_i, \tag{10.2}$$
che chiarisce l'interpretazione del parametro θ_i quale media del trattamento i-esimo.
2. Senza l'ipotesi *(2)* sarebbe possibile calcolare stime puntuali dei parametri o costruire stimatori di varianza minima all'interno di una certa classe, ma non sarebbe possibile effettuare verifiche di ipotesi o stime per intervalli. Inoltre, ipotizzare per gli errori leggi diverse da quella Normale (sebbene possibile e talora legittimo) condurrebbe a calcoli estremamente complessi.
3. L'ipotesi di *omoschedasticità* (ovvero $\sigma_i^2 = \sigma^2$ per ogni indice i) è strettamente legata all'ipotesi *(2)*, e influenza la robustezza delle tecniche ANOVA. In generale, nel caso in cui si sospetti che le osservazioni disponibili violino le assunzioni del modello, si cerca anzitutto di trasformare i dati (tipicamente in maniera non lineare) così da fare in modo che il nuovo campione rispetti le ipotesi standard enunciate.

Si noti che non si assume che il numero di osservazioni sia lo stesso per ciascun trattamento.

Schematicamente le osservazioni possono essere raccolte in una tabella come quella che segue.

Trattamenti o Livelli				
1	**2**	**3**	**...**	$\boldsymbol{k}$
y_{11}	y_{21}	y_{31}	$\cdots$	y_{k1}
y_{12}	y_{22}	y_{32}	$\cdots$	y_{k2}
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$
$\vdots$	$\vdots$	y_{3n_3}	$\ldots$	$\vdots$
y_{1n_1}	$\vdots$		$\ldots$	$\vdots$
	y_{2n_2}		$\cdots$	y_{kn_k}

Nell'ambito delle tecniche ANOVA il test tipico riguarda la verifica dell'ipotesi $\mathcal{H}_0 \colon \theta_1 = \cdots = \theta_k$, una ipotesi che, nella pratica, spesso è di scarso interesse se non addirittura palesemente falsa: in effetti, assumere che trattamenti *differenti* producano in media gli *stessi* effetti può essere irragionevole. Il vero interesse nelle tecniche ANOVA non risiede tanto nella costruzione di test, quanto nella stima, tipicamente per intervalli; tuttavia, data l'equivalenza delle due, è possibile (e più facile) affrontare il problema in termini di verifica di ipotesi.

10.1.1 Inferenze su combinazioni lineari

Rifiutare l'ipotesi (ANOVA) $\mathcal{H}_0$ significa accettare il fatto che vi siano differenze tra le componenti del vettore $\boldsymbol{\theta}$, senza tuttavia poter sapere quali siano gli elementi dissimili. Un modo per aggirare l'ostacolo può essere quello di scrivere le ipotesi in termini di *contrasti*.

DEFINIZIONE 10.2 (Contrasto)**.** Sia $\boldsymbol{t} = (t_1, \ldots, t_k)$ un vettore di variabili (aleatorie o deterministiche), e sia $\boldsymbol{a} = (a_1, \ldots, a_k)$ un vettore di costanti note. Se $\sum_{i=1}^k a_i = 0$, allora la combinazione lineare

$$\sum_{i=1}^{k} a_i \, t_i = \langle \boldsymbol{a}, \boldsymbol{t} \rangle \tag{10.3}$$

è detta *contrasto*.

I contrasti sono importanti in quanto, scrivendo le ipotesi in termini di opportuni contrasti, è possibile fare inferenze circa i singoli elementi di $\boldsymbol{\theta}$. Ad esempio, volendo confrontare θ_1 e θ_k, si potrebbe usare il vettore $\boldsymbol{a} = (1, 0, \ldots, 0, -1)$, che genera il contrasto $\langle \boldsymbol{a}, \boldsymbol{\theta} \rangle = \theta_1 - \theta_k$. Il risultato che segue è importante e chiarisce il ruolo giocato dai contrasti.

PROPOSIZIONE 10.1. *Sia $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ un vettore di parametri. Allora*

$$\theta_1 = \cdots = \theta_k \Leftrightarrow \langle \boldsymbol{a}, \boldsymbol{\theta} \rangle = 0 \quad \textit{per ogni } \boldsymbol{a} \in \mathfrak{A}, \tag{10.4}$$

dove $\mathfrak{A} = \left\{ \boldsymbol{a} = (a_1, \ldots, a_k) \colon \sum_{i=1}^k a_i = 0 \right\}$, ovvero i parametri $(\theta_1, \ldots, \theta_k)$ sono uguali se, e solo se, tutti i contrasti soddisfanno alla relazione $\langle \boldsymbol{a}, \boldsymbol{\theta} \rangle = 0$, con $\boldsymbol{a} \in \mathfrak{A}$.

DIM. Se $\theta_1 = \cdots = \theta_k = \theta$ allora, essendo $\boldsymbol{a} \in \mathfrak{A}$,

$$\langle \boldsymbol{a}, \boldsymbol{\theta} \rangle = \sum_{i=1}^k a_i \, \theta_i = \theta \sum_{i=1}^k a_i = 0.$$

Viceversa, si considerino i vettori k-dimensionali di $\mathfrak{A}$

$$\boldsymbol{a}_1 = (1, -1, 0, \ldots, 0), \quad \boldsymbol{a}_2 = (0, 1, -1, 0, \ldots, 0), \quad \ldots \quad \boldsymbol{a}_{k-1} = (0, \ldots, 0, 1, -1),$$

che formano una *base* per $\mathfrak{A}$. Infatti $\mathfrak{A}$ è uno spazio lineare di dimensione $k-1$ (ciò a causa del vincolo $\sum_{i=1}^k a_i = 0$), e ogni suo elemento può essere scritto quale opportuna combinazione lineare dei vettori $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{k-1}$: se $\boldsymbol{a} = \sum_{i=1}^{k-1} c_i \boldsymbol{a}_i$, $c_i \in \mathbb{R}$, allora $\boldsymbol{a} = (c_1, c_2 - c_1, \ldots, c_{k-1} - c_{k-2}, -c_{k-1})$ e dunque $\sum_{i=1}^k a_i = 0$, cioè $\boldsymbol{a} \in \mathfrak{A}$; viceversa, se $\boldsymbol{a} = (a_1, \ldots, a_k) \in \mathfrak{A}$, è sufficiente porre $c_i = \sum_{j=1}^i a_j$, $i = 1, \ldots, k-1$, per ottenere $\boldsymbol{a} = \sum_{i=1}^{k-1} c_i \boldsymbol{a}_i$, dovendo essere $a_k = -\sum_{i=1}^{k-1} a_i = -c_{k-1}$.

Costruendo contrasti con gli elementi di tale base si ottiene

$$\boldsymbol{a}_1 \Rightarrow \theta_1 = \theta_2, \quad \boldsymbol{a}_2 \Rightarrow \theta_2 = \theta_3, \quad \ldots \quad \boldsymbol{a}_{k-1} \Rightarrow \theta_{k-1} = \theta_k,$$

che, congiuntamente, danno $\theta_1 = \cdots = \theta_k$. □

Evidentemente l'ipotesi (ANOVA) nulla $\mathcal{H}_0 \colon \theta_1 = \cdots = \theta_k$ è vera se, e solo se, è vera l'ipotesi $\mathcal{H}_0 \colon \langle \boldsymbol{a}, \boldsymbol{\theta} \rangle = 0$ per ogni $\boldsymbol{a} \in \mathfrak{A}$. Inoltre, se $\mathcal{H}_0$ fosse falsa allora esisterebbe almeno un contrasto non nullo. Il vantaggio dell'uso dei contrasti risiede nella maggiore semplicità di interpretazione dei risultati, potendo spesso operare in maniera *univariata*: infatti, molte sono le inferenze che possono essere espresse in termini di un contrasto o di insiemi di contrasti, come vedremo meglio nel seguito.

Quale conseguenza delle assunzioni del modello ANOVA ad una via si ha che

$$Y_{ij} \sim \mathtt{N}\big(\theta_i, \sigma^2\big), \quad i = 1, \ldots, k, \quad j = 1, \ldots, n_i,$$

e dunque

$$\overline{Y}_{i\bullet} = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij} \sim \mathtt{N}\left(\theta_i, \frac{\sigma^2}{n_i}\right), \quad i = 1, \ldots, k.$$

Inoltre, dato un insieme qualsiasi di costanti $\boldsymbol{a} = (a_1, \ldots, a_k)$, la v.a. $\sum_{i=1}^k a_i \, \overline{Y}_{i\bullet}$ ha legge Normale, con

$$\mathbf{E}\left(\sum_{i=1}^k a_i \, \overline{Y}_{i\bullet} \right) = \sum_{i=1}^k a_i \, \theta_i \quad \text{e} \quad \mathbf{V}\left(\sum_{i=1}^k a_i \, \overline{Y}_{i\bullet} \right) = \sigma^2 \sum_{i=1}^k \frac{a_i^2}{n_i}.$$

Infine,

$$\frac{\sum_{i=1}^k a_i \, \overline{Y}_{i\bullet} - \sum_{i=1}^k a_i \, \theta_i}{\sqrt{\sigma^2 \sum_{i=1}^k \frac{a_i^2}{n_i}}} \sim \mathtt{N}(0, 1)\,.$$

Volendo fare inferenze circa le componenti del vettore $\boldsymbol{\theta}$, è anzitutto necessario fornire una stima del parametro incognito σ^2, ad esempio tramite la varianza campionaria:

$$\overline{S}_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} \big(Y_{ij} - \overline{Y}_{i\bullet}\big)^2, \quad i = 1, \ldots, k,$$

da cui risulta $(n_i - 1)\overline{S}_i^2/\sigma^2 \sim \chi^2(n_i - 1)$. Tuttavia, poiché ciascuna v.a. $\overline{S}_i^2$ è uno stimatore di σ^2, è possibile migliorare la stima usando la statistica

$$\overline{S}_P^2 = \frac{1}{N-k} \sum_{i=1}^{k} (n_i - 1)\, \overline{S}_i^2,$$

dove $N - k = \sum_{i=1}^{k}(n_i - 1)$. Dalla indipendenza delle v.a. $\overline{S}_i^2$ scende che $(N-k)\overline{S}_P^2/\sigma^2 \sim \chi^2(N-k)$. Inoltre, poiché $\overline{S}_P^2$ è indipendente da ciascuna v.a. $\overline{Y}_{i\bullet}$, si ha che

$$\frac{\sum_{i=1}^{k} a_i\, \overline{Y}_{i\bullet} - \sum_{i=1}^{k} a_i\, \theta_i}{\sqrt{\overline{S}_P^2 \, \sum_{i=1}^{k} a_i^2/n_i}} \sim \mathtt{t}(N-k)\,.$$

Dato un insieme qualsiasi di costanti $\boldsymbol{a} = (a_1, \ldots, a_k)$, e volendo procedere alla verifica delle ipotesi

$$\mathcal{H}_0\colon \langle \boldsymbol{a}, \boldsymbol{\theta} \rangle = 0 \quad \text{contro} \quad \mathcal{H}_1\colon \langle \boldsymbol{a}, \boldsymbol{\theta} \rangle \neq 0, \tag{10.5}$$

al livello $\alpha \in (0,1)$ si rifiuterà $\mathcal{H}_0$ se

$$\left| \frac{\sum_{i=1}^{k} a_i\, \overline{Y}_{i\bullet}}{\sqrt{\overline{S}_P^2 \, \sum_{i=1}^{k} a_i^2/n_i}} \right| > t_{N-k,1-\frac{\alpha}{2}}. \tag{10.6}$$

Inoltre, è possibile calcolare un IC di livello $\alpha \in (0,1)$ per $\sum_{i=1}^{k} a_i\, \theta_i$:

$$\sum_{i=1}^{k} a_i\, \overline{Y}_{i\bullet} - t_{N-k,1-\frac{\alpha}{2}} \sqrt{\overline{S}_P^2 \sum_{i=1}^{k} a_i^2/n_i} \leq \sum_{i=1}^{k} a_i\, \theta_i \leq \sum_{i=1}^{k} a_i\, \overline{Y}_{i\bullet} + t_{N-k,1-\frac{\alpha}{2}} \sqrt{\overline{S}_P^2 \sum_{i=1}^{k} a_i^2/n_i}. \tag{10.7}$$

L'Esempio che segue mostra la generalità e la potenza dell'uso dei contrasti.

ESEMPIO 10.1. ►

Si supponga di voler confrontare i trattamenti 1 e 2, ovvero di voler procedere alla verifica delle ipotesi

$$\mathcal{H}_0\colon \theta_1 = \theta_2 \quad \text{contro} \quad \mathcal{H}_1\colon \theta_1 \neq \theta_2.$$

A tale fine è sufficiente scegliere $\boldsymbol{a} = (1, -1, 0, \ldots, 0)$, e si rifiuterà $\mathcal{H}_0$, al livello $\alpha \in (0,1)$, se

$$\left| \frac{\overline{Y}_{1\bullet} - \overline{Y}_{2\bullet}}{\sqrt{\overline{S}_P^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \right| > t_{N-k,1-\frac{\alpha}{2}}.$$

Si noti che l'uso dello stimatore $\overline{S}_P^2$ consente di sfruttare anche le "informazioni" eventualmente fornite dai trattamenti $3, \ldots, k$ ai fini della stima del parametro σ^2. Si supponga poi di voler confrontare il trattamento 1 (che, ad esempio, potrebbe essere quello di controllo) con la media dei trattamenti 2 e 3, ovvero di voler procedere alla verifica delle ipotesi

$$\mathcal{H}_0\colon \theta_1 = \frac{\theta_2 + \theta_3}{2} \quad \text{contro} \quad \mathcal{H}_1\colon \theta_1 \neq \frac{\theta_2 + \theta_3}{2}.$$

A tale fine è sufficiente scegliere $\boldsymbol{a} = (1, -\frac{1}{2}, -\frac{1}{2}, 0, \ldots, 0)$, e si rifiuterà $\mathcal{H}_0$, al livello $\alpha \in (0,1)$, se

$$\left| \frac{\overline{Y}_{1\bullet} - \frac{1}{2}\overline{Y}_{2\bullet} - \frac{1}{2}\overline{Y}_{3\bullet}}{\sqrt{\overline{S}_P^2 \left(\frac{1}{n_1} + \frac{1}{4n_2} + \frac{1}{4n_3}\right)}} \right| > t_{N-k,1-\frac{\alpha}{2}}.$$

Infine, grazie all'*equazione (10.7)*, la costruzione di IC è immediata in entrambi i casi illustrati. ◄

10.1.2 Il test ANOVA ad una via

Dalla *Proposizione (10.1)* sappiamo che l'ipotesi (ANOVA) nulla $\mathcal{H}_0\colon \theta_1 = \cdots = \theta_k$ è equivalente all'ipotesi $\mathcal{H}_0\colon \langle \boldsymbol{a}, \boldsymbol{\theta} \rangle = 0$ per ogni $\boldsymbol{a} \in \mathfrak{A}$. Quest'ultima versione può essere foriera di risultati interessanti per quanto concerne la verifica di ipotesi, sfruttando opportunamente il *Principio di Inclusione-Esclusione* (o di *Unione-Intersezione*) e la *Disuguaglianza di Bonferroni*, che richiamiamo brevemente.

TEOREMA 10.1 (Principio di Inclusione-Esclusione)**.** *Siano $A_1, \ldots, A_k$ eventi qualsiasi. Allora*

$$\begin{aligned} \mathbf{P}\left\{\bigcup_{i=1}^{k} A_i\right\} = \sum_{i=1}^{k} \mathbf{P}\{A_i\} - \sum_{i_1<i_2} \mathbf{P}\{A_{i_1} \cap A_{i_2}\} \\ + \sum_{i_1<i_2<i_3} \mathbf{P}\{A_{i_1} \cap A_{i_2} \cap A_{i_3}\} + \cdots + (-1)^{k+1}\, \mathbf{P}\left\{\bigcap_{i=1}^{k} A_i\right\}. \end{aligned} \tag{10.8}$$

TEOREMA 10.2 (Disuguaglianza di Bonferroni)**.** *Siano A, B eventi qualsiasi. Allora*

$$\mathbf{P}\{A \cap B\} \geq \mathbf{P}\{A\} + \mathbf{P}\{B\} - 1. \tag{10.9}$$

Procedendo per induzione è possibile generalizzare il *Teorema (10.2).*

COROLLARIO 10.1. *Siano $A_1, \ldots, A_m$, $m > 1$, eventi qualsiasi. Allora*

$$\mathbf{P}\left\{\bigcap_{i=1}^{m} A_i\right\} \geq \sum_{i=1}^{m} \mathbf{P}\{A_i\} - (m-1). \tag{10.10}$$

Sia ora $\boldsymbol{a} = (a_1, \ldots, a_k)$ un generico vettore di $\mathfrak{A}$, e sia

$$\Theta_{\boldsymbol{a}} = \{\boldsymbol{\theta}\colon\ \langle \boldsymbol{a}, \boldsymbol{\theta} \rangle = 0\}\,.$$

Ne segue che

$$\theta_1 = \cdots = \theta_k \Leftrightarrow \boldsymbol{\theta} \in \Theta_{\boldsymbol{a}} \text{ per ogni } \boldsymbol{a} \in \mathfrak{A} \ \Leftrightarrow \boldsymbol{\theta} \in \bigcap_{\boldsymbol{a} \in \mathfrak{A}} \Theta_{\boldsymbol{a}},$$

da cui scende che l'ipotesi (ANOVA) nulla può essere scritta come una intersezione. In particolare, $\mathcal{H}_0$ è da rifiutare se, e solo se, si rifiuta l'ipotesi $\mathcal{H}_{\Theta_{\boldsymbol{a}}}\colon \boldsymbol{\theta} \in \Theta_{\boldsymbol{a}}$ per qualche vettore $\boldsymbol{a} \in \mathfrak{A}$. Per quanto già visto, la verifica della singola ipotesi $\mathcal{H}_{\Theta_{\boldsymbol{a}}}$ può essere effettuata utilizzando la statistica test

$$T_{\boldsymbol{a}} = \left| \frac{\sum_{i=1}^{k} a_i\, \overline{Y}_{i\bullet} - \sum_{i=1}^{k} a_i\, \theta_i}{\sqrt{\overline{S}_P^2\ \sum_{i=1}^{k} a_i^2/n_i}} \right|, \tag{10.11}$$

e si rifiuta $\mathcal{H}_{\Theta_{\boldsymbol{a}}}$ se $T_{\boldsymbol{a}} > t^*$ per un opportuno valore t^*; qui $C_{\boldsymbol{a}} = \{T_{\boldsymbol{a}} > t^*\}$ rappresenta la regione critica. Ne segue che la regione critica C per la verifica dell'ipotesi (ANOVA) nulla è data dall'unione $\bigcup_{\boldsymbol{a} \in \mathfrak{A}} C_{\boldsymbol{a}}$. Tuttavia, poiché t^* *non dipende* da $\boldsymbol{a}$, è possibile dimostrare che l'ipotesi (ANOVA) nulla può essere rifiutata se

$$\sup_{\boldsymbol{a} \in \mathfrak{A}} T_{\boldsymbol{a}} > t^*, \tag{10.12a}$$

dove t^* è tale che

$$\mathbf{P}_{\mathcal{H}_0}\left\{\sup_{\boldsymbol{a} \in \mathfrak{A}} T_{\boldsymbol{a}} > t^*\right\} = \alpha \tag{10.12b}$$

per un livello $\alpha \in (0, 1)$ prefissato. Quale risultato fondamentale si ha il seguente Teorema.

TEOREMA 10.3. *Per la statistica $T_{\boldsymbol{a}}$ definita dalla equazione (10.11) vale*

$$\sup_{\boldsymbol{a} \in \mathfrak{A}} \left(T_{\boldsymbol{a}}^2\right) = \frac{1}{\overline{S}_P^2} \sum_{i=1}^{k} n_i \left(\left(\overline{Y}_{i\bullet} - \overline{\overline{Y}}\right) - \left(\theta_i - \overline{\theta}\right)\right)^2, \tag{10.13}$$

dove, se $N = \sum_{i=1}^{k} n_i$,

$$\overline{\overline{Y}} = \frac{1}{N} \sum_{i=1}^{k} n_i \overline{Y}_{i\bullet} \quad \textit{(detta} \text{ grande media}\textit{) e} \quad \overline{\theta} = \frac{1}{N} \sum_{i=1}^{k} n_i \theta_i.$$

*Inoltre, nell'ipotesi (*ANOVA*) nulla,*

$$\frac{1}{k-1} \sup_{\boldsymbol{a} \in \mathfrak{A}} \left(T_{\boldsymbol{a}}^2 \right) \sim \mathtt{F}(k-1, N-k)\,. \tag{10.14}$$

In pratica, la verifica di ipotesi si svolge introducendo la seguente statistica test.

DEFINIZIONE 10.3 (***F***-ANOVA ad una via)**.** La v.a. F data da

$$F = \frac{1}{(k-1)\,\overline{S}_P^2} \sum_{i=1}^{k} n_i \left(\overline{Y}_{i\bullet} - \overline{\overline{Y}} \right)^2 = \frac{N-k}{(k-1)} \frac{\sum_{i=1}^{k} n_i \left(\overline{Y}_{i\bullet} - \overline{\overline{Y}} \right)^2}{\sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(Y_{ij} - \overline{Y}_{i\bullet} \right)^2} \tag{10.15}$$

è detta *statistica* F-ANOVA *ad una via*.

Si noti come, nell'ipotesi (ANOVA) nulla, $F \sim \mathtt{F}(k-1, N-k)$. Dunque, al livello $\alpha \in (0,1)$, si rifiuta $\mathcal{H}_0$ se $f > f_{k-1,N-k,1-\alpha}$.

NOTA 10.1.

Si osservi che si sarebbe ottenuto lo stesso risultato, sebbene attraverso calcoli ben più complessi, utilizzando il *test del rapporto di verosimiglianza generalizzato* e gli stimatori SMV delle quantità coinvolte.

La metodologia ANOVA ad una via consente di distinguere in maniera efficace in quale modo trattamenti differenti influenzino le variabili osservate. In particolare, essa fornisce una partizione delle somme dei quadrati che definiscono la statistica F-ANOVA ad una via (talora dette *corrette*, in quanto vi compaiono sottrazioni di valori medi). La somma dei quadrati che compare al numeratore dell'*equazione (10.15)* è detta *somma esterna (between)* (*SEQ*), mentre quella al denominatore è detta *somma interna (within)* (*SIQ*): le due descrivono aspetti (cause) complementari che contribuiscono, in maniera diversa, alle fluttuazioni presenti nel modello ANOVA ad una via; in particolare, SEQ tiene conto degli effetti dei diversi trattamenti, e SIQ di quelli dovuti alla presenza di fluttuazioni aleatorie.

OSSERVAZIONE 10.2 (Partizionamento delle fluttazioni)**.**

Si consideri la relazione

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(Y_{ij} - \overline{\overline{Y}} \right)^2 = \sum_{i=1}^{k} n_i \left(\overline{Y}_{i\bullet} - \overline{\overline{Y}} \right)^2 + \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(Y_{ij} - \overline{Y}_{i\bullet} \right)^2 = \text{SEQ} + \text{SIQ}. \tag{10.16}$$

Il termine di sinistra, che "misura" globalmente le fluttuazioni senza distinguere tra cause *interne* ed *esterne*, può in effetti essere scritto come somma di tali sorgenti distinte di variabilità.

Tradizionalmente i risultati dei test F-ANOVA ad una via sono presentati come in *Tabella (10.1).*

Gli Esempi che seguono illustrano l'applicazione del test F-ANOVA ad una via.

ESEMPIO 10.2. ▶

Tre diverse aziende producono lo stesso componente microscopico. I diametri di alcuni pezzi (espressi in μm) sono mostrati nella *Tabella (10.2).*

L'ipotesi che si vuole verificare è che le tre aziende producano pezzi con lo stesso diametro medio. L'analisi della varianza ad una via conduce ai seguenti risultati.

Causa	**Somma dei Quadrati**	**Gradi di Libertà**	**Media**	***F*-ANOVA**
Esterna	1140	2	570	4.28
Interna	1600	12	133.33	

Causa	Somma dei Quadrati	Gradi di Libertà	Media	F-ANOVA
Esterna	$\text{SEQ} = \sum_{i=1}^{k} n_i \left(\overline{Y}_{i\bullet} - \overline{\overline{Y}}\right)^2$	$k-1$	$\frac{\text{SEQ}}{k-1}$	$F = \frac{\text{SEQ}/(k-1)}{\text{SIQ}/(N-k)}$
Interna	$\text{SIQ} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(Y_{ij} - \overline{Y}_{i\bullet}\right)^2$	$N-k$	$\frac{\text{SIQ}}{N-k}$	

Tabella 10.1: presentazione standard dei risultati del test F-ANOVA ad una via.

Azienda A	Azienda B	Azienda C
40	60	60
30	40	50
50	55	70
50	65	65
40		75
		40

Tabella 10.2: i dati dell'*Esempio (10.2)*.

Al livello $\alpha = 0.05$ si ha $f_{2,12,0.95} \approx 3.89$, e dunque si rifiuterà l'ipotesi $\mathcal{H}_0 : \theta_1 = \theta_2 = \theta_3$; invece, al livello $\alpha = 0.01$ si ha $f_{2,12,0.99} \approx 6.93$, e dunque si accetterà l'ipotesi $\mathcal{H}_0$. ◀

ESEMPIO 10.3. ▶

In una università il corso di *Statistica* è suddiviso in tre sezioni distinte, ciascuna con un docente diverso. La *Tabella (10.3)* mostra i voti finali degli studenti di ciascun corso (espressi in centesimi).

Docente A	Docente B	Docente C
95	88	68
33	78	79
48	91	91
76	51	71
89	85	87
82	77	68
60	31	79
77	62	16
	96	35
	81	

Tabella 10.3: i dati dell'*Esempio (10.3)*.

L'ipotesi che si vuole verificare è se i voti medi θ_i siano gli stessi. L'analisi della varianza ad una via conduce ai seguenti risultati.

Causa	Somma dei Quadrati	Gradi di Libertà	Media	*F*-ANOVA
Esterna	303.41	2	151.70	0.31
Interna	11752.00	24	489.67	

Al livello $\alpha = 0.05$ si ha $f_{2,24,0.95} \approx 3.40$, e dunque si accetterà l'ipotesi $\mathcal{H}_0 \colon \theta_1 = \theta_2 = \theta_3$. ◀

10.1.3 Stima simultanea di contrasti

Nella pratica è spesso di interesse fare inferenze simultanee su più contrasti. Purtroppo si sa che operando una verifica di ipotesi simultanea, utilizzando test di livello α, non conduce necessariamente ad un test ancora di livello α. Si pensi, ad esempio, di voler fare inferenze sulle differenze $\theta_i - \theta_j$, come spesso è richiesto nelle applicazioni. Usando una notazione intuitiva, si definiscano i seguenti IC per le differenze $\theta_i - \theta_j$:

$$I_{ij} = \left((\overline{Y}_{i\bullet} - \overline{Y}_{j\bullet}) - t_{N-k,1-\frac{\alpha}{2}} \sqrt{\overline{S}_P^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}, (\overline{Y}_{i\bullet} - \overline{Y}_{j\bullet}) + t_{N-k,1-\frac{\alpha}{2}} \sqrt{\overline{S}_P^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)} \right),$$

dove $i, j = 1, \ldots, k$ e $i \neq j$. Allora sarebbe $\mathbf{P}\{\theta_i - \theta_j \in I_{ij}\} = 1-\alpha$ per ciascun intervallo I_{ij} ma, considerando più intervalli contemporaneamente, si avrebbe, ad esempio, $\mathbf{P}\{\{\theta_1 - \theta_2 \in I_{12}\} \cap \{\theta_2 - \theta_3 \in I_{23}\}\} < 1 - \alpha$.

Il problema può essere affrontato in diversi modi. Grazie alla *Disuguaglianza di Bonferroni* sappiamo che, volendo fare inferenza simultanea su m eventi distinti, dovremmo fare in modo che ciascun IC abbia un livello γ tale che

$$1 - \alpha = \sum_{i=1}^{m} \gamma - (m-1),$$

cioè $\gamma = 1 - \frac{\alpha}{m}$. In realtà non è necessario che tutti gli intervalli abbiano lo stesso livello: infatti, è sufficiente che tra i diversi livelli γ_i sussista la relazione

$$1 - \alpha = \sum_{i=1}^{m} \gamma_i - (m-1).$$

Dunque, in un modello ANOVA con k trattamenti, volendo fare inferenza simultanea su tutte le $m = \binom{k}{2}$ differenze ad un livello $\alpha \in (0,1)$, è sufficiente che ciascun intervallo abbia livello $1 - \frac{\alpha}{m}$.

Una procedura alternativa, ideata da *Scheffé* nel 1959 e detta *Metodo S*, consente di fare inferenze simultanee su *tutti* i contrasti. Tale metodologia è da preferire alla precedente solo se il numero di contrasti è elevato: infatti, il prezzo che spesso si deve pagare per ottenere una così ampia generalità risiede, inevitabilmente, nella eccessiva lunghezza degli IC derivati o nella scarsa potenza dei test prodotti. Il Teorema che segue sancisce le caratteristiche del *Metodo S*.

TEOREMA 10.4. *Nelle ipotesi* ANOVA*, se*

$$M = \sqrt{(k-1)\, f_{k-1,N-k,1-\alpha}}, \tag{10.17}$$

dove $\alpha \in (0,1)$*, allora è* $1 - \alpha$ *la probabilità che*

$$\sum_{i=1}^{k} a_i \overline{Y}_{i\bullet} - M \sqrt{\overline{S}_P^2 \sum_{i=1}^{k} a_i^2/n_i} \leq \sum_{i=1}^{k} a_i \theta_i \leq \sum_{i=1}^{k} a_i \overline{Y}_{i\bullet} + M \sqrt{\overline{S}_P^2 \sum_{i=1}^{k} a_i^2/n_i}$$

simultaneamente per ogni $\boldsymbol{a} \in \mathfrak{A}$.

Come già accennato, per quanto i risultati del *Metodo S* siano validi per tutti i contrasti, può accadere che le inferenze derivate siano poco interessanti: riprova ne è il fatto che, confrontando i valori di $t_{r,1-\alpha/2}$ e $\sqrt{(k-1)\, f_{k-1,r,1-\alpha}}$, si ha $t_{r,1-\alpha/2} \leq \sqrt{(k-1)\, f_{k-1,r,1-\alpha}}$ per qualsiasi combinazione di r, α, k, ciò che determina un allungamento degli IC e una diminuzione della potenza dei test.

10.2 Analisi della varianza a due vie

In molte applicazioni è di interesse lo studio dell'effetto combinato di due fattori, diciamo A e B: ad esempio, la resa di un appezzamento agricolo può dipendere sia dalla varietà di grano coltivato sia dal tipo di fertilizzante utilizzato, oppure la qualità di un corso universitario può dipendere sia dal docente sia dal numero di studenti frequentanti. La tecnica ANOVA *a due vie* (dal termine anglosassone *two-ways* — si veda, ad esempio, [24]) è stata ideata per affrontare problemi di questo tipo. Al fine di poter effettuare stime statistiche e verifiche di ipotesi è necessario chiarire quale sia il contesto probabilistico nel quale il problema si colloca.

DEFINIZIONE 10.4 (ANOVA a due vie)**.** Le v.a. Y_{ij} sono osservate secondo il modello

$$Y_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ij}, \quad i = 1, \ldots, a, \quad j = 1, \ldots b, \tag{10.18}$$

dove:

1. μ rappresenta il livello medio globale;
2. α_i sono gli effetti indotti dal fattore A;
3. β_j sono gli effetti indotti dal fattore B;
4. γ_{ij} rappresentano le interazioni tra A e B;
5. ϵ_{ij} rappresentano le fluttuazioni (errori), che si assumono i.i.d. con legge $\mathtt{N}\big(0, \sigma^2\big)$.

Si assume inoltre, senza perdita di generalità, che

$$\sum_{i=1}^{a} \alpha_i = 0, \tag{10.19a}$$

$$\sum_{j=1}^{b} \beta_j = 0, \tag{10.19b}$$

$$\sum_{j=1}^{b} \gamma_{ij} = 0 \quad \text{per ogni } i = 1, \ldots, a, \tag{10.19c}$$

$$\sum_{i=1}^{a} \gamma_{ij} = 0 \quad \text{per ogni } j = 1, \ldots, b. \tag{10.19d}$$

Nel seguito indicheremo con Y_{ijs}, $s = 1, \ldots, m$, la s-esima osservazione quando il fattore A è al livello i e il fattore B è al livello j; supporremo inoltre che $m > 1$. Schematicamente le osservazioni possono essere raccolte in una tabella come quella che segue.

Livello del Fattore A	**Livello del Fattore B**			
	1	**2**	**...**	***b***
1	y_{111} $\vdots$ y_{11m}	y_{121} $\vdots$ y_{12m}	$\cdots$ $\ldots$ $\cdots$	y_{1b1} $\vdots$ y_{1bm}
2	y_{211} $\vdots$ y_{21m}	y_{221} $\vdots$ y_{22m}	$\cdots$ $\ldots$ $\cdots$	y_{2b1} $\vdots$ y_{2bm}
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$
a	y_{a11} $\vdots$ y_{a1m}	y_{a21} $\vdots$ y_{a2m}	$\cdots$ $\ldots$ $\cdots$	y_{ab1} $\vdots$ y_{abm}

Dunque il modello di riferimento è

$$Y_{ijs} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijs}, \quad i = 1, \dots, a, \quad j = 1, \dots b, \quad s = 1, \dots m, \tag{10.20}$$

con gli stessi vincoli dati nella *Definizione (10.4)*. Si noti che nel modello ANOVA a due vie si assume che gli effetti prodotti dai fattori A e B siano *additivi*. Tra le numerose ipotesi che si possono verificare, quelle che seguono giocano un ruolo di grande importanza nelle applicazioni:

$$\mathcal{H}_A \colon \alpha_1 = \cdots = \alpha_a = 0, \tag{10.21a}$$

$$\mathcal{H}_B \colon \beta_1 = \cdots = \beta_b = 0, \tag{10.21b}$$

$$\mathcal{H}_\gamma \colon \gamma_{ij} = 0 \quad \text{per ogni indice } i, j. \tag{10.21c}$$

Le prime due ipotesi mirano a verificare che, rispettivamente, il fattore A o B non produca effetti sul fenomeno in esame, mentre la terza riguarda l'assenza di interazioni tra i due fattori. Per procedere alla verifica delle ipotesi date utilizzeremo le seguenti statistiche, dove $n = abm$:

$$\overline{Y} = \frac{1}{n} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{s=1}^{m} Y_{ijs}, \tag{10.22a}$$

$$\overline{Y}_{ij\bullet} = \frac{1}{m} \sum_{s=1}^{m} Y_{ijs}, \tag{10.22b}$$

$$\overline{Y}_{i\bullet\bullet} = \frac{1}{mb} \sum_{j=1}^{b} \sum_{s=1}^{m} Y_{ijs}, \tag{10.22c}$$

$$\overline{Y}_{\bullet j\bullet} = \frac{1}{ma} \sum_{i=1}^{a} \sum_{s=1}^{m} Y_{ijs}. \tag{10.22d}$$

I parametri del modello si possono stimare utilizzando, ad esempio, i rispettivi SMV:

$$\hat{\mu} = \overline{Y}, \tag{10.23a}$$

$$\hat{\gamma}_{ij} = \overline{Y}_{ij\bullet} - \overline{Y}_{i\bullet\bullet} - \overline{Y}_{\bullet j\bullet} + \overline{Y}, \quad \text{per ogni indice } i, j, \tag{10.23b}$$

$$\hat{\alpha}_i = \overline{Y}_{i\bullet\bullet} - \overline{Y}, \quad i = 1, \dots, a, \tag{10.23c}$$

$$\hat{\beta}_j = \overline{Y}_{\bullet j\bullet} - \overline{Y}, \quad j = 1, \dots, b. \tag{10.23d}$$

Procedendo come nella *Sezione 10.1*, possiamo pensare di verificare ciascuna delle ipotesi date tramite opportuni test F-ANOVA a due vie; i risultati di cui ci avvarremo sono raccolti nel Teorema che segue.

TEOREMA 10.5 (***F***-ANOVA a due vie)**.** *Per il modello* ANOVA *a due vie valgono i seguenti risultati.*

1. *Nell'ipotesi $\mathcal{H}_A$, la statistica F_A data da*

$$F_A = \frac{ab(m-1)}{a-1} \frac{mb \sum_i \left(\overline{Y}_{i\bullet\bullet} - \overline{Y}\right)^2}{\sum_i \sum_j \sum_s \left(Y_{ijs} - \overline{Y}_{ij\bullet}\right)^2} \tag{10.24a}$$

ha legge $\mathtt{F}(a-1, ab(m-1))$; dunque, al livello $\alpha \in (0,1)$, si rifiuta $\mathcal{H}_A$ se

$$f_A > f_{a-1, ab(m-1), 1-\alpha}. \tag{10.24b}$$

2. *Nell'ipotesi $\mathcal{H}_B$, la statistica F_B data da*

$$F_B = \frac{ab(m-1)}{b-1} \frac{ma \sum_j \left(\overline{Y}_{\bullet j\bullet} - \overline{Y}\right)^2}{\sum_i \sum_j \sum_s \left(Y_{ijs} - \overline{Y}_{ij\bullet}\right)^2} \tag{10.25a}$$

ha legge $\mathtt{F}(b-1, ab(m-1))$; dunque, al livello $\alpha \in (0,1)$, si rifiuta $\mathcal{H}_B$ se

$$f_B > f_{b-1, ab(m-1), 1-\alpha}. \tag{10.25b}$$

3. Nell'ipotesi $\mathcal{H}_\gamma$, la statistica F_γ data da

$$F_\gamma = \frac{ab(m-1)}{(a-1)(b-1)} \frac{m \sum_i \sum_j \left(\overline{Y}_{ij\bullet} - \overline{Y}_{i\bullet\bullet} - \overline{Y}_{\bullet j\bullet} + \overline{Y}\right)^2}{\sum_i \sum_j \sum_s \left(Y_{ijs} - \overline{Y}_{ij\bullet}\right)^2} \tag{10.26a}$$

ha legge $\mathtt{F}((a-1)(b-1), ab(m-1))$*; dunque, al livello* $\alpha \in (0,1)$*, si rifiuta* $\mathcal{H}_\gamma$ *se*

$$f_\gamma > f_{(a-1)(b-1),ab(m-1),1-\alpha}. \tag{10.26b}$$

Tradizionalmente i risultati dei test F-ANOVA a due vie sono presentati come in *Tabella (10.4)*, dove le variabili utilizzate valgono:

$$\begin{aligned}
\text{SQA} &= mb \sum_i \left(\overline{Y}_{i\bullet\bullet} - \overline{Y}\right)^2, \\
\text{SQB} &= ma \sum_j \left(\overline{Y}_{\bullet j\bullet} - \overline{Y}\right)^2, \\
\text{SQI} &= m \sum_i \sum_j \left(\overline{Y}_{ij\bullet} - \overline{Y}_{i\bullet\bullet} - \overline{Y}_{\bullet j\bullet} + \overline{Y}\right)^2, \\
\text{SQE} &= \sum_i \sum_j \sum_s \left(Y_{ijs} - \overline{Y}_{ij\bullet}\right)^2.
\end{aligned}$$

Causa	**Somma dei Quadrati**	**Gradi di Libertà**	**Media**	$\boldsymbol{F}$**-ANOVA**
A	SQA	$a-1$	$\text{MSQA} = \frac{\text{SQA}}{a-1}$	$F_A = \frac{\text{MSQA}}{\text{MSQE}}$
B	SQB	$b-1$	$\text{MSQB} = \frac{\text{SQB}}{b-1}$	$F_B = \frac{\text{MSQB}}{\text{MSQE}}$
γ	SQI	$(a-1)(b-1)$	$\text{MSQI} = \frac{\text{SQI}}{(a-1)(b-1)}$	$F_\gamma = \frac{\text{MSQI}}{\text{MSQE}}$
Errore	SQE	$ab(m-1)$	$\text{MSQE} = \frac{\text{SQE}}{ab(m-1)}$	

Tabella 10.4: presentazione standard dei risultati dei test F-ANOVA a due vie.

L'Esempio che segue illustra una applicazione del test F-ANOVA a due vie.

ESEMPIO 10.4. ►

A tre docenti diversi (D_1, D_2 e D_3) viene chiesto di impartire lo stesso corso di *Statistica* a gruppi di quattro studenti secondo tre differenti metodi di insegnamento (I, II e III). La *Tabella (10.5)-(a)* mostra i voti finali degli studenti (espressi in centesimi).

Le ipotesi che si vogliono verificare sono:

- $\mathcal{H}_A$*: le tre metodologie sono ugualmente efficaci*;
- $\mathcal{H}_B$*: i tre docenti sono didatticamente equivalenti*;
- $\mathcal{H}_\gamma$*: non vi sono interazioni tra metodologie e docenti*.

L'analisi della varianza a due vie conduce ai risultati mostrati in *Tabella (10.5)-(b)*. La Tabella F-ANOVA a due vie risulta perciò essere:

(a)

	Docente		
Metodo	D_1	D_2	D_3
I	95	60	86
	85	90	77
	74	80	75
	74	70	70
II	90	89	83
	80	90	70
	92	91	75
	82	86	72
III	70	68	74
	80	73	86
	85	78	91
	85	93	89

(b)

	$\overline{y}_{ij\bullet}$			
	D_1	D_2	D_3	$\overline{y}_{i\bullet\bullet}$
I	82	75	77	78.0
II	86	89	75	83.3
III	80	78	85	81.0
$\overline{y}_{\bullet j\bullet}$	82.7	80.7	79.0	$\overline{y} = 80.8$

Tabella 10.5: *(a)* i dati dell'*Esempio (10.4)*; *(b)* i risultati dell'analisi della varianza a due vie.

Causa	**Somma dei Quadrati**	**Gradi di Libertà**	**Media**	F**-ANOVA**
Metodo	SQA $= 169.56$	2	MSQA $= 84.78$	$F_A = 1.25$
Docente	SQB $= 82.32$	2	MSQB $= 41.16$	$F_B = 0.61$
Interaz.	SQI $= 561.80$	4	MSQI $= 140.45$	$F_\gamma = 2.07$
Errore	SQE $= 1830.00$	27	MSQE $= 67.78$	

Al livello $\alpha = 0.05$ si ha $f_{2,27,0.95} \approx 3.35$ e $f_{4,27,0.95} \approx 2.73$; dunque si accetteranno tutte le ipotesi enunciate. ◀

Capitolo 11

REGRESSIONE LINEARE

Un problema statistico molto frequente nelle applicazioni è quello in cui si considera una variabile Y che è funzione di altre variabili $X_1, \ldots, X_k$, in aggiunta ad una perturbazione aleatoria. Ad esempio, X potrebbe rappresentare la temperatura e Y la viscosità di un fluido, oppure X potrebbe essere la concentrazione di fertilizzante nel suolo e Y la quantità di sostanza assorbita da una pianta; in entrambi i casi è opportuno tenere conto, tramite l'introduzione di una perturbazione aleatoria, di possibili fluttuazioni casuali nei valori della variabile Y causate, ad esempio, dalle variazioni aleatorie della pressione a cui è sottoposto il fluido o dalla concentrazione disomogenea del fertilizzante nel suolo. Gli esempi si potrebbero facilmente moltiplicare.

Le tecniche di *regressione* (e, in particolare, quelle di *regressione lineare*) sono uno strumento ampiamente usato in Statistica Matematica (si vedano, ad esempio, [24, 33, 6, 1]). Lo scopo principale della regressione è quello di studiare la dipendenza di una variabile rispetto ad un'altra (o ad altre). Esistono numerose forme di regressione: lineare, non lineare, semplice, multipla, parametrica, non parametrica, ecc... Nel seguito ci occuperemo di regressione lineare, sia semplice sia multipla.

11.1 Regressione lineare semplice

Il modello di base risulta essere come segue.

DEFINIZIONE 11.1 (Regressione lineare semplice)**.** Nella regressione lineare semplice il modello è:

$$Y_i = \theta_0 + \theta_1 x_i + \epsilon_i, \quad i = 1, \ldots, n, \tag{11.1}$$

dove Y_i è una v.a. osservata detta *risposta*, x_i è una variabile osservata detta *predittore*, e ϵ_i è una v.a. detta *fluttuazione casuale* (o *errore*); i parametri reali θ_0 e θ_1, detti *intercetta* e *pendenza* della regressione, sono incogniti ma costanti.

NOTA 11.1.

Nel linguaggio comune solitamente Y è detta *variabile dipendente*, mentre x è detta *variabile indipendente*. Tale terminologia è evidentemente imprecisa, almeno da un punto di vista probabilistico: infatti, non è detto che le variabili x_i siano aleatorie, anzi tipicamente supporremo che siano deterministiche. Perciò è preferibile usare per Y il termine di *risposta* e per x quello di *predittore*.

Senza perdita di generalità si suppone che sia

$$\mathbf{E}(\epsilon_i) = 0 \tag{11.2}$$

per ogni indice i, da cui risulta

$$\mathbf{E}(Y_i) = \theta_0 + \theta_1 x_i. \tag{11.3}$$

Ciò conduce alla seguente definizione.

DEFINIZIONE 11.2 (Funzione di regressione)**.** Nel modello di regressione la funzione

$$r(x) = \mathbf{E}(Y) = \theta_0 + \theta_1 x \tag{11.4}$$

è detta *funzione di regressione della popolazione*.

OSSERVAZIONE 11.1.

Il termine "regressione lineare" si riferisce al fatto che il modello è *lineare nei parametri* θ_0 *e* θ_1, ovvero che r è una funzione affine dei due. Ad esempio, entrambi i modelli

$$r(x) = \mathbf{E}(Y) = \theta_0 + \theta_1 x^2 \quad \text{e} \quad r(x) = \mathbf{E}(\ln Y) = \theta_0 + \theta_1 \frac{1}{x}$$

ricadono nell'ambito della regressione lineare: nel primo caso si ipotizza una relazione lineare tra Y e x^2, nel secondo tra $\ln Y$ e $\frac{1}{x}$. Al contrario, la specifica $r(x) = \mathbf{E}(Y) = \theta_0 + \theta_1^2 x$ non definisce una regressione lineare.

Uno degli scopi principali della regressione è quello di *predirre* i valori di Y per particolari valori di x, tipicamente sfruttando la funzione di regressione r. Per quanto detto nell'*Osservazione (11.1)* è chiaro che numerose sono le relazioni tra la risposta e il predittore (o i predittori, come vedremo più avanti nella *Sezione 11.2*) che è possibile modellizzare attraverso una regressione lineare.

Nei problemi di regressione lineare semplice le osservazioni del predittore e della risposta consistono di coppie di valori $(x_1, y_1), \ldots, (x_n, y_n)$. Per l'analisi dei dati osservati diverse sono le ipotesi che si possono formulare, a seconda che i valori x o i valori y, o entrambi, siano interpretati quali osservazioni di v.a. X o Y. In generale le n osservazioni non cadono esattamente su di una retta nel piano (x, y); tuttavia, saremo interessati a calcolare la retta che "meglio" si adatti ai dati disponibili, secondo criteri che enunceremo di volta in volta. Evidentemente il problema principale risiede nella stima dei parametri θ_0 e θ_1, che affronteremo in diversi modi. Definiamo ora alcune quantità che utilizzeremo nel seguito, ove confonderemo spesso le notazioni y e Y. Le variabili

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \quad \text{e} \quad \overline{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \tag{11.5a}$$

sono dette *medie campionarie*. Le variabili

$$S_{xx} = \sum_{i=1}^{n} (x_i - \overline{x})^2 \quad \text{e} \quad S_{yy} = \sum_{i=1}^{n} (y_i - \overline{y})^2 \tag{11.5b}$$

sono dette *somme dei quadrati*. La variabile

$$S_{xy} = \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) \tag{11.5c}$$

è detta *somma incrociata dei prodotti*.

11.1.1 Il metodo dei Minimi Quadrati (Interpolazione)

Un primo modo per stimare i parametri θ_0 e θ_1 è il cosiddetto *metodo dei Minimi Quadrati* (MQ), che non richiede alcuna ipotesi di natura probabilistica: in effetti, tale metodologia è assai generale, ed è comunemente utilizzata al di fuori del contesto statistico; anzi, proprio per queste ragioni, i risultati forniti dal metodo MQ sono spesso apostrofati dagli statistici come "soluzioni" anziché come "stime" (termine, quest'ultimo, di chiara valenza probabilistica).

Si considerino n osservazioni $(x_1, y_1), \ldots, (x_n, y_n)$, ove si assume che vi siano almeno due coppie con $x_i \neq x_j$. Il criterio che si adotta per la ricerca di θ_0 e θ_1 (a partire da *Gauss* che lo introdusse in un contesto ben più generale dell'attuale), è quello di calcolare la retta $\theta_0 + \theta_1 x$ che minimizzi la quantità

$$\sum_{i=1}^{n} (y_i - (\theta_0 + \theta_1 x_i))^2 = \sum_{i=1}^{n} r_i^2. \tag{11.6}$$

Evidentemente ciò corrisponde alla somma dei quadrati delle distanze (verticali) tra le osservazioni y_i e i punti della retta $\theta_0 + \theta_1 x$ di ascissa x_i, ovvero alla somma dei quadrati dei *residui* r_i: dunque, si tratta di un classico problema di ricerca del minimo di una funzione quadratica. I valori $\hat{\theta}_0$ e $\hat{\theta}_1$ che rendono minima la funzione quadratica $\varphi(\theta_0, \theta_1) = \sum_{i=1}^{n} (y_i - (\theta_0 + \theta_1 x_i))^2$ sono quelli che risolvono le equazioni

$$\frac{\partial \varphi}{\partial \theta_0} = -2 \sum_{i=1}^{n} (y_i - (\theta_0 + \theta_1 x_i)) = 0, \tag{11.7a}$$

$$\frac{\partial \varphi}{\partial \theta_1} = -2 \sum_{i=1}^{n} x_i \left(y_i - (\theta_0 + \theta_1 x_i) \right) = 0. \tag{11.7b}$$

La soluzione[1] è data da

$$\hat{\theta}_0 = \overline{y} - \frac{S_{xy}}{S_{xx}} \overline{x} = \overline{y} - \hat{\theta}_1 \overline{x}, \tag{11.8a}$$

$$\hat{\theta}_1 = \frac{S_{xy}}{S_{xx}}. \tag{11.8b}$$

Possiamo raccogliere tale risultato nella seguente Definizione.

DEFINIZIONE 11.3 (Retta di regressione MQ)**.** Con $\hat{\theta}_0$ e $\hat{\theta}_1$ dati dalle *equazioni (11.8)*, la funzione

$$r(x) = \hat{\theta}_0 + \hat{\theta}_1 x \tag{11.9}$$

è detta *retta di regressione dei Minimi Quadrati*.

Come già accennato, il metodo MQ è da intendersi esclusivamente quale strumento per interpolare, tramite una retta, i dati disponibili, e non quale metodologia per effettuare inferenze statistiche: in effetti, per le ipotesi fatte, non saremmo in grado di costruire IC o di procedere a verifiche di ipotesi. Tuttavia, vedremo nella *Sezione 11.1.2* che le soluzioni fornite dal metodo MQ godono di proprietà di ottimalità in determinati contesti statistici.

ESEMPIO 11.1 (Acqua alta a Venezia)**.** ▶
La *Tabella (11.1)* mostra i livelli massimi dell'acqua alta a Venezia (in *cm*) osservati dal 1931 al 1980 (i dati sono tratti da [32]). Qui la variabile "Anno" è deterministica e gioca il ruolo di predittore, mentre il livello massimo dell'acqua alta è evidentemente aleatorio e corrisponde alla risposta del sistema.

Anno	cm	Anno	cm	Anno	cm	Anno	cm	Anno	cm
1931	103	1941	99	1951	151	1961	122	1971	122
1932	78	1942	91	1952	116	1962	114	1972	120
1933	121	1943	97	1953	107	1963	118	1973	114
1934	116	1944	106	1954	112	1964	107	1974	96
1935	115	1945	105	1955	97	1965	110	1975	125
1936	147	1946	136	1956	95	1966	194	1976	124
1937	119	1947	126	1957	119	1967	138	1977	120
1938	114	1948	132	1958	124	1968	144	1978	132
1939	89	1949	104	1959	118	1969	138	1979	166
1940	102	1950	117	1960	145	1970	123	1980	134

Tabella 11.1: i dati dell'*Esempio (11.1)*.

La *Figura (11.1)* mostra le $n = 50$ coppie di osservazioni disponibili e la retta di regressione di y su x avente equazione $y = -969.27 + 0.56x$: a prima vista l'interpolazione è da considerarsi soddisfacente. ◀

OSSERVAZIONE 11.2.
Il criterio sin qui adottato nella ricerca di θ_0 e θ_1 non è l'unico possibile. Infatti, ad esempio, si potrebbe pensare di minimizzare la somma dei quadrati delle distanze *orizzontali* tra le osservazioni e i punti della retta $\theta_0 + \theta_1 x$. La risoluzione di tale problema si ottiene in maniera analoga a quanto appena visto: semplicemente si tratta di scambiare i ruoli di x e y, ovvero il predittore diventa la risposta e viceversa. Dunque, non sorprende il fatto che, in generale, la soluzione cambi; in particolare, avremo

$$\hat{\theta}'_0 = \overline{x} - \frac{S_{xy}}{S_{yy}} \overline{y} = \overline{x} - \hat{\theta}'_1 \overline{y}, \tag{11.10a}$$

[1] Qui e nel seguito trascureremo di verificare che la soluzione trovata corrisponda effettivamente ad un minimo, un compito comunque non difficile trattandosi della minimizzazione di una funzione quadratica.

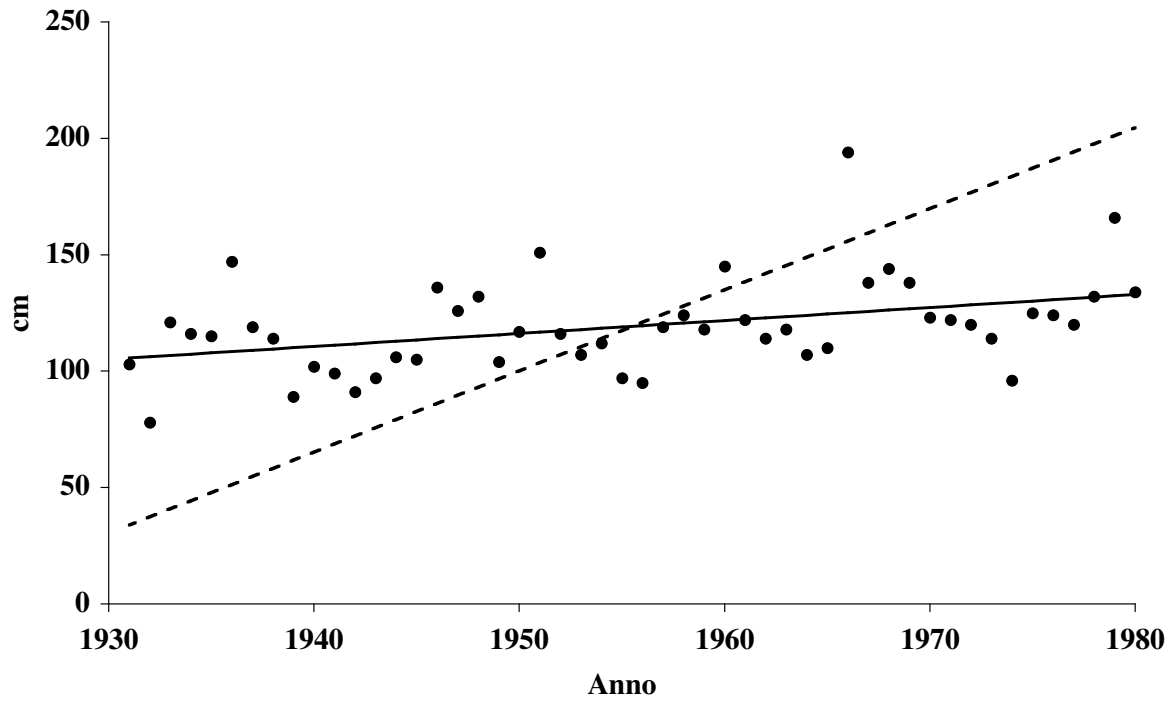

Figura 11.1: i dati dell'*Esempio (11.1)*, la retta di regressione di y su x (linea *continua*) e la retta di regressione di x su y (linea *tratteggiata*).

$$\hat{\theta}_1' = \frac{S_{xy}}{S_{yy}}, \tag{11.10b}$$

dove, tipicamente, $\hat{\theta}_0' \neq \hat{\theta}_0, \hat{\theta}_1' \neq \hat{\theta}_1$.

Tornando all'*Esempio (11.1)*, la *Figura (11.1)* mostra la retta di regressione di x su y, ottenuta scambiando tra loro risposta e predittore, avente equazione $y = -6694.35 + 3.48x$: evidentemente l'interpolazione non può considerarsi soddisfacente.

I risultati della regressione con il metodo MQ impongono alcune considerazioni metodologiche. Il fatto che il campione esaminato si riferisca a *livelli massimi* di acqua alta suggerirebbe l'uso di modelli GEV o GP per lo studio statistico del fenomeno. Tuttavia, la possibile presenza di un "*trend*" crescente potrebbe avvalorare l'ipotesi che i dati non provengano dalla stessa distribuzione "parente". In tal caso cadrebbe una delle ipotesi fondamentali alla base della Teoria dei Valori Estremi e l'utilizzo dei modelli GEV o GP sarebbe legittimo o opportuno solo se preceduto da opportune operazioni di "de-trending" dei dati disponibili (ma la spiegazione di tali metodologie va oltre gli intenti di questo Corso).

11.1.2 Stimatori BLUE

In questa Sezione mostreremo che le stime dei parametri θ_0 e θ_1 fornite dalle *equazioni (11.8)* sono ottimali nella classe degli stimatori lineari non distorti nell'ambito di un modello statistico alquanto generale.

Si considerino $n > 2$ osservazioni $(x_1, y_1), \ldots, (x_n, y_n)$, ove si assume che sia $x_i \neq x_j$ se $i \neq j$. I valori $x_1, \ldots, x_n$ sono noti e costanti. Invece, i valori $y_1, \ldots, y_n$ sono osservazioni di v.a. $Y_1, \ldots, Y_n$ che si assumono non correlate. Il modello è specificato dalle relazioni

$$\mathbf{E}(Y_i) = \theta_0 + \theta_1 x_i, \quad i = 1, \ldots, n, \tag{11.11a}$$

con

$$\mathbf{V}(Y_i) = \sigma^2, \quad i = 1, \ldots, n. \tag{11.11b}$$

Si noti che si assume che tutte le v.a. abbiano la stessa varianza incognita σ^2 (*omoschedasticità*); tale parametro è detto *varianza residua*. Le ipotesi specificate circa i momenti dei primi due ordini (ovvero, speranza, varianza e covarianza) sono le uniche di cui ci avverremo nel seguito. Il modello descritto dalle *equazioni (11.11)* può essere formulato anche nel seguente modo: presi $\theta_0, \theta_1 \in \mathbb{R}$ e $\sigma > 0$, si assume che

$$Y_i = \theta_0 + \theta_1 x_i + \sigma\epsilon_i, \quad i = 1, \dots, n, \tag{11.12a}$$

con

$$\mathbf{E}(\epsilon_i) = 0 \quad \text{e} \quad \mathbf{V}(\epsilon_i) = 1, \quad i = 1, \dots, n. \tag{11.12b}$$

Le v.a. $\epsilon_1, \dots, \epsilon_n$ sono dette *fluttuazioni casuali* e sono non correlate.

NOTA 11.2.

Essendo non correlate, le v.a. $\epsilon_1, \dots, \epsilon_n$ hanno matrice di varianza-covarianza $\boldsymbol{C} = \sigma^2 \boldsymbol{I}_n$. Se, invece, fossero correlate, sarebbe $\boldsymbol{C} = \sigma^2 \boldsymbol{G}$, con $\boldsymbol{G}$ una matrice $n \times n$ *definita positiva*. Dunque, sarebbe possibile definire la matrice $\boldsymbol{G}^{-1/2}$ e, operando una trasformazione opportuna, ricondursi al caso di v.a. non correlate.

Per la stima dei parametri θ_0 e θ_1 ci restringeremo alla classe degli *stimatori lineari*, ovvero a statistiche della forma $\sum_{i=1}^n c_i Y_i$, dove $c_1, \dots, c_n$ sono costanti note. Il problema diventa dunque quello di calcolare i valori c_i affinché gli stimatori siano non distorti e di varianza minima. A tale fine è fondamentale il Teorema che segue (per i risultati utilizzati nella dimostrazione si veda l'*Appendice G*). Prima di procedere è utile riscrivere il modello dato usando una notazione matriciale. Siano

$$\boldsymbol{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \boldsymbol{A} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \boldsymbol{\theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix}, \quad \boldsymbol{Z} = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}.$$

Dunque è

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{\theta} + \sigma\boldsymbol{Z}. \tag{11.13}$$

Si noti che $\boldsymbol{A}$ è iniettiva, essendo $x_i \neq x_j$ se $i \neq j$. Lo spazio dei parametri è

$$\Theta = \mathbb{R}^2 \times \mathbb{R}_+ = \{\theta_0 \in \mathbb{R}, \theta_1 \in \mathbb{R}, \sigma \in \mathbb{R}_+\}.$$

La legge $\mathbf{P}_\Theta$ di $\boldsymbol{Y}$ è la legge immagine di $\boldsymbol{Z}$ mediante l'applicazione affine $\boldsymbol{z} \mapsto \boldsymbol{A}\boldsymbol{\theta} + \sigma\boldsymbol{z}$, con $\boldsymbol{z} \in \mathbb{R}^n$.

TEOREMA 11.1 (Gauss–Markov (I))**.** *Nelle ipotesi date,*

$$\boldsymbol{U}\boldsymbol{Y} = (\boldsymbol{A}^\intercal \boldsymbol{A})^{-1} \boldsymbol{A}^\intercal \boldsymbol{Y} \tag{11.14}$$

è uno stimatore non distorto di $\boldsymbol{\theta}$ che è ottimale nella classe degli stimatori lineari non distorti.

DIM. Sia $\boldsymbol{V}\boldsymbol{Y}$ uno stimatore lineare di $\boldsymbol{\theta}$; qui $\boldsymbol{V}$ è una matrice $2 \times n$. La sua speranza rispetto a $\mathbf{P}_\Theta$ è

$$\mathbf{E}_\Theta(\boldsymbol{V}\boldsymbol{Y}) = \mathbf{E}_\Theta(\boldsymbol{V}\boldsymbol{A}\boldsymbol{\theta} + \sigma\boldsymbol{V}\boldsymbol{Z}) = \boldsymbol{V}\boldsymbol{A}\boldsymbol{\theta};$$

quindi $\boldsymbol{V}\boldsymbol{Y}$ è corretto se, e solo se, $\boldsymbol{V}\boldsymbol{A} = \boldsymbol{I}_2$. Ciò mostra che $\boldsymbol{U}\boldsymbol{Y}$ è uno stimatore corretto di $\boldsymbol{\theta}$; infatti

$$\boldsymbol{U}\boldsymbol{A} = (\boldsymbol{A}^\intercal \boldsymbol{A})^{-1} \boldsymbol{A}^\intercal \boldsymbol{A} = \boldsymbol{I}_2.$$

Se $\boldsymbol{V}\boldsymbol{Y}$ è corretto il suo Rischio vale, essendo le componenti di $\boldsymbol{Z}$ non correlate,

$$\begin{aligned} \mathbf{E}_\Theta\left(\|\boldsymbol{V}\boldsymbol{Y} - \boldsymbol{\theta}\|^2\right) &= \sigma^2 \mathbf{E}_\Theta\left(\|\boldsymbol{V}\boldsymbol{Z}\|^2\right) \\ &= \sigma^2 \mathbf{E}_\Theta(\langle \boldsymbol{V}\boldsymbol{Z}, \boldsymbol{V}\boldsymbol{Z} \rangle) \\ &= \sigma^2 \|\boldsymbol{V}\|^2. \end{aligned}$$

D'altro canto si ha, se $\boldsymbol{V}\boldsymbol{Y}$ è corretto e $\boldsymbol{P} = \boldsymbol{A}\boldsymbol{U}$ è la proiezione ortogonale di $\mathbb{R}^n$ su $\boldsymbol{A}\mathbb{R}^2$,

$$\boldsymbol{V}\boldsymbol{P} = \boldsymbol{V}(\boldsymbol{A}\boldsymbol{U}) = (\boldsymbol{V}\boldsymbol{A})\boldsymbol{U} = \boldsymbol{I}_2\boldsymbol{U} = \boldsymbol{U},$$

sicché, essendo la proiezione ortogonale simmetrica,

$$\boldsymbol{U}^\intercal = (\boldsymbol{VP})^\intercal = \boldsymbol{P}^\intercal \boldsymbol{V}^\intercal = \boldsymbol{PV}^\intercal.$$

Perciò, essendo $\boldsymbol{P}$ una proiezione ortogonale, è

$$\|\boldsymbol{U}\|^2 = \|\boldsymbol{U}^\intercal\|^2 = \|\boldsymbol{PV}^\intercal\|^2 \leq \|\boldsymbol{V}^\intercal\|^2 = \|\boldsymbol{V}\|^2,$$

ciò che mostra che $\boldsymbol{UY}$ è preferibile a $\boldsymbol{VY}$. □

OSSERVAZIONE 11.3.

Alla luce delle considerazioni illustrate nell'*Appendice G*, il risultato del *Teorema (11.1)* può essere visto sotto un'altra prospettiva. Si scomponga la matrice $\boldsymbol{A}$ nei suoi vettori colonna $\mathbf{1}_n, \boldsymbol{x}$; evidentemente questi sono linearmente indipendenti, essendo $\boldsymbol{A}$ iniettiva. Se $\boldsymbol{\theta}$ varia in $\mathbb{R}^2$, allora il vettore

$$\boldsymbol{v} = \boldsymbol{A\theta} = \mathbf{1}_n\theta_0 + \boldsymbol{x}\theta_1$$

varia in tutto il sottospazio $E = (\boldsymbol{A}\mathbb{R}^2)$ generato dai vettori $\mathbf{1}_n, \boldsymbol{x}$. Volendo stimare $\boldsymbol{\theta}$, si può pensare di utilizzare quel vettore che renda minima la distanza tra $\boldsymbol{v}$ e $\boldsymbol{Y}$. Dalla *Proposizione (G.4)* scende che il vettore di E avente distanza minima da $\boldsymbol{Y}$ è dato dalla proiezione ortogonale di $\boldsymbol{Y}$ su E; d'altro canto, dalla *Proposizione (G.3)* sappiamo che tale minimo si verifica in corrispondenza di $\hat{\boldsymbol{\theta}} = \boldsymbol{UY}$. Dunque, il calcolo di $\hat{\boldsymbol{\theta}}$ si riduce a quello della proiezione ortogonale di $\boldsymbol{Y}$ su E, che coincide con il risultato del *Teorema di Gauss–Markov (I)*. Volendo procedere al calcolo diretto dello stimatore $\hat{\boldsymbol{\theta}}$ di $\boldsymbol{\theta}$, sarebbe necessario minimizzare la funzione quadratica

$$\varphi(\theta_0, \theta_1) = \|\boldsymbol{Y} - \boldsymbol{A\theta}\|^2 = \sum_{i=1}^{n} (y_i - (\theta_0 + \theta_1 x_i))^2,$$

che rappresenta il quadrato della distanza tra $\boldsymbol{Y}$ e $\boldsymbol{A\theta}$. La soluzione al problema è già stata calcolata nella *Sezione 11.1.1* ed è data dalle *equazioni (11.8)*, cioè le stesse del metodo MQ, come mostrato nella *Proposizione (11.1)*.

Grazie al *Teorema (11.1)* è ora possibile calcolare gli stimatori dei parametri θ_0 e θ_1 che, per quanto già detto, godono di particolari proprietà di ottimalità.

PROPOSIZIONE 11.1 (Stimatori BLUE)**.** *Nelle ipotesi precedenti si ha*

$$\hat{\theta}_0 = \overline{y} - \frac{S_{xy}}{S_{xx}}\,\overline{x} = \overline{y} - \hat{\theta}_1\,\overline{x}, \tag{11.15a}$$

$$\hat{\theta}_1 = \frac{S_{xy}}{S_{xx}}. \tag{11.15b}$$

Tali stimatori sono detti BLUE*, dall'acronimo anglosassone "Best Linear Unbiased Estimators" (stimatori lineari non distorti ottimali).*

Evidentemente, gli stimatori BLUE sono formalmente gli stessi del metodo MQ classico, operando una regressione di y su x: la differenza risiede nel fatto che ora siamo in presenza di un modello statistico ben definito, e tali stimatori sono quelli di Rischio minimo nella classe degli stimatori lineari non distorti. Tuttavia, non essendo stata specificata una precisa legge di probabilità (ma solo i momenti dei primi due ordini) non siamo ancora in grado di poter costruire IC o procedere a verifiche di ipotesi, ciò che faremo nel seguito; per questa stessa ragione non ci interessiamo qui della stima del parametro σ, che tratteremo in un contesto più specifico.

Tornando all'*Esempio (11.1)* relativo allo studio dell'acqua alta a Venezia, nelle ipotesi fatte si ottiene che la retta di regressione BLUE è la stessa di quella del metodo MQ (operando una regressione di y su x) mostrata in *Figura (11.1)*.

11.1.3 Il modello Normale condizionale

In questa Sezione illustraremo il più comune tra i modelli di regressione lineare semplice ove, a differenza delle Sezioni precedenti, il contesto statistico sia definito specificando la legge di probabilità di riferimento.

Si considerino $n > 2$ osservazioni $(x_1, y_1), \ldots, (x_n, y_n)$, ove si assume che sia $x_i \neq x_j$ se $i \neq j$. I valori $x_1, \ldots, x_n$ sono noti e costanti. Invece, i valori $y_1, \ldots, y_n$ sono osservazioni di v.a. Normali $Y_1, \ldots, Y_n$ che si assumono indipendenti. Il *modello Normale condizionale* è specificato dalle relazioni

$$Y_i = \theta_0 + \theta_1 x_i + \sigma\epsilon_i, \quad i = 1, \ldots, n, \tag{11.16a}$$

dove le fluttuazioni casuali $\{\epsilon_i\}$ sono v.a. i.i.d. aventi legge

$$\epsilon_i \sim \mathtt{N}(0, 1)\,, \quad i = 1, \ldots, n. \tag{11.16b}$$

Si noti che si assume che tutte le v.a. abbiano la stessa varianza incognita σ^2 (*omoschedasticità*). Ora la legge di probabilità è ben specificata; in particolare, si ha che

$$Y_i \sim \mathtt{N}\big(\theta_0 + \theta_1 x_i, \sigma^2\big)\,, \quad i = 1, \ldots, n, \tag{11.17}$$

e inoltre le v.a. Y_i sono indipendenti. Utilizzando la notazione matriciale già vista nella *Sezione 11.1.2* si ha che, formalmente, il modello Normale condizionale è lo stesso di quello dato dall'*equazione (11.13)*. Ora, tuttavia, il vettore $\boldsymbol{Z}$ ha legge $\mathtt{N}(\boldsymbol{0}_n, \boldsymbol{I}_n)$, e dunque la legge $\mathbf{P}_\Theta$ di $\boldsymbol{Y}$ è $\mathbf{P}_\Theta = f_\Theta \cdot \lambda_n$ ove, al solito, λ_n è la misura n-dimensionale di Lebesgue e la densità f_Θ vale

$$f_\Theta(\boldsymbol{y}) = \frac{1}{(2\pi)^{n/2}\,\sigma^n}\, e^{-\frac{1}{2\sigma^2}\,\|\boldsymbol{y} - \boldsymbol{A\theta}\|^2}$$

per $\boldsymbol{y} \in \mathbb{R}^n$. Anche in questo caso lo spazio dei parametri è $\Theta = \mathbb{R}^2 \times \mathbb{R}_+ = \{\theta_0 \in \mathbb{R}, \theta_1 \in \mathbb{R}, \sigma \in \mathbb{R}_+\}$.

OSSERVAZIONE 11.4 (Statistiche sufficienti e complete nel modello Normale condizionale)**.** Grazie alla *Proposizione (G.1)* sappiamo che

$$\|\boldsymbol{y} - \boldsymbol{A\theta}\|^2 = \|\boldsymbol{y}\|^2 + \|\boldsymbol{A\theta}\|^2 - 2\,\langle \boldsymbol{Py}, \boldsymbol{A\theta}\rangle\,,$$

dove $\boldsymbol{P}$ è la proiezione ortogonale di $\mathbb{R}^n$ sul sottospazio $\big(\boldsymbol{A}\mathbb{R}^2\big)$. Pertanto

$$f_\Theta(\boldsymbol{y}) = \frac{1}{(2\pi)^{n/2}\,\sigma^n}\, e^{-\frac{1}{2\sigma^2}\,\|\boldsymbol{y}\|^2 - \frac{1}{2\sigma^2}\,\|\boldsymbol{A\theta}\|^2 + \frac{1}{\sigma^2}\,\langle \boldsymbol{Py}, \boldsymbol{A\theta}\rangle}.$$

Siamo dunque in presenza di un *modello esponenziale*, nel quale è possibile riconoscere in $(\boldsymbol{PY}, \|\boldsymbol{Y}\|^2)$ delle statistiche (congiuntamente) sufficienti e complete.

Ripetendo i ragionamenti già visti nel caso del *Teorema (11.1)*, si dimostra il seguente risultato, che si può (parzialmente) ottenere anche ricorrendo al metodo della Massima Verosimiglianza.

TEOREMA 11.2 (Gauss–Markov (II))**.** *Nelle ipotesi date,*

$$\boldsymbol{UY} = (\boldsymbol{A}^\intercal \boldsymbol{A})^{-1}\, \boldsymbol{A}^\intercal \boldsymbol{Y} \tag{11.18}$$

è uno stimatore non distorto di $\boldsymbol{\theta}$, e

$$\|\boldsymbol{AUY} - \boldsymbol{Y}\|^2 \tag{11.19}$$

è uno stimatore non distorto di $(n-2)\sigma^2$. Tutti gli stimatori sono ottimali nella classe degli stimatori non distorti di quadrato sommabile dei rispettivi parametri.

DIM. La parte relativa all'*equazione (11.18)* è già stata dimostrata nel *Teorema (11.1)*. Ora, sfruttando i risultati dell'*Appendice G*, si ha

$$\begin{aligned}
\boldsymbol{AUY} - \boldsymbol{Y} &= \boldsymbol{PY} - \boldsymbol{Y} \\
&= \boldsymbol{PA\theta} + \sigma\boldsymbol{PZ} - \boldsymbol{A\theta} - \sigma\boldsymbol{Z} \\
&= \boldsymbol{A\theta} + \sigma\boldsymbol{PZ} - \boldsymbol{A\theta} - \sigma\boldsymbol{Z} \\
&= \sigma\,(\boldsymbol{PZ} - \boldsymbol{Z}) \\
&= -\sigma\boldsymbol{P}^\perp\boldsymbol{Z},
\end{aligned}$$

che è una proiezione su un sottospazio di dimensione $n-2$. Essendo $\boldsymbol{Z}$ un vettore aleatorio Normale standard, dal *Teorema di Cochran* sappiamo che $\left\|\boldsymbol{P}^{\perp}\boldsymbol{Z}\right\|^2 \sim \chi^2(n-2)$, la cui speranza vale $n-2$; dunque

$$\begin{aligned}\mathbf{E}_\Theta\left(\|\boldsymbol{AUY}-\boldsymbol{Y}\|^2\right) &= \sigma^2\,\mathbf{E}_\Theta\left(\left\|\boldsymbol{P}^{\perp}\boldsymbol{Z}\right\|^2\right)\\ &= (n-2)\,\sigma^2.\end{aligned}$$

Le statistiche $(\boldsymbol{PY}, \|\boldsymbol{Y}\|^2)$ sono (congiuntamente) sufficienti e complete. Ora, $\boldsymbol{UP}=\boldsymbol{U}$: infatti

$$\begin{aligned}\boldsymbol{UP} &= (\boldsymbol{A}^\intercal\boldsymbol{A})^{-1}\,\boldsymbol{A}^\intercal\boldsymbol{P}\\ &= (\boldsymbol{A}^\intercal\boldsymbol{A})^{-1}\,\boldsymbol{A}^\intercal\boldsymbol{P}^\intercal\\ &= (\boldsymbol{A}^\intercal\boldsymbol{A})^{-1}\,(\boldsymbol{PA})^\intercal\\ &= (\boldsymbol{A}^\intercal\boldsymbol{A})^{-1}\,\boldsymbol{A}^\intercal,\end{aligned}$$

che è uguale a $\boldsymbol{U}$. Inoltre

$$\|\boldsymbol{AUY}-\boldsymbol{Y}\|^2 = \|\boldsymbol{PY}-\boldsymbol{Y}\|^2 = \|\boldsymbol{Y}\|^2 - \|\boldsymbol{PY}\|^2.$$

Ne scende che gli stimatori $\boldsymbol{PY}$ e $\boldsymbol{AUY}$ sono non distorti e funzioni di statistiche (congiuntamente) sufficienti e complete, e dunque ottimali. Ciò completa la dimostrazione. □

NOTA 11.3.

I risultati dei *Teoremi di Gauss–Markov (I-II)* (opportunamente modificati) continuano a valere anche quando la matrice $\boldsymbol{A}$ (iniettiva) ha dimensione $n \times k$, $2 < k < n$, ovvero quando c'è più di un predittore, come vedremo nella *Sezione 11.2* trattando della *regressione lineare multipla*.

Utilizzando il metodo della Massima Verosimiglianza, il calcolo degli stimatori porta al seguente risultato.

PROPOSIZIONE 11.2. *Nel modello Normale condizionale gli* SMV *dei parametri sono:*

$$\hat\theta_0 = \overline{y} - \frac{S_{xy}}{S_{xx}}\,\overline{x} = \overline{y} - \hat\theta_1\,\overline{x}, \tag{11.20a}$$

$$\hat\theta_1 = \frac{S_{xy}}{S_{xx}}, \tag{11.20b}$$

$$\hat\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}\Big(y_i - (\hat\theta_0 + \hat\theta_1 x_i)\Big)^2. \tag{11.20c}$$

Inoltre, $\left(\hat\theta_0, \hat\theta_1\right)$ *e* $\hat\sigma^2$ *sono indipendenti.*

Vedremo nel seguito quali siano le leggi degli stimatori appena calcolati.

OSSERVAZIONE 11.5.

Per quanto già visto, $\hat\theta_0$ e $\hat\theta_1$ sono stimatori lineari non distorti dei rispettivi parametri. Invece $\hat\sigma^2$ è distorto, essendo

$$\mathbf{E}_\Theta\big(\hat\sigma^2\big) = \frac{n-2}{n}\,\sigma^2. \tag{11.21}$$

Per questa ragione, nella pratica si usa il seguente stimatore corretto:

$$S^2 = \frac{n}{n-2}\,\hat\sigma^2 = \frac{1}{n-2}\sum_{i=1}^{n}\Big(y_i - (\hat\theta_0 + \hat\theta_1 x_i)\Big)^2. \tag{11.22}$$

Si noti come anche nel modello Normale condizionale la stima $\hat{\boldsymbol{\theta}}$ di $\boldsymbol{\theta}$ si riduca al calcolo della proiezione ortogonale di $\boldsymbol{Y}$ su $E = (\boldsymbol{A}\mathbb{R}^2)$, ovvero al calcolo del vettore che minimizzi il quadrato della distanza tra $\boldsymbol{Y}$ e $\boldsymbol{A\theta}$ (si veda l'*Osservazione (11.3)*): la soluzione è la stessa del metodo MQ o degli stimatori BLUE.

I risultati precedenti suggeriscono di introdurre alcune nuove quantità.

DEFINIZIONE 11.4. Le variabili

$$\hat{y}_i = \hat{\theta}_0 + \hat{\theta}_1 x_i, \quad i = 1, \ldots, n, \tag{11.23}$$

sono dette *valori stimati*. Le variabili

$$r_i = y_i - \hat{y}_i, \quad i = 1, \ldots, n, \tag{11.24}$$

sono dette *residui*.

NOTA 11.4.

Dalla *Definizione (11.4)* scende che

$$S^2 = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \frac{1}{n-2} \sum_{i=1}^{n} r_i^2.$$

Dunque, usando una notazione ovvia,

$$(n-2)\, S^2 = \left\| \boldsymbol{Y} - \hat{\boldsymbol{Y}} \right\|^2 = \left\| \boldsymbol{Y} - \boldsymbol{A}\hat{\boldsymbol{\theta}} \right\|^2 = \| \boldsymbol{Y} - \boldsymbol{A}\boldsymbol{U}\boldsymbol{Y} \|^2 ,$$

come deve essere in base al *Teorema di Gauss–Markov (II)*.

I residui godono di alcune proprietà interessanti.

PROPOSIZIONE 11.3. *Per il modello Normale condizionale valgono i seguenti risultati:*

$$\sum_{i=1}^{n} r_i = \langle \boldsymbol{r}, \boldsymbol{1}_n \rangle = 0, \tag{11.25a}$$

$$\sum_{i=1}^{n} r_i \hat{y}_i = \langle \boldsymbol{r}, \hat{\boldsymbol{y}} \rangle = 0. \tag{11.25b}$$

Inoltre il vettore dei residui $\boldsymbol{r}$ e quello dei valori stimati $\hat{\boldsymbol{y}}$ sono indipendenti.

DIM. L'*equazione (11.25a)* segue dalla definizione di residui e dall'*equazione (11.7a)*, mentre per l'*equazione (11.25b)* basta osservare che

$$\sum_{i=1}^{n} r_i \hat{y}_i = \hat{\theta}_0 \sum_{i=1}^{n} r_i + \hat{\theta}_1 \sum_{i=1}^{n} r_i x_i,$$

e la prima somma a destra è nulla grazie all'*equazione (11.25a)*, mentre la seconda lo è per via dell'*equazione (11.7b)*. Infine, l'indipendenza segue dal *Teorema di Cochran*. □

NOTA 11.5.

Indicando con $\boldsymbol{P}$ la proiezione ortogonale su E e con $\boldsymbol{P}^{\perp}$ quella su $E^{\perp}$ si ha

$$\boldsymbol{r} = \boldsymbol{y} - \hat{\boldsymbol{y}} = (\boldsymbol{I}_n - \boldsymbol{P})\, \boldsymbol{y} = \boldsymbol{P}^{\perp} \boldsymbol{y}. \tag{11.26}$$

Dunque $\boldsymbol{r}$ è ortogonale a E, e corrisponde alla proiezione ortogonale di $\boldsymbol{y}$ su $E^{\perp}$. In particolare, $\boldsymbol{r}$ è ortogonale ai vettori $\boldsymbol{1}_n, \boldsymbol{x}$ che formano le colonne della matrice $\boldsymbol{A}$; ciò fornisce sia una dimostrazione alternativa della *Proposizione (11.3)* sia una interpretazione geometrica dei risultati esposti. Inoltre, dall'*equazione (11.25a)* si ricava che

$$\frac{1}{n} \sum_{i=1}^{n} \hat{y}_i = \frac{1}{n} \sum_{i=1}^{n} y_i = \overline{y}. \tag{11.27}$$

Nelle ipotesi date, e grazie al *Teorema di Gauss–Markov (II)* e al *Teorema di Cochran*, si ottiene quanto segue.

TEOREMA 11.3. *Nel modello Normale condizionale si ha:*

$$\hat{\theta}_0 \sim \mathtt{N}\left(\theta_0, \frac{\sigma^2}{nS_{xx}} \sum_{i=1}^{n} x_i^2\right) \sim \mathtt{N}\left(\theta_0, \sigma^2 \left(\frac{1}{n} + \frac{\overline{x}^2}{S_{xx}}\right)\right), \tag{11.28a}$$

$$\hat{\theta}_1 \sim \mathtt{N}\left(\theta_1, \frac{\sigma^2}{S_{xx}}\right), \tag{11.28b}$$

$$\mathbf{C}\left(\hat{\theta}_0, \hat{\theta}_1\right) = -\frac{\sigma^2}{S_{xx}}\,\overline{x}. \tag{11.28c}$$

Inoltre, $\left(\hat{\theta}_0, \hat{\theta}_1\right)$ *e* S^2 *sono indipendenti, e*

$$\frac{n-2}{\sigma^2}\, S^2 \sim \chi^2(n-2)\,. \tag{11.28d}$$

Grazie al *Teorema (11.3)* è possibile fare inferenze circa i parametri stimati. Infatti le leggi delle statistiche che seguono sono di derivazione immediata:

$$T_{\theta_0} = \frac{\hat{\theta}_0 - \theta_0}{S\sqrt{\sum_{i=1}^{n} x_i^2/(nS_{xx})}} \sim \mathtt{t}(n-2)\,, \tag{11.29a}$$

$$T_{\theta_1} = \frac{\hat{\theta}_1 - \theta_1}{S/\sqrt{S_{xx}}} \sim \mathtt{t}(n-2)\,. \tag{11.29b}$$

Dunque è ora facile procedere alla costruzione di IC e alla verifica di ipotesi. Solitamente c'è maggiore interesse nei confronti del parametro θ_1. Infatti, θ_0 non rappresenta altro che il valore atteso di Y quando $x = 0$, e che tale quantità sia o no utile dipende dal problema in esame (addirittura $x = 0$ potrebbe non essere nemmeno un valore ammissibile per il predittore). Al contrario, θ_1 rappresenta il tasso di variazione del valore atteso di Y quale funzione di x: dunque, esso contiene informazioni circa la relazione lineare esistente tra la risposta e il predittore su tutto l'insieme dei valori ammissibili di quest'ultimo.

Di particolare interesse è l'analisi del caso $\theta_1 = 0$: in tale situazione si avrebbe $Y \sim \mathtt{N}\big(\theta_0, \sigma^2\big)$ e $\mathbf{E}(Y) = \theta_0$, ovvero la risposta non dipenderebbe (linearmente) dal predittore. Nella verifica dell'ipotesi

$$\mathcal{H}_0 \colon \theta_1 = 0 \quad \text{contro} \quad \mathcal{H}_1 \colon \theta_1 \neq 0 \tag{11.30}$$

si utilizza l'*equazione (11.29b)* e si rifiuta $\mathcal{H}_0$, al livello $\alpha \in (0,1)$, se

$$\left|\frac{\hat{\theta}_1}{S/\sqrt{S_{xx}}}\right| > t_{n-2,1-\alpha/2} \quad \text{o, equivalentemente, se} \quad \frac{\hat{\theta}_1^2}{S^2/S_{xx}} > f_{1,n-2,1-\alpha}. \tag{11.31}$$

Tradizionalmente i risultati sono presentati come mostrato in *Tabella (11.2)*, del tutto simile alla *Tabella (10.1)* relativa al test F-ANOVA ad una via.

Causa	**Somma dei Quadrati**	**Gradi di Libertà**	**Media**	***F***
Regressione	$\text{SQR} = S_{xy}^2/S_{xx}$	1	SQR	$F = \frac{\text{SQR}}{\text{SRQ}/(n-2)}$
Residui	$\text{SRQ} = \sum_{i=1}^{n} r_i^2$	$n-2$	$\text{SRQ}/(n-2)$	

Tabella 11.2: presentazione standard dei risultati della verifica dell'ipotesi $\mathcal{H}_0 \colon \theta_1 = 0$.

OSSERVAZIONE 11.6 (Partizionamento delle fluttazioni)**.**
In effetti, la verifica dell'ipotesi $\mathcal{H}_0 : \theta_1 = 0$ è simile a quella dell'ipotesi (ANOVA) nulla che tutti i trattamenti siano uguali: nel caso ANOVA tale assunzione comporta che la risposta non sia legata in modo specifico ad alcun trattamento, mentre nel caso della regressione lineare semplice si afferma che non esiste una *relazione lineare* tra la risposta Y e i predittori x. Tuttavia le analogie non si fermano qui: si consideri la relazione

$$\sum_{i=1}^{n} (y_i - \overline{y})^2 = \sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2 + \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \text{SQR} + \text{SRQ}, \tag{11.32}$$

simile all'*equazione (10.16)*. Il termine di sinistra "misura" globalmente le deviazioni presenti nel modello, mentre *SQR* tiene conto solamente di quelle dovute alla regressione e *SRQ* di quelle dovute ai residui.

Una quantità che è spesso utile considerare nelle applicazioni è la seguente.

DEFINIZIONE 11.5 (Coefficiente di determinazione)**.** La variabile data da

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2} \tag{11.33}$$

è detta *coefficiente di determinazione*.

Tradizionalmente si intende che R^2 rappresenti *la proporzione di variazione di Y che viene spiegata dal modello*. Dall'*equazione (11.32)* scende che $0 \leq R^2 \leq 1$: valori di R^2 vicini a 1 indicano un buon modello (in effetti, se tutti i punti (x_i, y_i) giacessero sulla retta di regressione si avrebbe $y_i = \hat{y}_i$, e dunque sarebbe $R^2 = 1$); valori di R^2 vicini a 0 possono essere spiegati da valori elevati di σ^2, oppure suggeriscono che il modello non spieghi completamente il comportamento della risposta Y.

NOTA 11.6.
Il vincolo $0 \leq R^2 \leq 1$ può essere derivato anche dal fatto che R^2 rappresenta il quadrato del coefficiente di correlazione campionario delle n coppie $(x_1, y_1), \ldots, (x_n, y_n)$, o delle n coppie $(y_1, \hat{y}_1), \ldots, (y_n, \hat{y}_n)$. Infatti vale

$$R^2 = \left(\frac{S_{xy}}{\sqrt{(S_{xx} S_{yy})}} \right)^2, \tag{11.34}$$

che è uno stimatore del quadrato del coefficiente di correlazione lineare $\boldsymbol{\rho_{x,y}}$.

Tornando all'*Esempio (11.1)* relativo allo studio dell'acqua alta a Venezia, nelle ipotesi fatte si ottiene che la retta di regressione del modello Normale condizionale è la stessa di quella del metodo MQ (operando una regressione di y su x) o di quella BLUE, mostrata in *Figura (11.1)*. Ora, tuttavia, è possibile fare considerazioni di natura statistica. Tipicamente i risultati dell'analisi sono raccolti in una tabella come quella che segue.

Parametro	**Stima**	**Errore standard**	$\boldsymbol{T}$	$\boldsymbol{p}$**-value**
$\boldsymbol{\theta_0}$	-969.2669	360.3277	-2.6900	0.0098
$\boldsymbol{\theta_1}$	0.5566	0.1843	3.0210	0.0040
$\boldsymbol{\sigma}$	18.8021	$\boldsymbol{R^2} = 0.1598$		

Tabella 11.3: stime fornite dal modello Normale condizionale per i dati dell'*Esempio (11.1)*.

Le stime dei parametri θ_0 e θ_1 sono calcolate tramite le *equazioni (11.20a)–(11.20b)*, la stima di σ è data dall'*equazione (11.22)*. L'*errore standard* rappresenta la radice quadrata della varianza degli stimatori calcolata tramite le *equazioni (11.28a)–(11.28b)*: qui il valore di σ^2 è stimato con S^2. La colonna "T" mostra, rispettivamente, i valori t_{θ_0} e t_{θ_1} delle statistiche T_{θ_0} e T_{θ_1} calcolati tramite le *equazioni (11.29a)–(11.29b)*: qui è, rispettivamente,

$\theta_0 = 0$ e $\theta_1 = 0$, che corrispondono alle ipotesi che solitamente si vogliono verificare. La colonna "p-value" contiene, rispettivamente, i valori delle probabilità $\mathbf{P}\{|T| > t_{\theta_0}\}$ e $\mathbf{P}\{|T| > t_{\theta_1}\}$, dove $T \sim \mathtt{t}(n-2)$, ovvero $T \sim \mathtt{t}(48)$ nel caso in esame; in Letteratura tali probabilità sono dette *p-value*, e rappresentano le quantità che occorre conoscere volendo effettuare la verifica delle ipotesi $\mathcal{H}_0\colon \theta_0 = 0$ o $\mathcal{H}_0\colon \theta_1 = 0$: affinché tali ipotesi siano respinte occorre che i rispettivi p-value siano minori del livello $\alpha \in (0,1)$ prescelto (si veda l'*Appendice H*). Qui l'ipotesi $\mathcal{H}_0\colon \theta_0 = 0$ può essere respinta a tutti i livelli $\alpha > 0.0098$, cioè quelli comunemente adottati nella pratica statistica; la stessa conclusione può essere tratta considerando l'ipotesi $\mathcal{H}_0\colon \theta_1 = 0$. Dunque, nell'ipotesi che il modello adottato sia corretto, statisticamente si accetta il fatto che sia $\theta_1 > 0$, da cui si deduce che i massimi di acqua alta a Venezia siano andati aumentando nel periodo considerato. In effetti, calcolando le quantità mostrate in *Tabella (11.2)* si ottiene:

Causa	Somma dei Quadrati	Gradi di Libertà	Media	F
Regressione	3226.2777	1	3226.2777	9.1263
Residui	16968.8424	48	353.5176	

Perciò, essendo $f_{1,48,0.99} \approx 7.19$, in base all'*equazione (11.31)* l'ipotesi $\mathcal{H}_0\colon \theta_1 = 0$ è da rifiutare (almeno) al livello $\alpha = 0.01$.

In realtà, nei problemi di regressione lineare la questione più delicata è la verifica che i dati disponibili soddisfino alle ipotesi che sono alla base del modello. In particolare, non è detto a priori che la risposta sia una funzione affine del predittore (o dei predittori), né che la varianza della perturbazione aletoria sia la stessa per tutte le osservazioni (e in tal caso si parla di *eteroschedasticità*). Uno strumento importante per controllare la validità del modello di regressione è dato dall'*analisi dei residui*. Grazie alla *Proposizione (11.3)* sappiamo che, se il modello è corretto, il vettore dei residui $\boldsymbol{r}$ e quello dei valori stimati $\hat{\boldsymbol{y}}$ sono indipendenti. Dunque, intuitivamente, il grafico dei punti $(\hat{\boldsymbol{y}}, \boldsymbol{r})$ dovrebbe chiarire *(i)* se esiste una dipendenza di una coordinata dall'altra e *(ii)* se l'omoschedasticità sia un'ipotesi ragionevole.

La *Figura (11.2)* mostra l'analisi dei residui per i dati dell'*Esempio (11.1)* relativo allo studio dell'acqua alta a Venezia. Apparentemente i punti si dispongono uniformemente attorno alla retta orizzontale, e non si individua alcuna tendenza particolare ad assumere andamenti anomali (ad esempio curvature); inoltre, l'entità delle fluttuazioni attorno alla retta orizzontale non sembra dipendere dai valori stimati. Invece, può risultare sospetto il basso valore di $R^2 \approx 16\%$, che sembrerebbe indicare una scarsa propensione del modello a "spiegare" le fluttuazioni della risposta: ciò, tuttavia, può essere semplicemente originato dall'elevato valore di σ^2, dovuto alla intrinseca variabilità dei dati. In conclusione, il modello di regressione lineare semplice utilizzato sembra fornire una descrizione accettabile del fenomeno in esame.

11.1.4 Stima e predizione nel modello Normale condizionale

Uno degli obiettivi principali dell'uso dei modelli di regressione lineare consiste nella previsione dei valori della risposta per assegnati valori del predittore. Nel seguito assumeremo che siano disponibili n osservazioni $(x_1, Y_1), \ldots, (x_n, Y_n)$ soddisfacenti alle ipotesi del modello Normale condizionale, e che $\hat{\theta}_0$, $\hat{\theta}_1$ e S^2 rappresentino le stime dei parametri del modello basate su tali dati.

Sia ora x_0 un valore assegnato del predittore, al quale è naturale associare una risposta Y_0. Due sono gli obiettivi che tipicamente si vogliono raggiungere:

1. stimare la media della popolazione dalla quale l'osservazione Y_0 è estratta;
2. fornire una predizione del valore di Y_0.

Come vedremo, i due problemi hanno solo formalmente la stessa soluzione. Prima di affrontare la questione è tuttavia opportuno premettere alcune considerazioni: a priori nulla garantisce che x_0 sia un valore ammissibile del predittore, né che il modello Normale condizionale sia ancora valido in corrispondenza di x_0; anzi, queste sono ipotesi aggiuntive di cui l'analista deve tener conto prima di procedere al calcolo di stime e predizioni.

Consideriamo dapprima la stima di $\mathbf{E}(Y_0)$. Per le ipotesi fatte, è ovvio prendere quale stimatore

$$\hat{y}_0 = \hat{\theta}_0 + \hat{\theta}_1 x_0, \tag{11.35}$$

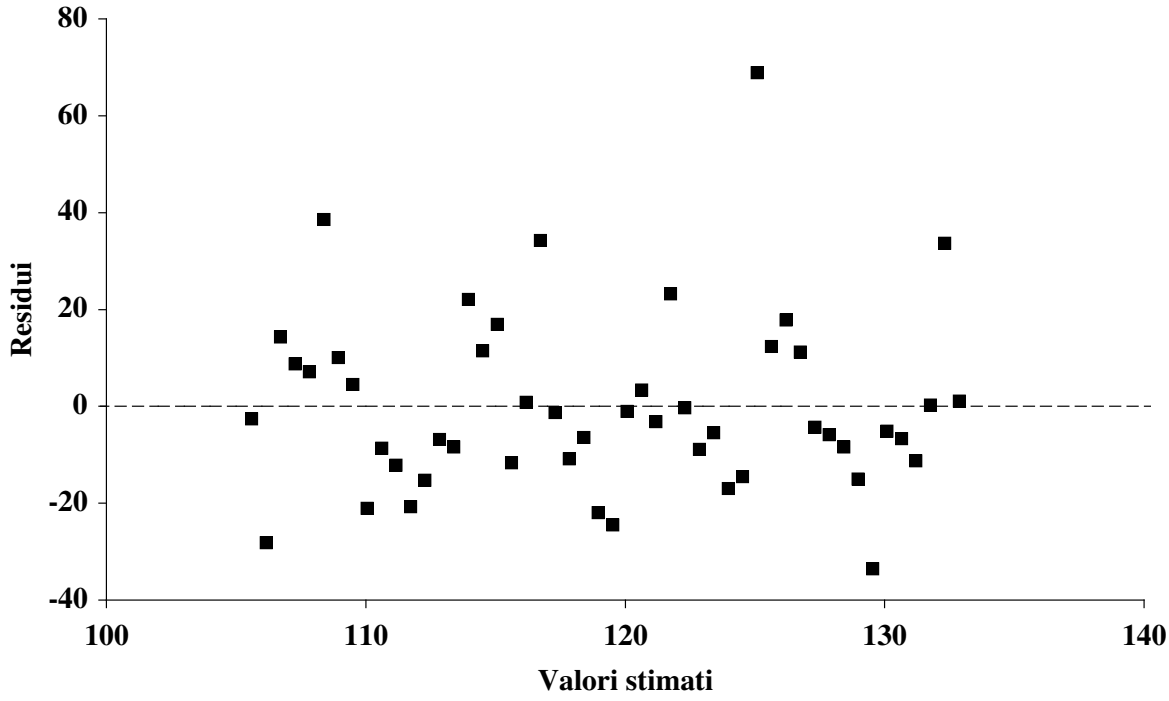

Figura 11.2: analisi dei residui per i dati dell'*Esempio (11.1).*

che risulta essere non distorto e di legge

$$\hat{y}_0 \sim \mathtt{N}\left(\theta_0 + \theta_1 x_0, \sigma^2 \left(\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}\right)\right). \tag{11.36}$$

Poiché in base al *Teorema (11.3)* S^2 e $\left(\hat{\theta}_0, \hat{\theta}_1\right)$ sono indipendenti, anche $\hat{y}_0$ e S^2 lo sono; dunque

$$T_0 = \frac{\hat{y}_0 - (\theta_0 + \theta_1 x_0)}{S\sqrt{\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}}} \sim \mathtt{t}(n-2)\,, \tag{11.37}$$

con cui è possibile costruire IC e procedere alla verifica di ipotesi. In particolare, si noti che la lunghezza dell'IC basato sulla statistica pivotale T_0 è proporzionale a $\frac{(x_0 - \overline{x})^2}{S_{xx}}$; dunque, a parità di livello α, tale lunghezza è minima quando $x_0 = \overline{x}$, un risultato che non sorprende: infatti è ragionevole pensare che la stima sia più accurata nei pressi del "baricentro" $\overline{x}$ dei dati disponibili.

Invece, sia ora Y_0 una nuova osservazione presa in corrispondenza di x_0, avente legge $\mathtt{N}\big(\theta_0 + \theta_1 x_0, \sigma^2\big)$ e indipendente dai dati $(x_1, Y_1), \ldots, (x_n, Y_n)$; dunque Y_0 è indipendente da $\hat{\theta}_0$, $\hat{\theta}_1$ e S^2. Ne segue che

$$Y_0 - (\hat{\theta}_0 + \hat{\theta}_1 x_0) \sim \mathtt{N}\left(0, \sigma^2 \left(1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}\right)\right), \tag{11.38}$$

e inoltre

$$T_{Y_0} = \frac{Y_0 - (\hat{\theta}_0 + \hat{\theta}_1 x_0)}{S\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}}} \sim \mathtt{t}(n-2)\,, \tag{11.39}$$

con cui è possibile costruire IC e procedere alla verifica di ipotesi. Si noti che la lunghezza dell'IC basato sulla statistica pivotale T_{Y_0} è maggiore di quella dell'IC basato su T_0: intuitivamente ciò si spiega con il fatto che $\mathbf{E}(Y_0)$ è un parametro (costante), mentre Y_0 è una v.a., e dunque esprime una "variabilità" maggiore.

Una questione ulteriore è rappresentata dal seguente problema: si supponga di voler stimare $\mathbf{E}(Y_0)$ in corrispondenza di diversi valori del predittore, diciamo $\{x_{0,i}\}$, $i = 1, \ldots, m$. Nelle applicazioni è spesso di interesse la costruzione di un *unico* IC di livello $\alpha \in (0, 1)$ per i valori stimati, ossia tale che valga, con probabilità almeno pari a $1 - \alpha$,

$$\hat{y}_{0,i} - t_{n-2,\gamma}\, S \sqrt{\frac{1}{n} + \frac{(x_{0,i} - \overline{x})^2}{S_{xx}}} < \theta_0 + \theta_1 x_{0,i} < \hat{y}_{0,i} + t_{n-2,\gamma}\, S \sqrt{\frac{1}{n} + \frac{(x_{0,i} - \overline{x})^2}{S_{xx}}}$$

contemporaneamente per tutti gli indici $i = 1, \ldots, m$. Si tratta dunque di calcolare un livello γ opportuno. Un problema simile è già stato affrontato e risolto nella *Sezione 10.1.3*, e la stessa soluzione può essere adottata nel presente contesto. Infatti, sfruttando la *Disuguaglianza di Bonferroni*, si ottiene $\gamma = 1 - \frac{\alpha}{2m}$, mentre applicando il *Metodo S* di *Scheffé* si ha che $t_{n-2,\gamma}$ deve essere sostituito da $M = \sqrt{2\, f_{2,n-2,1-\alpha}}$; inoltre valgono le considerazioni già espresse in quella sede. Si noti tuttavia che applicando il *Metodo S* si ha

$$\hat{\theta}_0 + \hat{\theta}_1 x - M\, S \sqrt{\frac{1}{n} + \frac{(x - \overline{x})^2}{S_{xx}}} < \theta_0 + \theta_1 x < \hat{\theta}_0 + \hat{\theta}_1 x + M\, S \sqrt{\frac{1}{n} + \frac{(x - \overline{x})^2}{S_{xx}}},$$

ovvero un IC valido simultaneamente *per ogni* x, che dunque costituisce una *banda di confidenza*. Come già accennato, il prezzo che si deve pagare a tanta generalità risiede nell'ampiezza, spesso eccessiva e poco informativa, di tale banda.

Tornando all'*Esempio (11.1)* relativo allo studio dell'acqua alta a Venezia, la *Figura (11.3)* mostra due diverse bande di confidenza di *Scheffé*: quella interna ha livello $\alpha = 0.05$ e quella esterna ha livello $\alpha = 0.01$; si noti come l'ampiezza di entrambe sia minima quando $x = \overline{x} = 1955.5$. Infine, a titolo d'esempio, volendo fare una predizione del valore Y_0 di Y per l'anno $x_0 = 1981$, è naturale usare quale stimatore la v.a. $\hat{Y}_0 = \hat{\theta}_0 + \hat{\theta}_1 x_0$, che fornisce $\hat{Y}_0 \approx 133.43$; grazie alle *equazioni (11.38)–(11.39)* si ottiene poi un IC di livello $\alpha = 0.05$ dato da $(94.10, 172.77)$. Purtroppo l'ampiezza di tale intervallo lo rende poco utile, come del resto era prevedibile dato l'elevato valore di S^2 (per confronto, la corrispondente banda di confidenza di *Scheffé* al livello $\alpha = 0.05$ fornisce l'intervallo $(119.80, 147.07)$); inoltre, a priori non è detto che la presunta validità del modello Normale condizionale sia estendibile anche al 1981. La *Figura (11.3)* mostra la stima di Y_0 unitamente all'IC calcolato; per confronto, il valore massimo di acqua alta effettivamente osservato nel 1981 fu pari a 138 cm, anch'esso indicato in Figura.

11.2 Regressione lineare multipla

Come già accennato, la *regressione lineare multipla* costituisce un valido strumento per lo studio di fenomeni in cui la risposta dipenda da più di un predittore: in agricoltura, ad esempio, è fondamentale per l'analisi del raccolto medio quando siano impiegati diversi tipi di fertilizzante; ancora, essa è impiegata con successo nello studio del consumo regionale di carburante in funzione di fattori quali il numero di abitanti, il reddito pro capite, le tasse automobilistiche e lo sviluppo del sistema autostradale. Il modello di base risulta essere come segue.

DEFINIZIONE 11.6 (Regressione lineare multipla)**.** Nella regressione lineare multipla il modello è:

$$Y_i = \theta_0 + \theta_1 x_i^{(1)} + \cdots + \theta_p x_i^{(p)} + \sigma\epsilon_i, \quad i = 1, \ldots, n, \tag{11.40}$$

dove le variabili Y_i sono v.a. osservate dette *risposta*, le variabili $x_i^{(j)}$, $j = 1, \ldots, p$, $p < n - 1$, sono variabili osservate dette *predittori*, e le variabili $\epsilon_i \sim \mathtt{N}(0, 1)$ sono v.a. i.i.d. dette *fluttuazioni casuali* (o *errori*). I parametri reali θ_j, $j = 0, \ldots, p$, sono incogniti ma costanti, così come il parametro $\sigma > 0$.

Prima di procedere è utile riscrivere il modello dato usando una notazione matriciale. Siano

$$\boldsymbol{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \boldsymbol{A} = \begin{pmatrix} 1 & x_1^{(1)} & \ldots & x_1^{(p)} \\ \vdots & \vdots & & \\ 1 & x_n^{(1)} & \ldots & x_n^{(p)} \end{pmatrix}, \quad \boldsymbol{\theta} = \begin{pmatrix} \theta_0 \\ \vdots \\ \theta_p \end{pmatrix}, \quad \boldsymbol{Z} = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}.$$

Dunque è, indicando con $\boldsymbol{x}^{(j)}$ i vettori colonna della matrice $\boldsymbol{A}$ e usando una notazione intuitiva,

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{\theta} + \sigma\boldsymbol{Z} = \theta_0 \boldsymbol{x}^{(0)} + \theta_1 \boldsymbol{x}^{(1)} + \cdots + \theta_p \boldsymbol{x}^{(p)} + \sigma\boldsymbol{Z}. \tag{11.41}$$

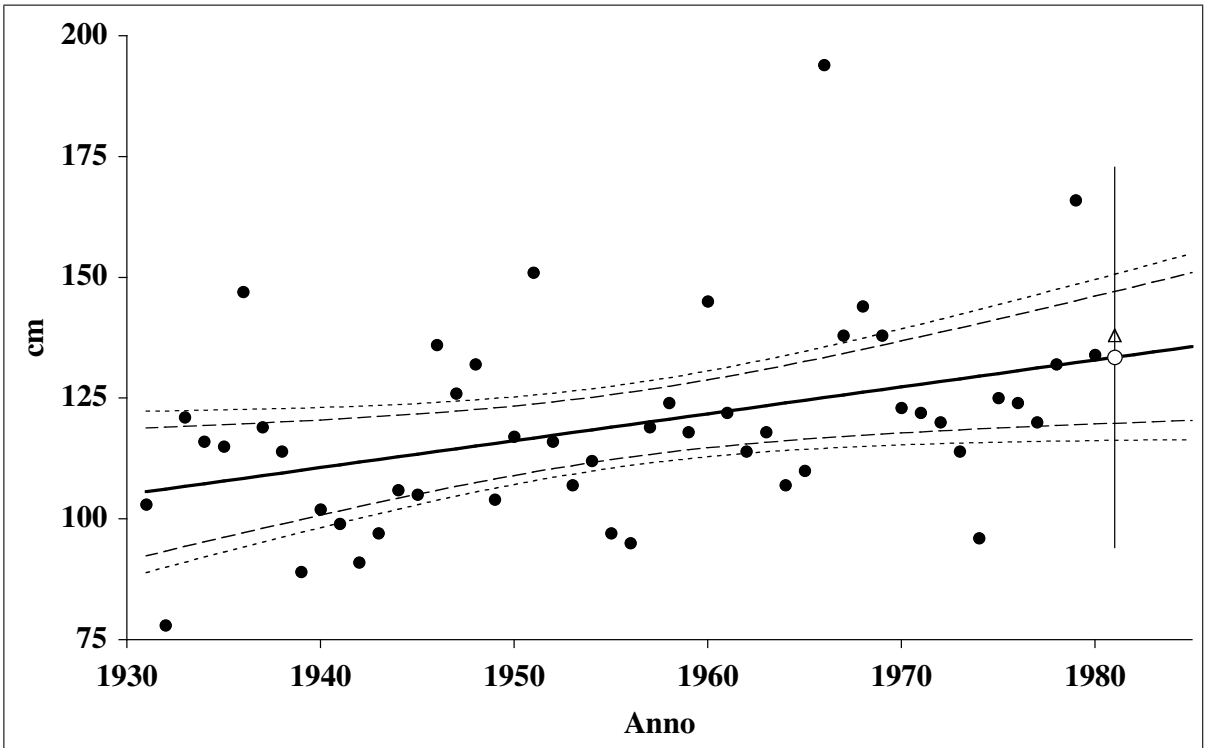

Figura 11.3: bande di confidenza di *Scheffé* per i dati dell'*Esempio (11.1)*; le linee a *tratteggio lungo* corrispondono ad un livello $\alpha = 0.05$, quelle a *tratteggio breve* ad un livello $\alpha = 0.01$. Il *cerchio vuoto* indica la stima per l'anno 1981; la banda di errore mostrata corrisponde all'IC di livello $\alpha = 0.05$. Il *triangolo* indica il valore massimo di acqua alta effettivamente osservato nel 1981.

Il vettore $\boldsymbol{x}^{(0)} = \mathbf{1}_n$ è detto *fattore costante*. Lo spazio dei parametri è

$$\Theta = \mathbb{R}^{p+1} \times \mathbb{R}_+ = \{\theta_0, \ldots, \theta_p \in \mathbb{R}, \sigma \in \mathbb{R}_+\}\,.$$

La legge $\mathbf{P}_\Theta$ di $\boldsymbol{Y}$ è la legge immagine di $\boldsymbol{Z}$ mediante l'applicazione affine $\boldsymbol{z} \mapsto \boldsymbol{A\theta} + \sigma\boldsymbol{z}$, con $\boldsymbol{z} \in \mathbb{R}^n$; dunque $\boldsymbol{Y} \sim \mathtt{N}\big(\boldsymbol{A\theta}, \sigma^2\, \boldsymbol{I}_n\big)$. Nel seguito supporremo che $\boldsymbol{A}$ sia iniettiva, ovvero che i vettori $\boldsymbol{x}^{(0)}, \ldots, \boldsymbol{x}^{(p)}$ siano linearmente indipendenti; d'altro canto, se così non fosse sarebbe possibile esprimere uno di tali vettori quale combinazione lineare degli altri, e dunque sarebbe impossibile distinguere l'effetto del predittore corrispondente da quello degli altri predittori.

NOTA 11.7.

Il modello di regressione lineare semplice è un caso particolare di quello di regressione lineare multipla (basta prendere $p = 1$), e (quasi) tutte le considerazioni espresse nel caso della regressione lineare semplice valgono anche per quella multipla.

Il Teorema che segue è fondamentale.

TEOREMA 11.4 (Gauss–Markov (III))**.** *Nelle ipotesi date,*

$$\boldsymbol{UY} = (\boldsymbol{A}^\intercal \boldsymbol{A})^{-1}\, \boldsymbol{A}^\intercal \boldsymbol{Y} \tag{11.42}$$

è uno stimatore non distorto di $\boldsymbol{\theta}$, e

$$\|\boldsymbol{AUY} - \boldsymbol{Y}\|^2 \tag{11.43}$$

è uno stimatore non distorto di $(n - p - 1)\sigma^2$. Tutti gli stimatori sono ottimali nella classe degli stimatori non distorti di quadrato sommabile dei rispettivi parametri.

DIM. È sufficiente generalizzare la dimostrazione del *Teorema (11.1)* e del *Teorema (11.2).* □

OSSERVAZIONE 11.7.

Alla luce delle considerazioni illustrate nell'*Appendice G* e nell'*Osservazione (11.3)*, il risultato del *Teorema (11.4)* può essere visto sotto un'altra prospettiva. Si scomponga la matrice $\boldsymbol{A}$ nei suoi vettori colonna linearmente indipendenti $\boldsymbol{x}^{(0)}, \dots, \boldsymbol{x}^{(p)}$ come mostrato nell'*equazione (11.41).* Se $\boldsymbol{\theta}$ varia in $\mathbb{R}^{p+1}$, allora il vettore

$$\boldsymbol{v} = \boldsymbol{A\theta} = \theta_0 \boldsymbol{x}^{(0)} + \theta_1 \boldsymbol{x}^{(1)} + \cdots + \theta_p \boldsymbol{x}^{(p)}$$

varia in tutto il sottospazio $E = (\boldsymbol{A}\mathbb{R}^{p+1})$ generato dai vettori $\boldsymbol{x}^{(0)}, \dots, \boldsymbol{x}^{(p)}$. Ancora una volta, volendo stimare $\boldsymbol{\theta}$, si può pensare di utilizzare quel vettore che renda minima la distanza tra $\boldsymbol{v}$ e $\boldsymbol{Y}$. Dalla *Proposizione (G.4)* scende che il vettore di E avente distanza minima da $\boldsymbol{Y}$ è dato dalla proiezione ortogonale di $\boldsymbol{Y}$ su E; d'altro canto, dalla *Proposizione (G.3)* sappiamo che tale minimo si verifica in corrispondenza di $\hat{\boldsymbol{\theta}} = \boldsymbol{UY}$. Dunque, il calcolo di $\hat{\boldsymbol{\theta}}$ si riduce a quello della proiezione ortogonale di $\boldsymbol{Y}$ su E, che coincide con il risultato del *Teorema di Gauss–Markov (III)*. Volendo procedere al calcolo diretto di $\hat{\boldsymbol{\theta}}$, sarebbe necessario minimizzare la funzione quadratica

$$\varphi(\boldsymbol{\theta}) = \|\boldsymbol{Y} - \boldsymbol{A\theta}\|^2 = \|\boldsymbol{Y}\|^2 + \langle \boldsymbol{A}^\intercal \boldsymbol{A\theta}, \boldsymbol{\theta} \rangle - 2 \langle \boldsymbol{A}^\intercal \boldsymbol{Y}, \boldsymbol{\theta} \rangle ,$$

che rappresenta il quadrato della distanza tra $\boldsymbol{Y}$ e $\boldsymbol{A\theta}$. La soluzione al problema si può ottenere calcolando i punti critici del gradiente. Ora, a meno di costanti, φ è la somma della funzione lineare $-2 \langle \boldsymbol{A}^\intercal \boldsymbol{Y}, \boldsymbol{\theta} \rangle$ e della funzione quadratica $\langle \boldsymbol{A}^\intercal \boldsymbol{A\theta}, \boldsymbol{\theta} \rangle$; dunque il suo gradiente vale

$$\operatorname{grad}(\varphi(\boldsymbol{\theta})) = 2\boldsymbol{A}^\intercal \boldsymbol{A\theta} - 2\boldsymbol{A}^\intercal \boldsymbol{Y}.$$

L'equazione $\operatorname{grad}(\varphi(\boldsymbol{\theta})) = 0$ si risolve facilmente, in quanto dalla *Proposizione (G.3)* sappiamo che la matrice $\boldsymbol{A}^\intercal \boldsymbol{A}$ è quadrata, di dimensione $(p+1) \times (p+1)$, ed è invertibile. La soluzione vale

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{A}^\intercal \boldsymbol{A})^{-1} \boldsymbol{A}^\intercal \boldsymbol{Y},$$

cioè la stessa del *Teorema di Gauss–Markov (III)*.

Il Teorema che segue riassume le proprietà degli stimatori individuati nel caso della regressione lineare multipla. Ricordiamo che il vettore $\boldsymbol{Y}$ ha legge $\mathtt{N}\big(\boldsymbol{A\theta}, \sigma^2 \boldsymbol{I}_n\big)$ e che, se $\boldsymbol{C_Y}$ rappresenta la matrice di varianza-covarianza di $\boldsymbol{Y}$, allora $\hat{\boldsymbol{\theta}} = \boldsymbol{UY}$ ha matrice di varianza-covarianza $\boldsymbol{C}_{\hat{\boldsymbol{\theta}}} = \boldsymbol{U}\boldsymbol{C_Y}\boldsymbol{U}^\intercal$: dunque, essendo $\boldsymbol{U} = (\boldsymbol{A}^\intercal \boldsymbol{A})^{-1} \boldsymbol{A}^\intercal$, se ne deduce che

$$\begin{aligned} \boldsymbol{C}_{\hat{\boldsymbol{\theta}}} &= (\boldsymbol{A}^\intercal \boldsymbol{A})^{-1} \boldsymbol{A}^\intercal \sigma^2 \boldsymbol{I}_n \left((\boldsymbol{A}^\intercal \boldsymbol{A})^{-1} \boldsymbol{A}^\intercal \right)^\intercal \\ &= \sigma^2 (\boldsymbol{A}^\intercal \boldsymbol{A})^{-1} \boldsymbol{A}^\intercal \boldsymbol{A} (\boldsymbol{A}^\intercal \boldsymbol{A})^{-1} \\ &= \sigma^2 (\boldsymbol{A}^\intercal \boldsymbol{A})^{-1} , \end{aligned} \tag{11.44}$$

dove si è usato il fatto che $(\boldsymbol{A}^\intercal \boldsymbol{A})^{-1}$ è una matrice simmetrica. Nel seguito $c_{\hat{\boldsymbol{\theta}}}(i,j)$ indicherà l'elemento della matrice $(\boldsymbol{A}^\intercal \boldsymbol{A})^{-1}$ di posto (i,j).

TEOREMA 11.5. *Nelle ipotesi date si ha:*

$$\hat{\boldsymbol{\theta}} \sim \mathtt{N}\Big(\boldsymbol{\theta}, \sigma^2 (\boldsymbol{A}^\intercal \boldsymbol{A})^{-1}\Big) ; \tag{11.45a}$$

la componente i-esima del vettore $\hat{\boldsymbol{\theta}}$ ha legge

$$\hat{\theta}_i \sim \mathtt{N}\big(\theta_i, \sigma^2 c_{\hat{\boldsymbol{\theta}}}(i,i)\big) , \tag{11.45b}$$

e la covarianza tra due stimatori $\hat{\theta}_i, \hat{\theta}_j$ vale

$$\mathbf{C}\Big(\hat{\theta}_i, \hat{\theta}_j\Big) = \sigma^2 c_{\hat{\boldsymbol{\theta}}}(i,j). \tag{11.45c}$$

Inoltre, $\hat{\boldsymbol{\theta}}$ e $S^2 = \frac{1}{n-p-1} \|\boldsymbol{r}\|^2$ (lo stimatore di σ^2) sono indipendenti, e

$$\frac{n-p-1}{\sigma^2} S^2 \sim \chi^2(n-p-1) . \tag{11.45d}$$

Grazie al *Teorema (11.5)* è possibile fare inferenze circa i parametri stimati. Infatti le leggi delle statistiche che seguono sono di derivazione immediata:

$$T_{\theta_i} = \frac{\hat\theta_i - \theta_i}{S\sqrt{c_{\hat{\boldsymbol\theta}}(i,i)}} \sim \mathtt{t}(n-p-1)\,, \quad i = 0, \dots, p. \tag{11.46}$$

Dunque è ora facile procedere alla costruzione di IC e alla verifica di ipotesi. In particolare,

$$I_i = \left(\hat\theta_i - S\sqrt{c_{\hat{\boldsymbol\theta}}(i,i)}\, t_{n-p-1,1-\alpha/2}, \hat\theta_i + S\sqrt{c_{\hat{\boldsymbol\theta}}(i,i)}\, t_{n-p-1,1-\alpha/2}\right), \quad i = 0, \dots, p, \tag{11.47}$$

è un IC per θ_i di livello $\alpha \in (0,1)$. Inoltre,

$$C_i = \left\{ \left| \frac{\hat\theta_i - t}{S\sqrt{c_{\hat{\boldsymbol\theta}}(i,i)}} \right| > t_{n-p-1,1-\alpha/2} \right\}, \quad i = 0, \dots, p, \tag{11.48}$$

è la regione critica di livello $\alpha \in (0,1)$ per il test

$$\mathcal{H}_0 \colon \theta_i = t \quad \text{contro} \quad \mathcal{H}_1 \colon \theta_i \neq t.$$

Tipicamente si usa $t = 0$. Volendo invece verificare l'ipotesi che i predittori x_i e x_j abbiano lo stesso effetto sulla risposta, è necessario eseguire il test dell'ipotesi

$$\mathcal{H}_0 \colon \theta_i = \theta_j \quad \text{contro} \quad \mathcal{H}_1 \colon \theta_i \neq \theta_j. \tag{11.49}$$

La v.a. $\hat\theta_{ij} = \hat\theta_i - \hat\theta_j$ ha legge $\hat\theta_{ij} \sim \mathtt{N}\big(\theta_i - \theta_j, \sigma^2\left(c_{\hat{\boldsymbol\theta}}(i,i) + c_{\hat{\boldsymbol\theta}}(j,j) - 2c_{\hat{\boldsymbol\theta}}(i,j)\right)\big)$. Dunque, nell'ipotesi $\mathcal{H}_0$,

$$T_{ij} = \frac{\hat\theta_i - \hat\theta_j}{S\sqrt{c_{\hat{\boldsymbol\theta}}(i,i) + c_{\hat{\boldsymbol\theta}}(j,j) - 2c_{\hat{\boldsymbol\theta}}(i,j)}} \sim \mathtt{t}(n-p-1)\,, \tag{11.50}$$

con cui è facile costruire IC opportuni, e

$$C_{ij} = \left\{ |T_{ij}| > t_{n-p-1,1-\alpha/2} \right\} \tag{11.51}$$

è la regione critica di livello $\alpha \in (0,1)$ per il test dell'ipotesi considerata.

Nelle applicazioni è spesso utile effettuare un cosiddetto *test globale di dipendenza lineare*, in cui si verifica l'ipotesi

$$\mathcal{H}_0 \colon \theta_1 = \dots = \theta_p = 0 \quad \text{contro} \quad \mathcal{H}_1 \colon \theta_i \neq 0 \quad \text{per qualche } i = 1, \dots, p. \tag{11.52}$$

Si noti che se l'ipotesi fosse vera non vi sarebbe alcuna dipendenza (lineare) della risposta dai predittori. Sia $E = (\boldsymbol{A}\mathbb{R}^{p+1})$ il sottospazio generato dai vettori $\boldsymbol{x}^{(0)}, \dots, \boldsymbol{x}^{(p)}$, e si considerino i seguenti sottospazi di $\mathbb{R}^n$:

- E_0, il sottospazio (di E) generato da $\boldsymbol{x}^{(0)}$, con $\dim(E_0) = 1$;
- $E_1 = E_0^\perp$, l'insieme dei vettori di E ortogonali a E_0, con $\dim(E_1) = p$;
- $E_2 = E^\perp$, l'insieme dei vettori di $\mathbb{R}^n$ ortogonali a E, con $\dim(E_2) = n - p - 1$.

Tali sottospazi sono evidentemente tutti ortogonali tra loro. Usando una notazione intuitiva, dove $\boldsymbol{P}_\cdot$ indica la proiezione ortogonale sul sottospazio corrispondente, è facile vedere che $\boldsymbol{P}_{E_0}\boldsymbol{y} = \overline{y}\boldsymbol{x}^{(0)}$: infatti, si tratta del vettore di E_0 avente distanza minima da $\boldsymbol{y}$. Inoltre, essendo $\boldsymbol{P}_E\boldsymbol{y} = \hat{\boldsymbol{y}}$ e $\boldsymbol{P}_{E_1} = \boldsymbol{P}_E - \boldsymbol{P}_{E_0}$, si ha $\boldsymbol{P}_{E_1}\boldsymbol{y} = \hat{\boldsymbol{y}} - \overline{y}\boldsymbol{x}^{(0)}$. Infine, $\boldsymbol{P}_{E_2}\boldsymbol{y} = \boldsymbol{y} - \hat{\boldsymbol{y}} = \boldsymbol{r}$. Se l'ipotesi $\mathcal{H}_0$ è vera allora $\boldsymbol{y} = \theta_0\boldsymbol{x}^{(0)} + \sigma\boldsymbol{Z}$, e poiché $\boldsymbol{P}_{E_0}\theta_0\boldsymbol{x}^{(0)} = \theta_0\boldsymbol{x}^{(0)}$ mentre $\boldsymbol{P}_{E_1}\theta_0\boldsymbol{x}^{(0)} = \boldsymbol{P}_{E_2}\theta_0\boldsymbol{x}^{(0)} = \boldsymbol{0}_n$, si ha

$$\begin{aligned}
\boldsymbol{P}_{E_0}\boldsymbol{y} &= \theta_0\boldsymbol{x}^{(0)} + \sigma\boldsymbol{P}_{E_0}\boldsymbol{Z}, \\
\boldsymbol{P}_{E_1}\boldsymbol{y} &= \sigma\boldsymbol{P}_{E_1}\boldsymbol{Z}, \\
\boldsymbol{P}_{E_2}\boldsymbol{y} &= \sigma\boldsymbol{P}_{E_2}\boldsymbol{Z}.
\end{aligned}$$

Essendo $\dim(E_1) = p$ e $\dim(E_2) = n - p - 1$, per il *Teorema di Cochran* si ha

$$\frac{1}{\sigma^2} \, \|\boldsymbol{P}_{E_1}\boldsymbol{y}\|^2 = \|\boldsymbol{P}_{E_1}\boldsymbol{Z}\|^2 \sim \chi^2(p)\,,$$
$$\frac{1}{\sigma^2} \, \|\boldsymbol{P}_{E_2}\boldsymbol{y}\|^2 = \|\boldsymbol{P}_{E_2}\boldsymbol{Z}\|^2 \sim \chi^2(n-p-1)\,,$$

e le due v.a. sono indipendenti. Dunque, nell'ipotesi $\mathcal{H}_0$, la v.a.

$$F = \frac{n-p-1}{p} \, \frac{\|\boldsymbol{P}_{E_1}\boldsymbol{y}\|^2}{\|\boldsymbol{P}_{E_2}\boldsymbol{y}\|^2} = \frac{n-p-1}{p} \, \frac{\left\|\hat{\boldsymbol{y}} - \overline{y}\boldsymbol{x}^{(0)}\right\|^2}{\|\boldsymbol{r}\|^2}$$

segue una legge $\mathtt{F}(p, n-p-1)$. A questo punto è immediato procedere alla verifica dell'ipotesi, utilizzando

$$C_0 = \left\{ \frac{n-p-1}{p} \, \frac{\sum_{i=1}^n (\hat{y}_i - \overline{y})^2}{\sum_{i=1}^n (y_i - \hat{y}_i)^2} > f_{p,n-p-1,1-\alpha} \right\} \tag{11.53}$$

quale regione critica di livello $\alpha \in (0,1)$. Tradizionalmente i risultati sono presentati come mostrato in *Tabella (11.4)*, del tutto simile alla *Tabella (11.2)* relativa alla regressione lineare semplice; valgono inoltre le considerazioni circa il partizionamento delle fluttuazioni espresse nell'*Osservazione (11.6)*.

Causa	**Somma dei Quadrati**	**Gradi di Libertà**	**Media**	$\boldsymbol{F}$
Regressione	$\text{SQR} = \sum_{i=1}^n (\hat{y}_i - \overline{y})^2$	p	SQR/p	$F = \frac{\text{SQR}/p}{\text{SRQ}/(n-p-1)}$
Residui	$\text{SRQ} = \sum_{i=1}^n (y_i - \hat{y}_i)^2$	$n-p-1$	$\text{SRQ}/(n-p-1)$	

Tabella 11.4: presentazione standard dei risultati della verifica dell'ipotesi $\mathcal{H}_0 \colon \theta_1 = \cdots = \theta_p = 0$.

ESEMPIO 11.2. ►

La qualità Y del vino di Bordeaux è valutata in una scala da -3 a $+3$. Quali predittori di Y si utilizzano quattro misurazioni meteorologiche del mese di aprile precedente la vendemmia: $x^{(1)}$, la somma della temperature medie giornaliere; $x^{(2)}$, le ore di insolazione; $x^{(3)}$, i millimetri di pioggia caduti; $x^{(4)}$, il numero di giorni in cui la temperatura ha superato la media stagionale. La *Tabella (11.5)* riporta i valori delle osservazioni disponibili per gli anni dal 1924 al 1955.

I risultati della regressione lineare multipla rispetto ai quattro predittori e al fattore costante sono mostrati di seguito; le quantità indicate sono le stesse di quelle discusse nello studio dell'acqua alta a Venezia.

Parametro	**Stima**	**Errore standard**	$\boldsymbol{T}$	$\boldsymbol{p}$**-value**
$\boldsymbol{\theta_0}$	-26.7737	5.3845	-4.9724	0.0000
$\boldsymbol{\theta_1}$	0.0077	0.0019	3.9712	0.0005
$\boldsymbol{\theta_2}$	0.0044	0.0014	3.0610	0.0049
$\boldsymbol{\theta_3}$	-0.0060	0.0015	-3.8753	0.0006
$\boldsymbol{\theta_4}$	-0.0237	0.0257	-0.9196	0.3659
$\boldsymbol{\sigma}$	0.7002	$\boldsymbol{R^2} = 0.8726$		

Apparentemente l'elevato valore di $R^2 \approx 88\%$ e le ridotte varianze degli stimatori sembrano indicare che il modello proposto sia ragionevole. In effetti, calcolando le quantità mostrate in *Tabella (11.4)* si ottiene:

Causa	**Somma dei Quadrati**	**Gradi di Libertà**	**Media**	$\boldsymbol{F}$
Regressione	90.6378	4	22.6595	46.2187
Residui	13.2372	27	0.4903	

Anno	Y	$x^{(1)}$	$x^{(2)}$	$x^{(3)}$	$x^{(4)}$	Anno	Y	$x^{(1)}$	$x^{(2)}$	$x^{(3)}$	$x^{(4)}$
1924	1	3064	1201	361	10	1940	0	3094	1329	339	11
1925	−1	3000	1053	338	11	1941	−2	3009	1210	536	15
1926	1	3155	1133	393	19	1942	1	3227	1331	414	21
1927	−2	3080	970	467	4	1943	2	3308	1368	282	24
1928	2	3245	1258	294	36	1944	1	3212	1289	302	17
1929	3	3267	1386	225	35	1945	3	3381	1444	253	25
1930	−2	3080	966	417	13	1946	0	3061	1175	261	12
1931	−2	2974	1185	488	12	1947	3	3478	1317	259	42
1932	−3	3038	1103	677	14	1948	1	3126	1248	315	11
1933	1	3318	1310	427	29	1949	3	3468	1508	286	43
1934	2	3317	1362	326	25	1950	1	3252	1361	346	26
1935	−1	3182	1171	326	28	1951	−2	3052	1186	443	14
1936	−2	2998	1102	349	9	1952	2	3270	1399	306	24
1937	2	3221	1424	382	21	1953	2	3198	1299	367	20
1938	0	3019	1239	275	16	1954	−2	2904	1164	311	6
1939	0	3022	1285	303	9	1955	2	3247	1277	375	19

Tabella 11.5: i dati dell'*Esempio (11.2).*

Perciò, essendo $f_{4,27,0.99} \approx 4.11$, in base all'*equazione (11.53)* l'ipotesi $\mathcal{H}_0 \colon \theta_1 = \theta_2 = \theta_3 = \theta_4 = 0$ è da rifiutare (almeno) al livello $\alpha = 0.01$. ◀

Un'attenta analisi dei risultati rivela però che il p-value relativo a $x^{(4)}$ è decisamente elevato, ciò che fa sospettare che, in realtà, tale predittore giochi un ruolo marginale: infatti, sarebbe possibile accettare l'ipotesi $\mathcal{H}_0 \colon \theta_4 = 0$ a tutti i livelli $\alpha < 0.3659$, cioè quelli comunemente adottati nella pratica statistica. Per verificare che $x^{(4)}$ non rappresenti una componente fondamentale del modello, tipicamente si effettua una nuova regressione scartando tale predittore. I risultati sono riportati di seguito.

Parametro	Stima	Errore standard	T	p-value
$\boldsymbol{\theta_0}$	−22.8693	3.3022	−6.9254	0.0000
$\boldsymbol{\theta_1}$	0.0063	0.0012	5.0714	0.0000
$\boldsymbol{\theta_2}$	0.0043	0.0014	3.0225	0.0053
$\boldsymbol{\theta_3}$	−0.0059	0.0015	−3.8538	0.0006
$\boldsymbol{\sigma}$	0.6983	$\boldsymbol{R^2} = 0.8686$		

Rispetto al modello originale, R^2 è rimasto praticamente invariato, così come le stime e le varianze dei parametri, mentre addirittura S^2 è diminuito: ciò spinge a ritenere che questo secondo modello fornisca una descrizione della risposta altrettanto adeguata, utilizzando però un numero ridotto di parametri. Calcolando le quantità mostrate in *Tabella (11.4)* per il modello ridotto si ottiene:

Causa	Somma dei Quadrati	Gradi di Libertà	Media	F
Regressione	90.2232	3	30.0744	61.6831
Residui	13.6518	28	0.4876	

Perciò, essendo $f_{3,28,0.99} \approx 4.57$, in base all'*equazione (11.53)* l'ipotesi $\mathcal{H}_0 \colon \theta_1 = \theta_2 = \theta_3 = 0$ è da rifiutare (almeno) al livello $\alpha = 0.01$.

Capitolo 12

SIMULAZIONE

Nella pratica si presenta spesso la necessità di *simulare* il comportamento aleatorio dei fenomeni presi in esame, ossia di generare (tramite opportuni algoritmi numerici) un insieme di valori aventi una legge di probabilità assegnata. Nei pacchetti statistici disponibili sul mercato esistono comandi che forniscono sequenze di numeri casuali *indipendenti* con le distribuzioni più comuni. Se però si presenta la necessità di simulare v.a. non standard, o di scrivere algoritmi in uno dei linguaggi di programmazione usuali (C, FORTRAN, ...), è necessario conoscere alcune tecniche fondamentali di simulazione numerica.

In questo Capitolo si illustrano alcuni metodi di base; per ulteriori dettagli si consultino [17, 1]. In ogni caso supporremo che sia disponibile un algoritmo in grado di generare una sequenza sufficientemente lunga di numeri casuali indipendenti con legge $\mathtt{U_c}(0,1)$. Per la descrizione di alcune routine nei linguaggi C e FORTRAN si consultino, rispettivamente, [23] e [22].

Nel seguito u indicherà il numero casuale restituito dalla routine di generazione aleatoria considerata, sicché u rappresenterà la realizzazione di una v.a. $U \sim \mathtt{U_c}(0,1)$ (e dunque $u \in (0,1)$); generazioni successive si intenderanno indipendenti.

12.1 Trasformazione Integrale di Probabilità

Un primo modo per generare numeri casuali Y, con distribuzione assegnata F_Y, consiste nel costruire una funzione τ tale che se $U \sim \mathtt{U_c}(0,1)$ allora $Y = \tau(U) \sim F_Y$. Se F_Y è *continua* e *invertibile*, la scelta $\tau = F_Y^{-1}$ risolve il problema. Infatti, essendo $F_U(u) = u$ per ogni $u \in [0,1]$, preso $t \in \mathbb{R}$ si ha

$$\mathbf{P}\{Y \le t\} = \mathbf{P}\{\tau(U) \le t\} = \mathbf{P}\big\{F_Y^{-1}(U) \le t\big\} = \mathbf{P}\{U \le F_Y(t)\} = F_U(F_Y(t)) = F_Y(t) \tag{12.1}$$

e dunque la v.a. Y così costruita ha la legge desiderata.

NOTA 12.1.

È interessante osservare che se $U \sim \mathtt{U_c}(0,1)$ allora $V = 1 - U \sim \mathtt{U_c}(0,1)$: infatti, se $v \in (0,1)$, si ha

$$\mathbf{P}\{V \le v\} = \mathbf{P}\{1 - U \le v\} = \mathbf{P}\{U > 1 - v\} = 1 - (1 - v) = v \tag{12.2}$$

e dunque $V \sim \mathtt{U_c}(0,1)$.

Nel seguito si illustra ora con alcuni esempi l'applicazione del metodo della *Trasformazione Integrale di Probabilità*. Qui sotto si riporta un programma elementare in MATLAB con il quale sono state preparate le figure dell'*Esempio (12.1)*.

```
% Simulazione della Legge Uniforme su (a,b)
function s = SIM_Uab(N,a,b);

% N: ampiezza del campione da simulare
% a: limite inferiore della Legge Uniforme su (a,b)
```

```
% b: limite superiore della Legge Uniforme su (a,b)
% s: simulazione

u = rand(N,1); % generazione di N numeri casuali i.i.d. U(0,1)
s = a + (b-a)*u; % simulazione della Legge Uniforme U(a,b)

return;
```

.....

ESEMPIO 12.1 (Legge $\mathtt{U_c}(a, b)$). ►

Si voglia simulare una v.a. $Y \sim \mathtt{U_c}(a, b)$, $a, b \in \mathbb{R}$, $a < b$. Poiché F vale

$$F(y) = \begin{cases} 0, & y \leq a \\ \frac{y-a}{b-a}, & a < y < b \\ 1, & y \geq b \end{cases}$$

ne scende che $y = F^{-1}(u) = a + (b - a)u$, $u \in (0, 1)$; dunque

$$Y = \tau(U) = a + (b - a)U \sim \mathtt{U_c}(a, b)\,.$$

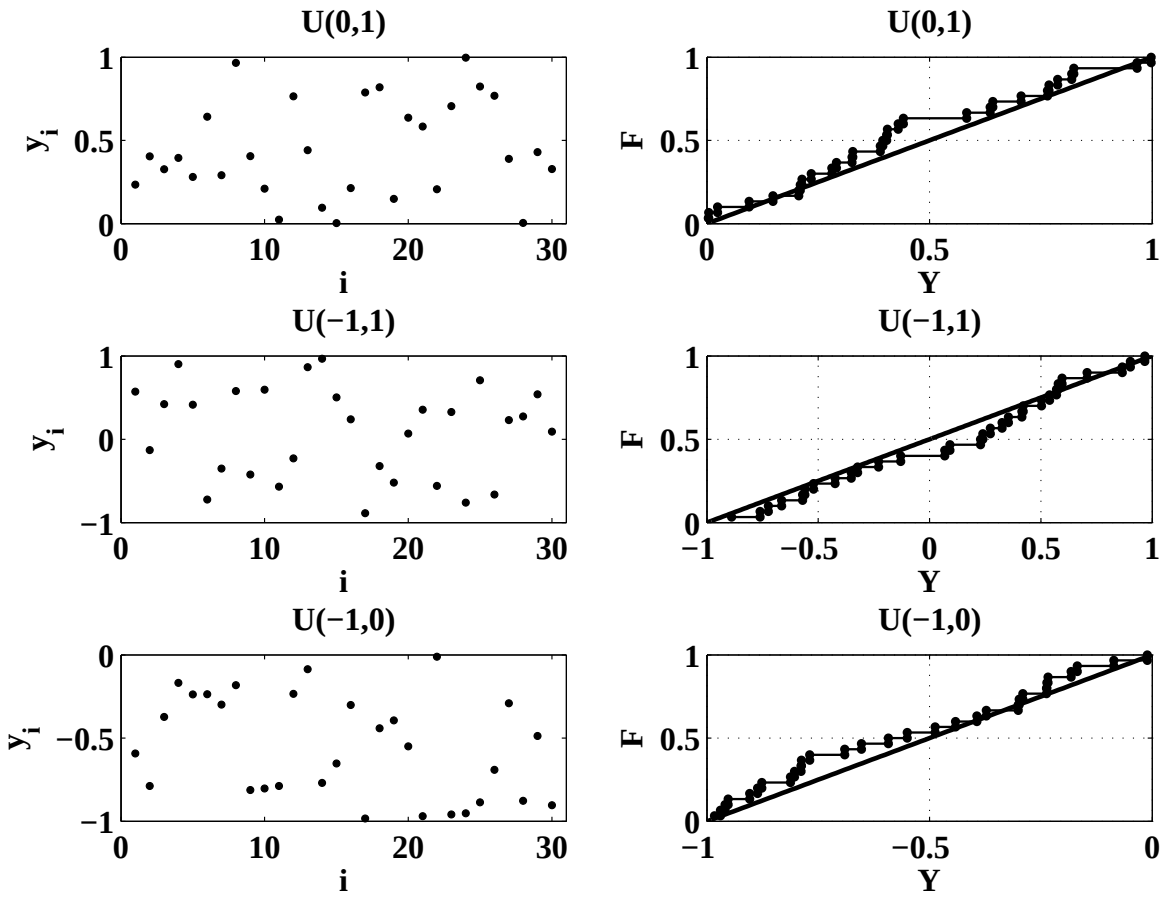

Figura 12.1: esempi di simulazione di una v.a. $Y \sim \mathtt{U_c}(a, b)$.

La *Figura (12.1)* mostra alcune simulazioni di 30 valori casuali estratti dalla distribuzione $\mathtt{U_c}(a, b)$ per diversi valori dei parametri; per il commento dei risultati si veda l'*Esempio (9.10).* ◄

ESEMPIO 12.2 (Legge $\mathtt{e}(\lambda)$). ►

Si voglia simulare una v.a. $Y \sim \mathtt{e}(\lambda) \sim \Gamma(1, \lambda)$, $\lambda > 0$. Poiché F vale

$$F(y) = \begin{cases} 0, & y < 0 \\ 1 - e^{-\lambda y}, & y \geq 0 \end{cases}$$

ne scende che $y = F^{-1}(u) = -\frac{1}{\lambda} \ln(1-u)$, $u \in (0,1)$. Grazie all'*equazione (12.2)* è immediato mostrare che

$$Y = \tilde{\tau}(U) = -\frac{1}{\lambda} \ln(U) \sim \mathtt{e}(\lambda)\,.$$

Si osservi che quest'ultima soluzione offre alcuni vantaggi di carattere numerico rispetto a quella standard data da $Y = \tau(U) = -\frac{1}{\lambda} \ln(1-U)$.

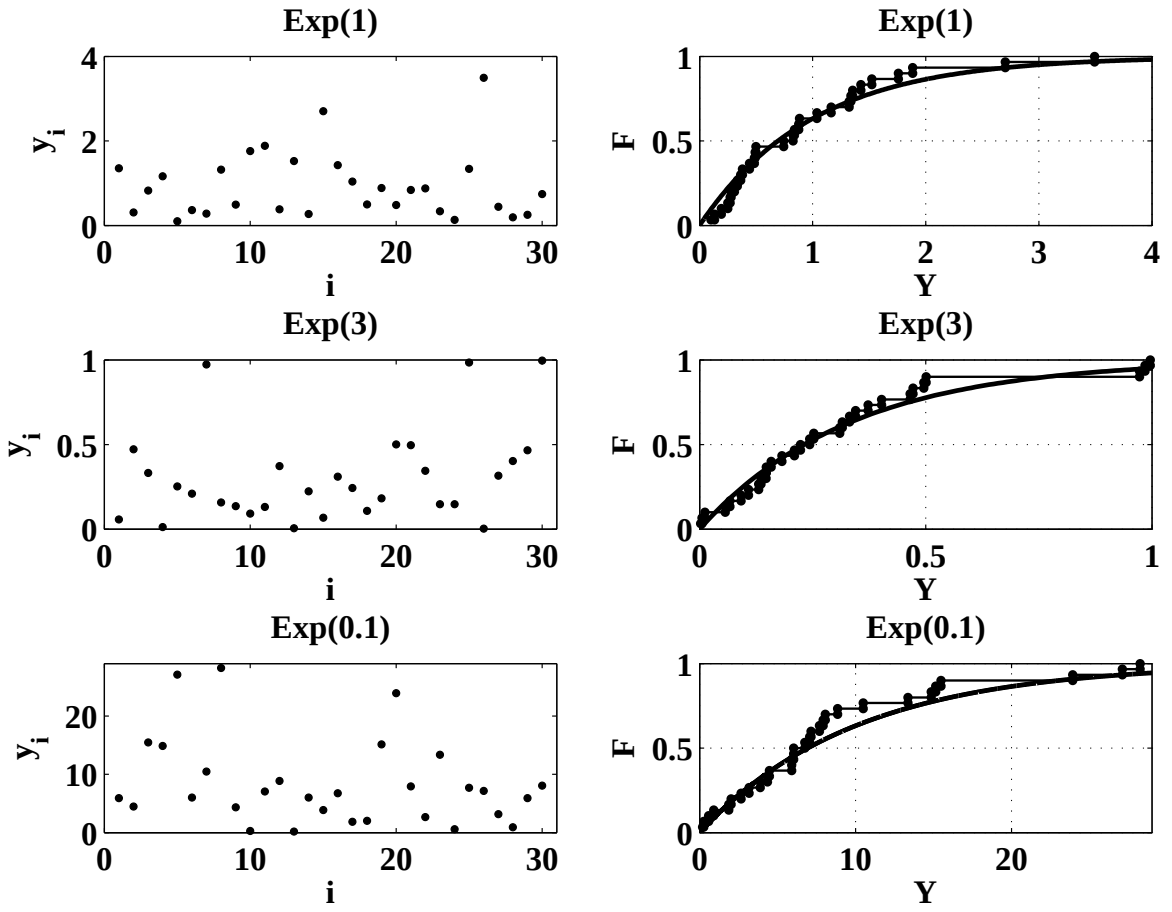

Figura 12.2: esempi di simulazione di una v.a. $Y \sim \mathtt{e}(\lambda)$.

La *Figura (12.2)* mostra alcune simulazioni di 30 valori casuali estratti dalla distribuzione $\mathtt{e}(\lambda)$ per diversi valori del parametro; per il commento dei risultati si veda l'*Esempio (9.10)*. ◄

ESEMPIO 12.3 (Legge *Pareto Generalizzata*). ►

La legge *Pareto Generalizzata* è di grande importanza nelle applicazioni, sia per le numerose proprietà di carattere matematico sia per la capacità di descrivere la statistica degli eventi estremi (si veda il *Capitolo 3*), fatto di notevole utilità, ad esempio, in finanza e in geofisica. Per una v.a. Y avente legge $\mathtt{PG}(b, c, k)$ la distribuzione F vale

$$F(y) = \begin{cases} 0, & y < b \\ 1 - \left(1 - k\,\frac{y-b}{c}\right)^{1/k}, & y \geq b \end{cases}$$

dove $b \in \mathbb{R}$ è un parametro di *posizione*, $c > 0$ è un parametro di *scala* e $k < 0$ è un parametro di *forma* (l'espressione data è solo formalmente diversa rispetto a quella dell'*equazione (3.28)*). Si osservi che, in generale, k può assumere valori su tutto $\mathbb{R}$ (con opportuna definizione del supporto della distribuzione). Ad esempio il caso $k = 0$ corrisponde alla legge *esponenziale* (a meno di una trasformazione affine), e il caso $k = 1$ corrisponde alla legge *uniforme* (a meno di una trasformazione affine). Nel caso $k < 0$, $\mathbf{E}(Y^m)$ esiste se, e solo se, $m < -1/k$, ciò che rende la distribuzione *Pareto Generalizzata* di interesse per la generazione di *valori estremi*. Ne scende che $y = F^{-1}(u) = b + \frac{c}{k}\left(1 - (1-u)^k\right)$, $u \in (0,1)$; dunque, grazie all'*equazione (12.2)*,

$$Y = \tilde{\tau}(U) = b + \frac{c}{k}\left(1 - U^k\right) \sim \texttt{PG}(b, c, k)\,.$$

Si osservi che quest'ultima soluzione offre alcuni vantaggi di carattere numerico rispetto a quella standard data da $Y = \tau(U) = b + \frac{c}{k}\left(1 - (1-U)^k\right)$.

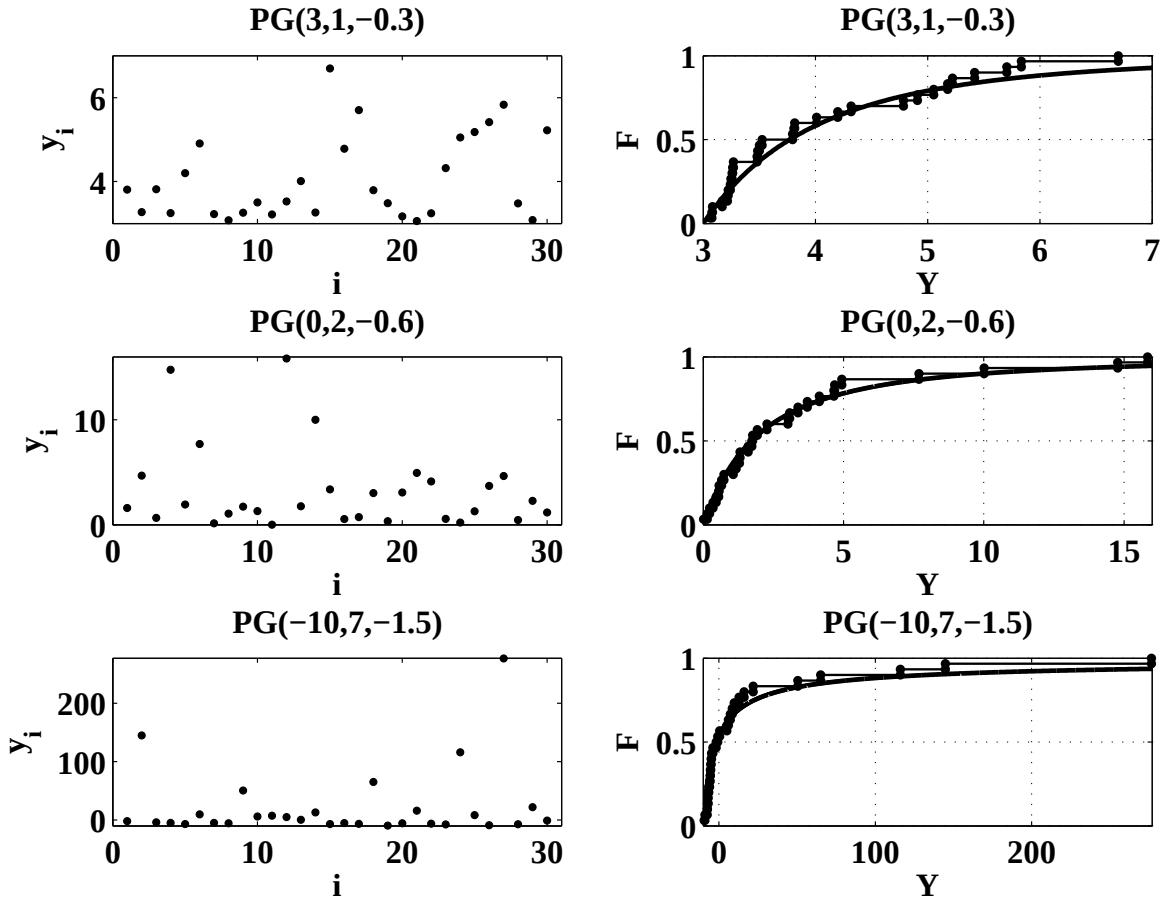

Figura 12.3: esempi di simulazione di una v.a. $Y \sim \texttt{PG}(b, c, k)$.

La *Figura (12.3)* mostra alcune simulazioni di 30 valori casuali estratti dalla distribuzione $\texttt{PG}(b, c, k)$ per diversi valori dei parametri; per il commento dei risultati si veda l'*Esempio (9.10)*. ◀

Il metodo della *Trasformazione Integrale di Probabilità* è poco efficace quando la funzione τ non abbia una espressione analitica semplice (come accade, ad esempio, per la legge Normale), oppure quando si voglia simulare una probabilità su $\mathbb{R}^d$, $d > 1$ — a tale proposito si vedano le *Sezioni 12.3–12.4*.

12.2 Ulteriori schemi di simulazione univariata

Negli esempi che seguono si presentano ulteriori schemi di simulazione univariata.

ESEMPIO 12.4 (Legge $\mathtt{Be}(p)$). ►
Si voglia simulare una v.a. $Y \sim \mathtt{Be}(p)$, con $p \in (0,1)$. Se $U \sim \mathtt{U_c}(0,1)$, allora

$$Y = \mathbb{I}_{(0,p]}(U)$$

è una v.a. di legge $\mathtt{Be}(p)$. ◄

ESEMPIO 12.5 (Legge $\mathtt{Bi}(n,p)$). ►
Si voglia simulare una v.a. $Y \sim \mathtt{Bi}(n,p)$, con $n \in \mathbb{N}$ e $p \in (0,1)$. Se $U_1, \ldots, U_n$ è un insieme di v.a. i.i.d. di legge $\mathtt{U_c}(0,1)$, allora

$$Y = \sum_{i=1}^{n} \mathbb{I}_{(0,p]}(U_i)$$

è una v.a. di legge $\mathtt{Bi}(n,p)$, essendo la somma di n v.a. i.i.d. di legge $\mathtt{Be}(p)$. ◄

ESEMPIO 12.6 (Legge $\mathtt{g}(p)$). ►
Si voglia simulare una v.a. $Y \sim \mathtt{g}(p)$, con $p \in (0,1)$. Se $U_1, U_2, \ldots$ è una successione di v.a. i.i.d. di legge $\mathtt{U_c}(0,1)$, e poiché il tempo di attesa dell'istante di primo successo in un processo di Bernoulli ha legge Geometrica, ne segue che se

$$Z = \inf\left\{ n : \sum_{i=1}^{n} \mathbb{I}_{(0,p]}(U_i) = 1 \right\}$$

allora $Y = Z - 1$ è una v.a. di legge $\mathtt{g}(p)$. ◄

ESEMPIO 12.7 (Legge $\mathtt{P}(\lambda)$). ►
Si voglia simulare una v.a. $Y \sim \mathtt{P}(\lambda)$, $\lambda > 0$. Consideriamo anzitutto una successione $X_1, \ldots, X_n, \ldots$ di v.a. i.i.d. di legge $\mathtt{e}(\lambda)$. Una (tediosa) integrazione per parti mostra che la somma $S_n = \sum_{i=1}^{n} X_i$, che è ovviamente non decrescente in n, ha legge $\Gamma(n,\lambda)$, con funzione di ripartizione

$$F_n(x) = 1 - \sum_{i=0}^{n-1} e^{-\lambda x} \frac{(\lambda x)^i}{i!}$$

per $x > 0$ e nulla altrove. Sia ora $T > 0$ una soglia prefissata e consideriamo la v.a. $Y = \sup\{n \colon S_n \leq T\}$. L'equivalenza degli eventi $\{Y \leq k\} \equiv \{S_{k+1} > T\}$, $k \in \mathbb{N}$, mostra che

$$\mathbf{P}\{Y \leq k\} = \mathbf{P}\{S_{k+1} > T\} = \sum_{i=0}^{k} e^{-\lambda T} \frac{(\lambda T)^i}{i!}.$$

Riconosciamo in questa espressione la funzione di ripartizione di una v.a. di legge $\mathtt{P}(\lambda T)$. Dunque è sufficiente porre $T = 1$, simulare una successione di v.a. i.i.d. di legge $\mathtt{e}(\lambda)$ (si veda l'*Esempio (12.2)*) e porre $Y = n$ se n è il più grande intero tale per cui $S_n \leq 1$. ◄

ESEMPIO 12.8 (Legge del dado). ►
Si voglia simulare un dado non truccato, cioè una v.a. Y avente legge uniforme sull'insieme $\{1,2,3,4,5,6\}$. Se $U \sim \mathtt{U_c}(0,1)$ allora la v.a. Y data da, per $i = 1, \ldots, 6$,

$$Y = i \quad \text{se} \quad \frac{i-1}{6} \leq U < \frac{i}{6}$$

è la v.a. cercata. ◄

12.3 Metodo del Rigetto

Al fine di ovviare ad alcune delle pecche del metodo della *Trasformazione Integrale di Probabilità* è possibile ricorrere ad altri schemi di simulazione. Tra questi un posto di primo piano è occupato dal cosiddetto *Metodo del Rigetto* o *di Reiezione*; esso si basa sul seguente risultato, che fornisce un metodo per generare v.a. con distribuzione uniforme su un arbitrario sottoinsieme (misurabile) $D \in \mathbb{R}^d$.

PROPOSIZIONE 12.1. *Siano $R \subseteq \mathbb{R}^d$ un sottoinsieme misurabile, $Z_1, \ldots, Z_n, \ldots$ una successione di v.a. i.i.d. a valori in R e $D \subseteq R$ un sottoinsieme misurabile tale che $\mathbf{P}\{Z_i \in D\} = p > 0$. Si definisca inoltre il* Tempo di Arresto $T = \inf\{i \colon Z_i \in D\}$ *e la v.a.*

$$X = \begin{cases} Z_k, & \text{se } T = k \text{ per qualche } k \in \mathbb{N} \\ \infty, & \text{se } T = +\infty \end{cases}$$

Allora, se $A \subseteq D$ è un sottoinsieme misurabile, vale

$$\mathbf{P}\{X \in A\} = \frac{\mathbf{P}\{Z \in A\}}{\mathbf{P}\{Z \in D\}}$$

In particolare, se Z è uniforme su R allora X è uniforme su D (e quindi su A).

DIM. Se $A \subseteq D$ allora, osservando che $X = Z_k$ se $T = k$,

$$\begin{aligned} \mathbf{P}\{X \in A\} &= \sum_{k=1}^{\infty} \mathbf{P}\{X \in A, T = k\} \\ &= \sum_{k=1}^{\infty} \mathbf{P}\{Z_k \in A, T = k\} \\ &= \sum_{k=1}^{\infty} \mathbf{P}\{Z_k \in A, Z_1 \notin D, \ldots, Z_{k-1} \notin D\} \\ &= \mathbf{P}\{Z \in A\} \sum_{k=1}^{\infty} (1-p)^{k-1} \end{aligned}$$

che vale $\mathbf{P}\{Z \in A\} / p$, da cui la tesi □

ESEMPIO 12.9. ►

Si voglia simulare una v.a. Y avente *densità* f limitata (cioè $f \leq M$, $M \in \mathbb{R}^+$) e a supporto compatto contenuto nell'intervallo $[a, b]$ — si veda la *Figura (12.4)*. Si supponga di disporre di un algoritmo in grado di fornire una v.a. bidimensionale $W = (U, V)$ *uniforme* nella regione di piano $D = \{(u, v) \colon a \leq u \leq b, 0 \leq v \leq f(u)\}$, cioè il sottografico di f. Si noti che U e V non sono indipendenti: infatti, deve essere $V \leq f(U)$. Ne segue che U ha densità f, in quanto

$$\mathbf{P}\{U \leq u\} = \mathbf{P}\{W \in D_u\} = \mathbf{P}\{(U, V) \in D_u\} = \int_a^u f(x)\, \mathrm{d}x = F_U(u)$$

dove D_u è l'intersezione del sottografico D di f e del semipiano $\{(x, y) \colon x \leq u\}$.

Dunque la simulazione di Y è risolta se si è in grado di simulare una v.a. W uniforme sul sottografico D. Grazie alla *Proposizione (12.1)* è possibile agire come segue: si simulano dapprima delle v.a. $Z_1, Z_2, \ldots$ i.i.d. uniformi sul rettangolo $[a, b] \times [0, M]$ (si ricordi che M è un maggiorante di f); per quanto visto nella *Sezione 12.1* è sufficiente prendere $Z_i = (X, Y)$, dove le v.a. X e Y sono indipendenti con leggi, rispettivamente, $X \sim \mathtt{U_c}(a, b)$ e $Y \sim \mathtt{U_c}(0, M)$. Si pone poi $W = Z_T$, dove T è il primo indice i per cui si abbia $\{Z_i \in D\}$, ovvero per cui succeda che $V_i \leq f(U_i)$. In base a quanto precedentemente illustrato la prima componente del vettore W avrà densità f. ◄

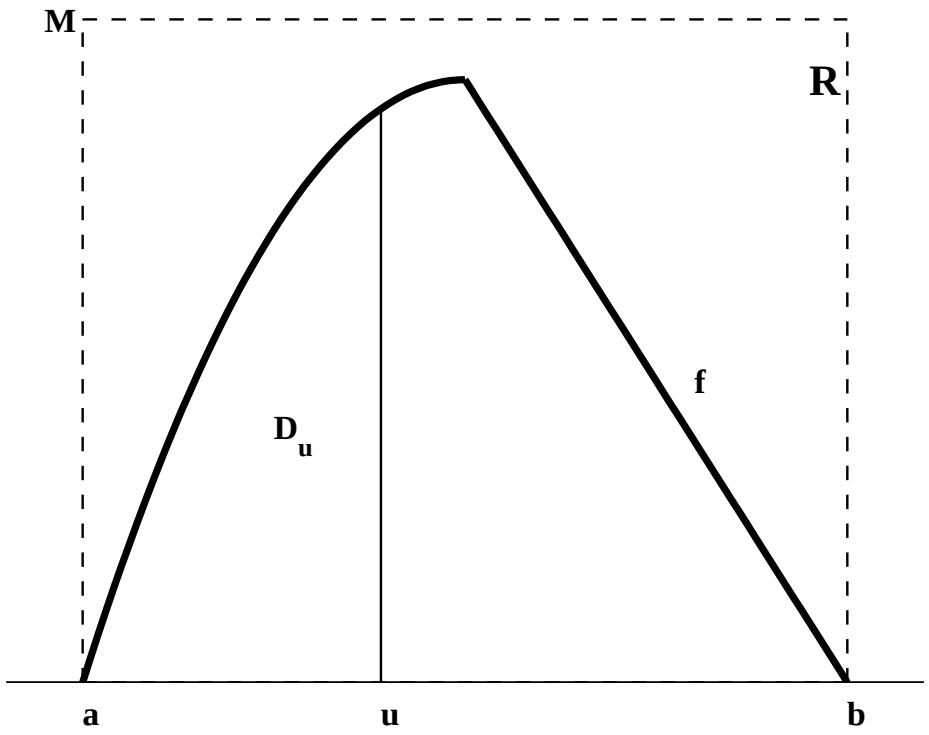

Figura 12.4: illustrazione del *Metodo del Rigetto*.

12.4 Copule e simulazione multivariata

La modellizzazione e simulazione di v.a. multivariate *non indipendenti* rappresenta da sempre un problema complesso nelle applicazioni [13]. Sviluppi recenti nel campo della Statistica Matematica hanno mostrato come, grazie all'utilizzo della *Teoria delle Copule* [30, 15, 21, 26], sia possibile aggirare e far fronte a molte delle difficoltà poste da tale problema. Nel seguito studieremo il caso bivariato, che può essere esteso a quello più generale di leggi multivariate attraverso tecniche opportune. Un software per l'applicativo "R" è disponibile in rete per lavorare con le copule [35].

Un modo semplice ed efficace di studiare eventi multivariati, ove le v.a. in gioco non siano indipendenti, consiste nello studiare separatamente le singole leggi marginali e la *struttura di dipendenza* (o *Copula*) che "lega" tra loro le variabili. Introduciamo anzitutto la nozione di 2-copula.

DEFINIZIONE 12.1 (2-Copula)**.** Sia $\mathbb{I} = [0, 1]$. Una 2-copula è una funzione bivariata $\mathbf{C}\colon \mathbb{I} \times \mathbb{I} \to \mathbb{I}$ tale che:

1. (marginali uniformi) per ogni $u, v \in \mathbb{I}$,

$$\mathbf{C}(u, 0) = 0, \quad \mathbf{C}(u, 1) = u, \quad \mathbf{C}(0, v) = 0, \quad \mathbf{C}(1, v) = v; \tag{12.3a}$$

2. (2-crescenza) per ogni $u_1, u_2, v_1, v_2 \in \mathbb{I}$ tali che $u_1 \leq u_2$ e $v_1 \leq v_2$,

$$\mathbf{C}(u_2, v_2) - \mathbf{C}(u_2, v_1) - \mathbf{C}(u_1, v_2) + \mathbf{C}(u_1, v_1) \geq 0. \tag{12.3b}$$

La *Figura (12.5)* illustra i vincoli enunciati nella *Definizione (12.1)*.

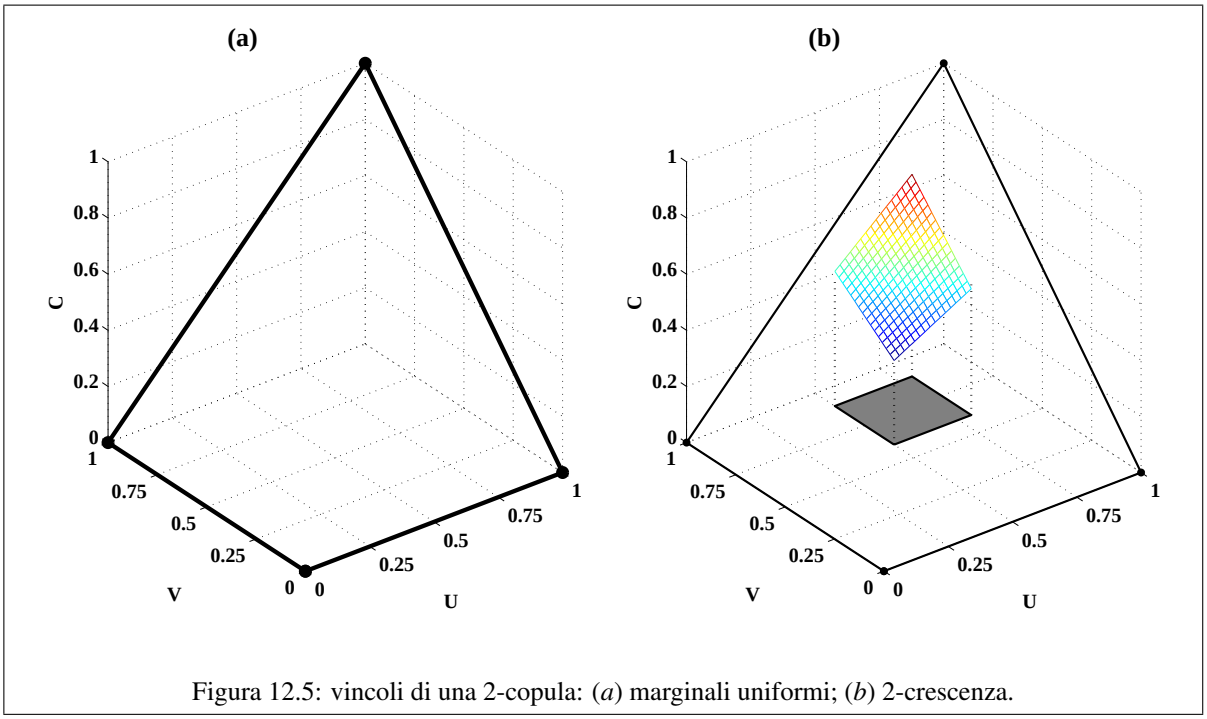

Figura 12.5: vincoli di una 2-copula: (*a*) marginali uniformi; (*b*) 2-crescenza.

Una 2-copula è uniformemente continua su $\mathbb{I}^2$; parimenti, sono uniformemente continue le sue *sezioni* orizzontali, verticali e lungo la diagonale principale del quadrato unitario. Da un punto di vista pratico $\mathbf{C}$ non è altro che una distribuzione di probabilità su $\mathbb{I}^2$ con marginali Uniformi. La definizione che segue è una conseguenza di questo fatto.

DEFINIZIONE 12.2 ($\mathbf{C}$-misura)**.** Una 2-copula $\mathbf{C}$ induce una misura di probabilità $\mu_{\mathbf{C}}$ su $\mathbb{I}^2$ data da

$$\mu_{\mathbf{C}}\left\{[0,u]\times[0,v]\right\} = \mathbf{C}\left(u,v\right) = \mathbf{P}\{U \leq u, V \leq v\}\,, \tag{12.4}$$

dove $u, v \in \mathbb{I}$ e $(U,V) \sim F_{UV} = \mathbf{C}$. Talora $\mu_{\mathbf{C}}$ è detta *misura doppiamente stocastica* (*doubly stochastic measure*).

Intuitivamente la misura $\mu_{\mathbf{C}}(A)$ di un insieme (misurabile) $A \subseteq \mathbb{I}^2$ è la probabilità che due v.a. U e V Uniformi su $\mathbb{I}$, e aventi legge congiunta $\mathbf{C}$, assumano valori in A. Il Teorema che segue è fondamentale [31] e sancisce il legame tra copule e distribuzioni bivariate.

TEOREMA 12.1 (Sklar (1959))**.** *Sia F_{XY} una distribuzione congiunta con marginali F_X e F_Y. Allora esiste una 2-copula $\mathbf{C}$ tale che*

$$F_{XY}(x,y) = \mathbf{P}\{X \leq x, Y \leq y\} = \mathbf{C}\left(F_X(x), F_Y(y)\right) \tag{12.5}$$

per ogni $(x,y) \in \mathbb{R}^2$. Se F_X, F_Y sono continue allora $\mathbf{C}$ è unica; altrimenti $\mathbf{C}$ è definita in maniera univoca su $\mathrm{Cod}(F_X) \times \mathrm{Cod}(F_Y)$.

*Viceversa, se $\mathbf{C}$ è una 2-copula e F_X, F_Y funzioni di ripartizione, allora la funzione F_{XY} data dall'*equazione *(12.5) è una distribuzione congiunta con marginali F_X e F_Y.*

Si noti che il *Teorema di Sklar* si può estendere facilmente al caso di dimensioni superiori a due. L'Esempio che segue mostra le proprietà di tre 2-copule fondamentali.

ESEMPIO 12.10. ►

Tre 2-copule speciali meritano una attenzione particolare; qui $(u,v) \in \mathbb{I}^2$.

1. Il *limite inferiore di Fréchet-Hoeffding* $\mathbf{W}_2$ dato da

$$\mathbf{W}_2(u,v) = \max\left\{u+v-1, 0\right\}. \tag{12.6}$$

2. Il *limite superiore di Fréchet-Hoeffding* $\mathbf{M}_2$ dato da

$$\mathbf{M}_2(u, v) = \min\{u, v\}. \tag{12.7}$$

3. La 2-copula *indipendenza* (o *prodotto*) $\mathbf{\Pi}_2$ data da

$$\mathbf{\Pi}_2(u, v) = uv. \tag{12.8}$$

La *Figura (12.6)* mostra i grafici di $\mathbf{W}_2$, $\mathbf{M}_2$ e $\mathbf{\Pi}_2$.

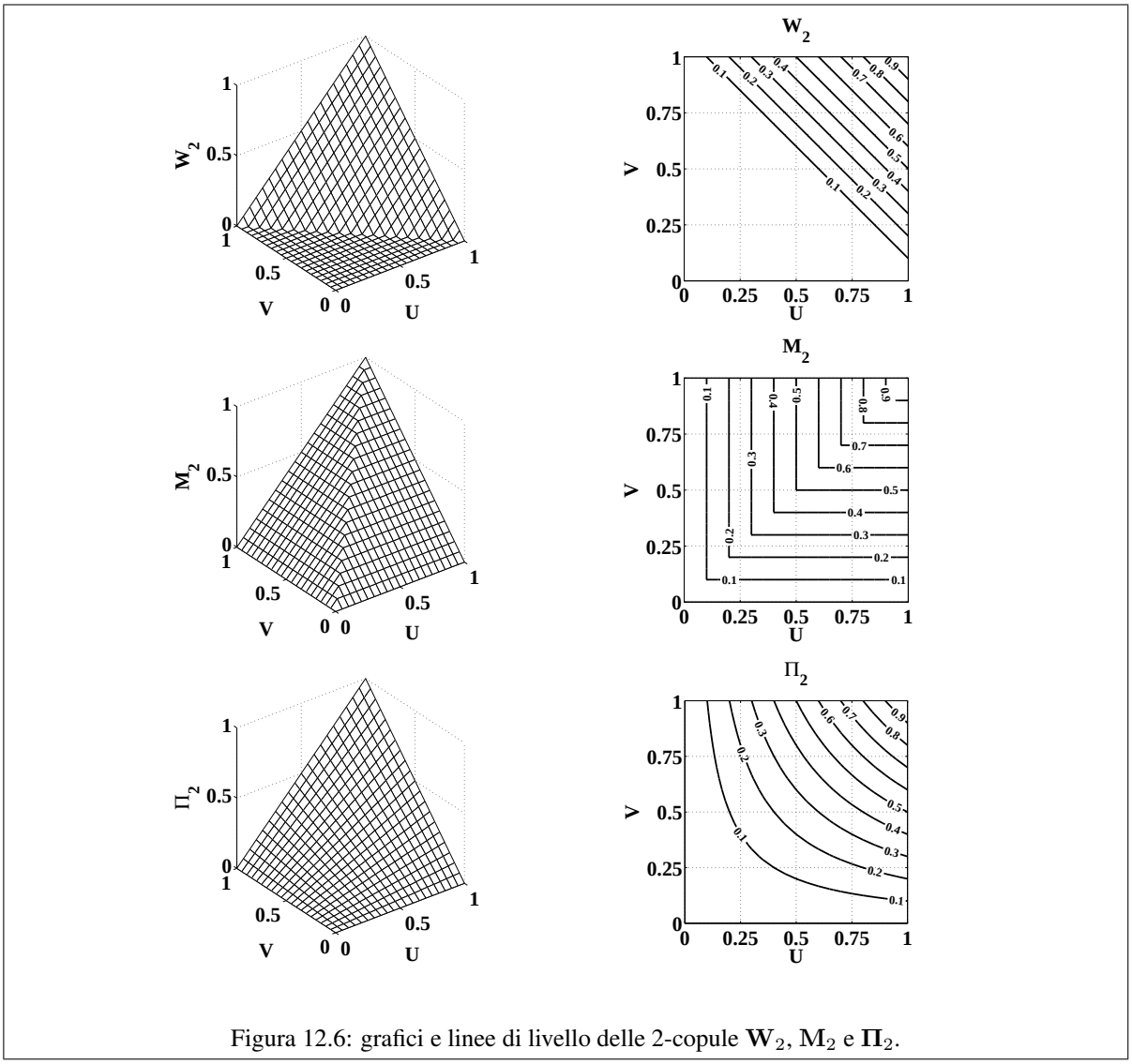

Figura 12.6: grafici e linee di livello delle 2-copule $\mathbf{W}_2$, $\mathbf{M}_2$ e $\mathbf{\Pi}_2$.

Le copule $\mathbf{W}_2$ e $\mathbf{M}_2$ forniscono limitazioni di carattere generale, in quanto per ogni 2-copula $\mathbf{C}$ e ogni coppia $(u, v) \in \mathbb{I}^2$ vale

$$\mathbf{W}_2(u, v) \leq \mathbf{C}(u, v) \leq \mathbf{M}_2(u, v). \tag{12.9}$$

In particolare, se X e Y sono v.a. continue, valgono le seguenti proprietà.

1. La variabile Y è quasi certamente una funzione strettamente decrescente di X se, e solo se, $\mathbf{C}_{XY} = \mathbf{W}_2$. Tali v.a. sono dette *contro-monotone*.

2. La variabile Y è quasi certamente una funzione strettamente crescente di X se, e solo se, $\mathbf{C}_{XY} = \mathbf{M}_2$. Tali v.a. sono dette *co-monotone*.

3. La 2-copula $\mathbf{\Pi}_2$ descrive l'assenza di dipendenza tra X e Y: infatti, se $F_{XY} = \mathbf{\Pi}_2(F_X, F_Y)$, allora $F_{XY} = F_X F_Y$. Quindi, X e Y sono indipendenti se, e solo se, la loro 2-copula è $\mathbf{\Pi}_2$.

Si noti che solo $\mathbf{M}_2$ e $\mathbf{\Pi}_2$ possono essere generalizzate al caso di più di due variabili. ◀

DEFINIZIONE 12.3 (Famiglia comprensiva)**.** Una famiglia di 2-copula che include $\mathbf{W}_2$, $\mathbf{M}_2$ e $\mathbf{\Pi}_2$ è detta *comprensiva*.

ESEMPIO 12.11 (Famiglia di Fréchet)**.** ▶

La famiglia di 2-copule data da

$$\mathbf{C}_{\alpha,\beta}(u,v) = \alpha\,\mathbf{M}_2(u,v) + (1-\alpha-\beta)\,\mathbf{\Pi}_2(u,v) + \beta\,\mathbf{W}_2(u,v), \tag{12.10}$$

dove $\alpha, \beta \geq 0$ e $\alpha + \beta \leq 1$, è detta *Famiglia di Fréchet*. Essa è evidentemente comprensiva, essendo una combinazione convessa di $\mathbf{W}_2$, $\mathbf{M}_2$ e $\mathbf{\Pi}_2$. ◀

Una proprietà importante delle 2-copule è l'*invarianza* rispetto a trasformazioni strettamente crescenti delle marginali, come mostrato dalla Proposizione che segue.

PROPOSIZIONE 12.2 (Invarianza)**.** *Siano X e Y v.a. continue con 2-copula $\mathbf{C}_{XY}$. Se α e β sono funzioni strettamente crescenti su, rispettivamente,* $\mathrm{Cod}(X)$ *e* $\mathrm{Cod}(Y)$*, allora*

$$\mathbf{C}_{\alpha(X)\beta(Y)} = \mathbf{C}_{XY}. \tag{12.11}$$

Dunque, usando la *Trasformazione Integrale di Probabilità* è possibile restringere l'attenzione alla coppia (non parametrica) di v.a. (U, V) date da

$$\begin{cases} U = F_X(X) \\ V = F_Y(Y) \end{cases} \iff \begin{cases} X = F_X^{[-1]}(U) \\ Y = F_Y^{[-1]}(V) \end{cases}, \tag{12.12}$$

dove $F_X^{[-1]}, F_Y^{[-1]}$ sono le *quasi-inverse* delle funzioni di ripartizione corrispondenti date da

1. se $t \in \mathrm{Cod}(F)$ allora $F^{[-1]}(t)$ è un qualsiasi valore $x \in \overline{\mathbb{R}}$ tale che $F(x) = t$, cioè, pe ogni $t \in \mathrm{Cod}(F)$,

$$F\left(F^{[-1]}(t)\right) = t; \tag{12.13a}$$

2. se $t \notin \mathrm{Cod}(F)$ allora

$$F^{[-1]}(t) = \inf\{x\colon F(x) \geq t\} = \sup\{x\colon F(x) \leq t\}. \tag{12.13b}$$

Si noti che le v.a. U e V dell'*equazione (12.12)* hanno legge $\mathtt{U_c}(0,1)$; infatti

$$\mathbf{P}\{U \leq u\} = \mathbf{P}\{F_X(X) \leq u\} = \mathbf{P}\left\{X \leq F_X^{[-1]}(u)\right\} = F_X\left(F_X^{[-1]}(u)\right) = u \tag{12.14}$$

e lo stesso vale per la v.a. V (si veda anche l'*equazione (12.1)*).

Evidentemente, se F è continua e strettamente crescente allora la sua quasi-inversa $F^{[-1]}$ coincide con l'ordinaria funzione inversa F^{-1}. In conclusione, le v.a. U e V date dall'*equazione (12.12)* sono Uniformi su $\mathbb{I}$ e grazie alla *Proposizione (12.2)* il vettore (U, V) ha la stessa 2-copula di (X, Y), cioè $(U, V) \sim \mathbf{C}_{UV} = \mathbf{C}_{XY}$.

Lo studio delle distribuzioni condizionali risulta facilitato grazie all'utilizzo delle copule. Ad esempio, considerando le v.a. U e V Uniformi su $\mathbb{I}$ e aventi 2-copula $\mathbf{C}$, si ha:

$$\mathbf{P}\{U \leq u \mid V \leq v\} = \frac{\mathbf{P}\{U \leq u, V \leq v\}}{v} = \frac{\mathbf{C}(u,v)}{v} \in \mathbb{I}, \tag{12.15a}$$

$$\mathbf{P}\{U \leq u \mid V > v\} = \frac{\mathbf{P}\{U \leq u, V > v\}}{1-v} = \frac{u - \mathbf{C}(u,v)}{1-v} \in \mathbb{I}, \tag{12.15b}$$

$$\mathbf{P}\{U \leq u \mid V = v\} = \frac{\partial}{\partial v}\,\mathbf{C}(u,v) \in \mathbb{I}, \tag{12.15c}$$

ed espressioni simili valgono per le leggi condizionali di V data U. Si noti che:

1. data $u \in \mathbb{I}$, la derivata parziale $\frac{\partial \mathbf{C}}{\partial v}$ esiste per (Lebesgue) quasi ogni $v \in \mathbb{I}$;
2. dove $\frac{\partial \mathbf{C}}{\partial v}$ esiste, il suo codominio è quello di una funzione di ripartizione, cioè

$$0 \leq \frac{\partial}{\partial v}\mathbf{C}\left(u, v\right) \leq 1; \tag{12.16}$$

3. la funzione $u \mapsto \frac{\partial \mathbf{C}(u,v)}{\partial v}$ è ben definita e non decrescente (Lebesgue) quasi ovunque in $\mathbb{I}$.

Le formule date sopra rappresentano il punto di partenza per la simulazione di copule e quindi di distribuzioni multivariate (si vedano anche [21, 26]). Infatti, un algoritmo generale per generare osservazioni (x, y) estratte dalla coppia di v.a. (X, Y) con marginali F_X, F_Y, legge congiunta F_{XY}, e 2-copula $\mathbf{C}_{XY}$ è come segue.

Grazie al *Teorema (12.1)* è sufficiente generare una coppia di osservazioni (u, v) estratte dal vettore bivariato (U, V) di v.a. Uniformi su $\mathbb{I}$ aventi 2-copula $\mathbf{C} = \mathbf{C}_{UV} = \mathbf{C}_{XY}$. Quindi, usando la *Trasformazione Integrale di Probabilità*, il campione (x, y) si ottiene tramite la formula

$$\begin{cases} x = F_X^{[-1]}(u) \\ y = F_Y^{[-1]}(v) \end{cases} . \tag{12.17}$$

Al fine di generare (u, v) usiamo la legge condizionale di V dato l'evento $\{U = u\}$:

$$c_u(v) = \mathbf{P}\{V \leq v \mid U = u\} = \frac{\partial}{\partial u}\,\mathbf{C}\left(u, v\right) \tag{12.18}$$

che richiede esclusivamente l'uso della (derivata parziale della) 2-copula $\mathbf{C} = \mathbf{C}_{UV} = \mathbf{C}_{XY}$. Un algoritmo possibile per ottenere la coppia (u, v) desiderata è come segue.

1. `Si generino osservazioni indipendenti` u, w `Uniformi su` $\mathbb{I}$.
2. `Si calcoli` $v = c_u^{[-1]}(w)$.
3. `La coppia desiderata è` (u, v).

ESEMPIO 12.12. ▶

Le copule appartenenti alle famiglie di *Frank* e *Gumbel-Hougaard* sono di notevole interesse nelle applicazioni (ad esempio in idrologia [26]). Le copule di Frank possono modellizzare v.a. sia concordanti sia discordanti, mentre quelle di Gumbel-Hougaard solo v.a. concordanti (si veda la *Sezione 9.4*).

Le 2-copule della famiglia di Frank hanno la seguente espressione:

$$\mathbf{C}_\theta\left(u, v\right) = \frac{1}{\ln \theta}\,\ln\left[1 + \frac{(\theta^u - 1)(\theta^v - 1)}{\theta - 1}\right], \tag{12.19}$$

dove $u, v \in \mathbb{I}$ e $\theta \geq 0$ è un parametro di associazione. In particolare, (U, V) sono concordanti per $\theta < 1$, indipendenti per $\theta = 1$ (cioè $\mathbf{C}_1 = \mathbf{\Pi}_2$), e discordanti per $\theta > 1$. Inoltre, $\mathbf{C}_0 = \mathbf{M}_2$ e $\mathbf{C}_\infty = \mathbf{W}_2$, ovvero la famiglia di Frank è comprensiva.

Le 2-copule della famiglia di Gumbel-Hougaard hanno la seguente espressione:

$$\mathbf{C}_\theta\left(u, v\right) = e^{-\left[(-\ln u)^\theta + (-\ln v)^\theta\right]^{1/\theta}}, \tag{12.20}$$

dove $u, v \in \mathbb{I}$ e $\theta \geq 1$ è un parametro di associazione. In particolare, (U, V) sono indipendenti per $\theta = 1$ (cioè $\mathbf{C}_1 = \mathbf{\Pi}_2$), e concordanti per $\theta > 1$; inoltre $\mathbf{C}_\infty = \mathbf{M}_2$.

Nel seguito si mostrano alcuni confronti tra simulazioni di coppie (U, V) estratte da copule della famiglie di Frank e Gumbel-Hougaard. In tutti i casi l'ampiezza del campione è pari a $N = 200$, e il generatore aleatorio è lo stesso in ogni figura: ciò che cambia è il grado di associazione tra U e V, misurato dal valore della τ di Kendall (si veda la *Sezione 9.4*).

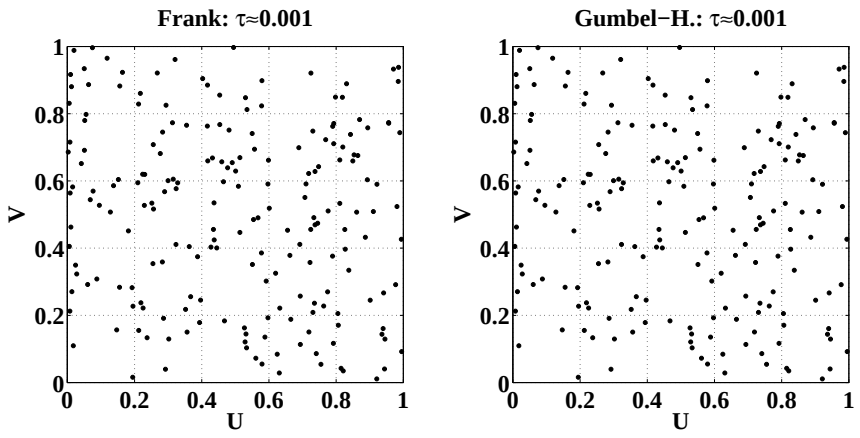

Figura 12.7: simulazione di campioni estratti dalle 2-copule di Frank e Gumbel-Hougaard per $\tau \approx 0.001$.

In *Figura (12.7)* si ha $\tau \approx 0.001$, cioè U e V sono debolmente concordanti (in pratica quasi indipendenti, cioè $\mathbf{C} \approx \mathbf{\Pi}_2$): in effetti i punti mostrano una distribuzione sostanzialmente uniforme nel quadrato unitario. I due campioni sono molto simili, con differenze minime nella disposizione dei punti.

In *Figura (12.8)* si ha $\tau \approx 0.5$, cioè U e V sono moderatamente concordanti: in effetti i punti tendono a disporsi lungo la diagonale principale. I due campioni sono molto simili: le differenze principali sono localizzate nell'angolo in alto a destra, ove il modello di Frank (avente coefficiente di *tail dependence* nullo [26]) organizza i punti in maniera differente rispetto a quello di Gumbel-Hougaard (che ha invece un coefficiente positivo).

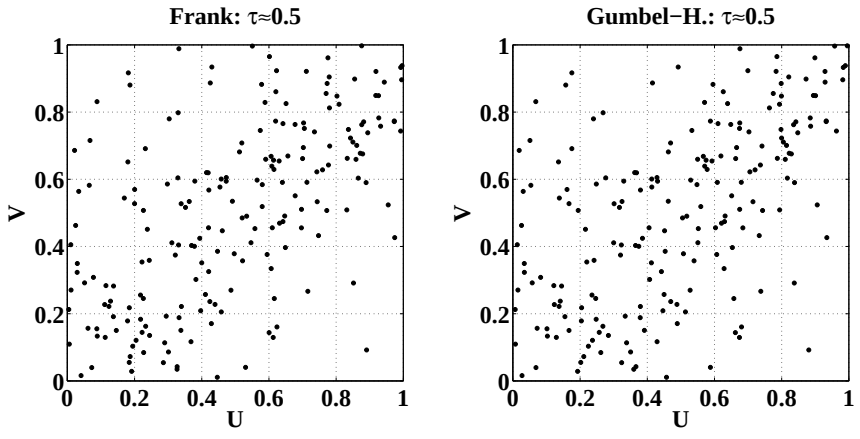

Figura 12.8: simulazione di campioni estratti dalle 2-copule di Frank e Gumbel-Hougaard per $\tau \approx 0.5$.

In *Figura (12.9)* si ha $\tau \approx 0.95$, cioè U e V sono fortemente concordanti: in effetti i punti sono praticamente disposti lungo la diagonale principale. I due campioni sono molto simili: le differenze principali sono localizzate nell'angolo in alto a destra, ove il modello di Gumbel-Hougaard sembra "aggregare" i punti in maniera maggiore rispetto a quello di Frank.

La *Figura (12.10)* mostra simulazioni estratte dalla copula di Frank, in cui le v.a. U e V sono discordanti: in effetti i valori della τ di Kendall sono negativi, e il grado di discordanza aumenta al decrescere di τ. In entrambi i casi i punti tendono a disporsi lungo la diagonale secondaria, e la loro "aggregazione" aumenta per $\tau \to -1$.

◀

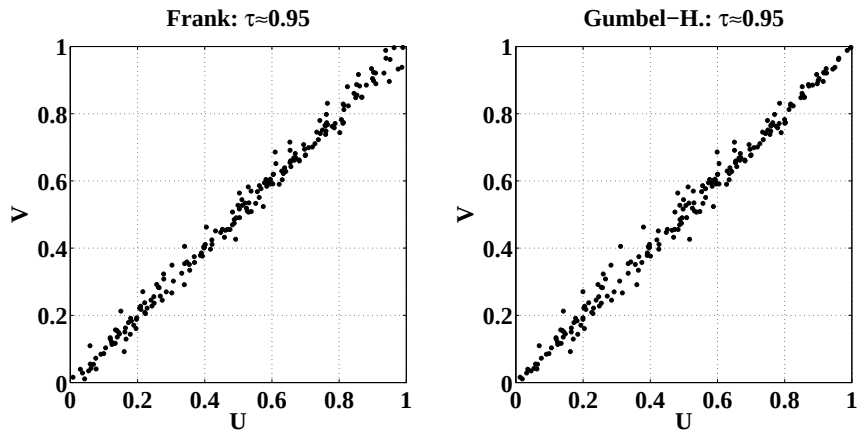

Figura 12.9: simulazione di campioni estratti dalle 2-copule di Frank e Gumbel-Hougaard per $\tau \approx 0.95$.

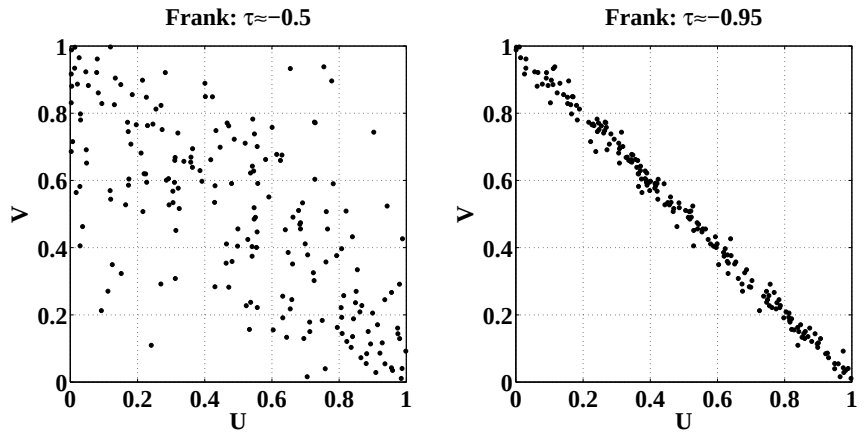

Figura 12.10: simulazione di campioni estratti dalla 2-copula di Frank per valori negativi della τ di Kendall.

In ambito multivariato, ovvero per dimensioni superiori a due, la modellizzazione attraverso le copule procede sostanzialmente lungo la falsariga dell'approccio bivariato, con alcune differenze che non è però possibile approfondire in questo Corso. A tale proposito si rimanda alla letteratura citata in Bibliografia [30, 15, 21, 26].

Capitolo 13

STATISTICA BAYESIANA

L'approccio Bayesiano alla statistica si basa sul convincimento che l'unico modo scientificamente fondato di fare statistica consista nell'apprendere dall'esperienza: chi investiga parte da certe conoscenze che traduce in una distribuzione di probabilità, osserva quindi i risultati dell'esperimento statistico e sulla base di questi modifica le proprie convinzioni, vale a dire la legge scelta inizialmente.

Nell'ambito della Statistica Matematica classica si presuppone che il parametro θ (scalare o vettoriale che sia), che caratterizza la distribuzione di probabilità in gioco, sebbene ignoto sia comunque una quantità fissa, ovvero deterministica e non aleatoria. L'approccio Bayesiano è fondamentalmente diverso ed è pervaso da una certa dose di soggettività e di arbitrarietà, in quanto si fonda sull'assunto che θ sia una v.a., la cui legge è inizialmente scelta (decisa) dall'analista in base a convincimenti propri o all'esperienza accumulata o, ancora, alla convenienza; in seguito, una volta osservato il fenomeno oggetto di studio, tale legge può essere modificata sulla base del campione raccolto. Ad esempio, considerando la percentuale θ di pezzi difettosi prodotti da una fabbrica, potrebbe essere utile supporre che tale parametro (incognito) non sia fisso, bensì possa subire variazioni aleatorie dovute al clima, all'abilità o alla condizione fisica dei diversi operai, e così via; ne segue che, ai fini della descrizione statistica del fenomeno, l'approccio Bayesiano risulti certamente più adatto e corretto di quello classico.

La *vexata quaestio* su quale dei due approcci sia preferibile è oggetto di un dibattito animato all'interno della comunità matematica internazionale. Probabilmente non esiste un modo di intendere (fare) la statistica migliore di un altro: alcuni problemi possono essere meglio risolti nell'ambito della statistica classica, altri sfruttando quella Bayesiana. Il punto importante di cui tenere conto è che i risultati possono essere assai diversi; inoltre, una soluzione Bayesiana può non essere ragionevole dal punto di vista classico e viceversa. Per ulteriori approfondimenti si vedano, ad esempio, [24, 6, 19, 5, 16].

13.1 Legge *a priori* e *a posteriori*

Come già accennato, l'approccio Bayesiano assume componenti di aleatorietà anche a livello dello spazio dei parametri Θ. Per questa ragione è necessario definire formalmente il *modello statistico Bayesiano*.

DEFINIZIONE 13.1 (Modello statistico Bayesiano)**.** Un modello statistico Bayesiano è composto di un modello statistico $\mathfrak{M} = \big(\Omega, \mathcal{F}, \{\mathbf{P}_\theta\}_{\theta\in\Theta}\big)$, dominato dalla misura σ-finita μ su $(\Omega, \mathcal{F})$, nel quale si suppone che lo spazio dei parametri Θ sia munito di una σ-algebra $\mathcal{T}$ di suoi sottoinsiemi. Inoltre, si assume che, per ogni $F \in \mathcal{F}$, l'applicazione $\theta \mapsto \mathbf{P}_\theta\{F\}$ sia misurabile rispetto alla σ-algebra $\mathcal{T}$.

I concetti di *legge a priori* e di *nucleo stocastico* sono fondamentali nella statistica Bayesiana.

DEFINIZIONE 13.2 (Legge *a priori*)**.** Sia π un'arbitraria misura di probabilità definita su $(\Theta, \mathcal{T})$; allora π è detta *legge a priori*.

DEFINIZIONE 13.3 (Nucleo stocastico)**.** Dati due spazi misurabili $(R, \mathcal{R})$ e $(S, \mathcal{S})$ si dice *nucleo stocastico*, o *probabilità di transizione* dallo spazio $(R, \mathcal{R})$ allo spazio $(S, \mathcal{S})$, ogni funzione $\mathbf{N}\colon R \times \mathcal{S} \to [0, 1]$ tale che:

1. per ogni $A \in \mathcal{S}$ la funzione definita in R da

$$x \mapsto \mathbf{N}(x, A)$$

sia misurabile rispetto a $\mathcal{R}$;

2. per ogni $x \in R$ la funzione definita in $\mathcal{S}$ da

$$A \mapsto \mathbf{N}(x, A)$$

sia una misura di probabilità su $(S, \mathcal{S})$.

OSSERVAZIONE 13.1.

In un modello statistico Bayesiano la funzione

$$\Theta \times \mathcal{F} \ni (\theta, F) \mapsto \mathbf{P}_\theta\{F\} \in [0, 1]$$

è una probabilità di transizione da $(\Theta, \mathcal{T})$ a $(\Omega, \mathcal{F})$.

Per un nucleo stocastico vale la seguente estensione del *Teorema di Fubini*.

TEOREMA 13.1. *Siano $(R, \mathcal{R})$ e $(S, \mathcal{S})$ due spazi misurabili, $\mathbf{N}$ una probabilità di transizione da $(R, \mathcal{R})$ a $(S, \mathcal{S})$, e π una misura di probabilità su $(R, \mathcal{R})$. Allora:*

1. *per ogni funzione $f\colon R \times S \to \mathbb{R}$ che sia limitata e misurabile rispetto alla σ-algebra prodotto $\mathcal{R} \otimes \mathcal{S}$, la funzione definita in R da*

$$x \mapsto \int_S f(x, y)\, \mathbf{N}(x, \mathrm{d}y)$$

è misurabile rispetto a $\mathcal{R}$;

2. *esiste un'unica misura di probabilità $\mathbf{Q}$ su $(R \times S, \mathcal{R} \otimes \mathcal{S})$ tale che, per ogni funzione limitata $f\colon R \times S \to \mathbb{R}$ che sia misurabile rispetto a $\mathcal{R} \otimes \mathcal{S}$, valga*

$$\int_{R \times S} f \, \mathrm{d}\mathbf{Q} = \int_R \mathrm{d}\pi(x) \int_S f(x, y)\, \mathbf{N}(x, \mathrm{d}y); \tag{13.1}$$

3. *l'equazione (13.1) vale anche se la funzione f è positiva e non necessariamente limitata;*

4. *la funzione $f\colon R \times S \to \mathbb{R}$ è integrabile rispetto a $\mathbf{Q}$, cioè $f \in \mathrm{L}^1(R \times S, \mathcal{R} \otimes \mathcal{S}, \mathbf{Q})$, se e solo se vale*

$$\int_{R \times S} |f| \, \mathrm{d}\mathbf{Q} = \int_R \mathrm{d}\pi(x) \int_S |f(x, y)| \, \mathbf{N}(x, \mathrm{d}y) < \infty; \tag{13.2}$$

in quest'ultimo caso vale anche l'equazione (13.1).

OSSERVAZIONE 13.2.

Dal *Teorema (13.1)* scende che

$$\mathbf{Q}\{D\} = \int_R \mathbf{N}(x, D_x) \, \mathrm{d}\pi(x), \tag{13.3}$$

dove $D \in \mathcal{R} \otimes \mathcal{S}$ e D_x indica la sezione di D in $x \in R$, definisce una misura di probabilità $\mathbf{Q}$ sullo spazio misurabile $(R \times S, \mathcal{R} \otimes \mathcal{S})$.

Dunque, nel modello statistico Bayesiano possiamo ora introdurre la misura di probabilità $\mathbf{Q}$ sullo spazio misurabile $(\Theta \times \Omega, \mathcal{T} \otimes \mathcal{F})$, la cui esistenza è assicurata dal *Teorema (13.1)*. Se π è la legge a priori su $(\Theta, \mathcal{T})$, l'*equazione (13.3)* mostra che, per ogni rettangolo misurabile $T \times F$, con $T \in \mathcal{T}$ e $F \in \mathcal{F}$, si ha

$$\mathbf{Q}\{T \times F\} = \int_T \mathbf{P}_\theta\{F\} \, \mathrm{d}\pi(\theta). \tag{13.4}$$

Si definiscano ora le σ-algebre di $\Theta \times \Omega$ date da

$$\tilde{\mathcal{T}} = \{T \times \Omega\colon T \in \mathcal{T}\} \quad \text{e} \quad \tilde{\mathcal{F}} = \{\Theta \times F\colon F \in \mathcal{F}\}.$$

Vale il seguente risultato.

LEMMA 13.1. *Per una funzione $X\colon \Theta \times \Omega \to \mathbb{R}$ sono equivalenti le affermazioni:*

1. *X è $\tilde{\mathcal{F}}$-misurabile, cioè $X \in \mathrm{L}^0\left(\tilde{\mathcal{F}}\right)$;*
2. *X è una funzione della sola variabile $\omega \in \Omega$ e, come funzione di ω, è $\mathcal{F}$-misurabile.*

La *speranza condizionale* gioca un ruolo fondamentale nella statistico Bayesiano (si veda la *Definizione (0.9)*). Il Teorema che segue consente di interpretare la probabilità $\mathbf{P}_\theta$ definita sullo spazio misurabile $(\Omega, \mathcal{F})$ come una *legge condizionata* dalla conoscenza del valore del parametro θ.

TEOREMA 13.2. *Nello spazio misurabile $(\Theta \times \Omega, \mathcal{T} \otimes \mathcal{F})$ si consideri una v.a. integrabile $X\colon \Theta \times \Omega \to \mathbb{R}$. Allora*

$$\mathbf{E}_{\mathbf{Q}}\left(X \mid \tilde{\mathcal{T}}\right)(\theta) = \int_\Omega X(\theta, \omega)\ \mathrm{d}\mathbf{P}_\theta(\omega) \quad \mathbf{Q}\text{-}q.c., \tag{13.5}$$

ove la speranza condizionale è calcolata rispetto alla misura di probabilità $\mathbf{Q}$ del Teorema (13.1).

ESEMPIO 13.1. ▶
Si consideri il caso particolare del parametro $\tilde{\theta} \in \Theta$ tale che $\left\{\tilde{\theta}\right\} \in \mathcal{T}$ e $\pi\left(\left\{\tilde{\theta}\right\}\right) > 0$. Allora, per ogni $F \in \mathcal{F}$ vale

$$\mathbf{P}_{\tilde{\theta}}\{F\} = \mathbf{Q}\left\{\Theta \times F \mid \left\{\tilde{\theta}\right\} \times \Omega\right\}. \tag{13.6}$$

Infatti, per definizione si ha

$$\mathbf{Q}\left\{\Theta \times F \mid \left\{\tilde{\theta}\right\} \times \Omega\right\} = \frac{\mathbf{Q}\left\{\left\{\tilde{\theta}\right\} \times F\right\}}{\mathbf{Q}\left\{\left\{\tilde{\theta}\right\} \times \Omega\right\}}. \tag{13.7}$$

Ora

$$\mathbf{Q}\left\{\left\{\tilde{\theta}\right\} \times F\right\} = \int_{\left\{\tilde{\theta}\right\}} \mathbf{P}_\theta\{F\}\ \mathrm{d}\pi(\theta) = \mathbf{P}_{\tilde{\theta}}\{F\}\ \pi\left(\left\{\tilde{\theta}\right\}\right),$$

mentre

$$\mathbf{Q}\left\{\left\{\tilde{\theta}\right\} \times \Omega\right\} = \mathbf{P}_{\tilde{\theta}}\{\Omega\}\ \pi\left(\left\{\tilde{\theta}\right\}\right),$$

che sostituite nell'*equazione (13.7)* danno l'*equazione (13.6)*. ◀

Come già si è accennato, lo statistico Bayesiano parte da una legge a priori π su $(\Theta, \mathcal{T})$, osserva il risultato dell'esperimento $\omega \in \Omega$, e si serve di quest'ultimo per "aggiornare" la sua conoscenza del parametro θ, ovvero per adottare una nuova misura di probabilità (detta *a posteriori*) che tenga conto dell'osservazione ω.

DEFINIZIONE 13.4 (Legge *a posteriori*)**.** Se esiste una probabilità di transizione $\mathbf{N}$ dallo spazio misurabile $(\Omega, \mathcal{F})$ allo spazio misurabile $(\Theta, \mathcal{T})$ tale che, per ogni v.a. limitata $X\colon \Theta \times \Omega \to \mathbb{R}$, sia

$$\mathbf{E}_{\mathbf{Q}}\left(X \mid \tilde{\mathcal{F}}\right)(\omega) = \int_\Theta X(\theta, \omega)\, \mathbf{N}(\omega, \mathrm{d}\theta), \tag{13.8}$$

ove $\mathbf{Q}$ è la misura di probabilità del *Teorema (13.1)*, allora $\mathbf{N}(\omega, \cdot)$ si dice *legge a posteriori* su $(\Theta, \mathcal{T})$ condizionata da $\omega \in \Omega$.

La Definizione appena data pone il problema dell'esistenza della legge a posteriori; tuttavia è possibile dimostrare che in un modello statistico dominato questa esiste sempre. Sia dunque $\mathfrak{M}$ un modello statistico dominato dalla misura σ-finita μ, e supponiamo che esista una versione della verosimiglianza $\mathbf{L}_\theta = \frac{\mathrm{d}\mathbf{P}_\theta}{\mathrm{d}\mu}$ che sia misurabile rispetto alla σ-algebra prodotto $\mathcal{T} \otimes \mathcal{F}$. Questa condizione (che si supporrà sempre verificata nel seguito) assicura, in virtù del *Teorema (13.1)*, che la funzione $\theta \mapsto \mathbf{P}_\theta\{F\}$ sia misurabile rispetto alla σ-algebra $\mathcal{T}$ per ogni $F \in \mathcal{F}$.

TEOREMA 13.3. *Nelle condizioni appena specificate, se π è la legge a priori su $(\Theta, \mathcal{T})$, si definisca la funzione*

$$\varphi(\theta, \omega) = \frac{\mathbf{L}(\theta, \omega)}{\int_\Theta \mathbf{L}(t, \omega)\ \mathrm{d}\pi(t)}\, \mathbb{I}_{D^{\complement}}(\omega) + \mathbb{I}_D(\omega),$$

dove si è posto

$$D = \left\{\omega \in \Omega\colon \int_\Theta \mathbf{L}(t, \omega)\ \mathrm{d}\pi(t) = 0\right\}.$$

Allora:

1. *l'insieme $D \in \mathcal{F}$ è trascurabile, ovvero $\mathbf{Q}\{\Theta \times D\} = 0$;*
2. *per quasi ogni $\omega \in \Omega$, rispetto a $\mathbf{Q}$, la funzione $\theta \mapsto \varphi(\theta, \omega)$ è una densità di probabilità sullo spazio $(\Theta, \mathcal{T})$;*
3. *la funzione $\mathbf{N}\colon \Omega \times \mathcal{T} \to \mathbb{R}_+$ data da*

$$\mathbf{N}(\omega, T) = \int_T \varphi(\theta, \omega)\ \mathrm{d}\pi(\theta), \tag{13.9}$$

dove $T \in \mathcal{T}$, è la legge a posteriori condizionata da $\omega \in \Omega$.

Nel seguito indicheremo con π_ω la legge a posteriori su $(\Theta, \mathcal{T})$ condizionata da $\omega \in \Omega$. Tale legge è *assolutamente continua* rispetto alla legge a priori π, e ha densità data da

$$\frac{\mathrm{d}\pi_\omega}{\mathrm{d}\pi}(\theta) = \frac{\mathbf{L}(\theta, \omega)}{\int_\Theta \mathbf{L}(t, \omega)\ \mathrm{d}\pi(t)} \quad \mathbf{Q}\text{-q.c.}. \tag{13.10}$$

Spesso nelle applicazioni la legge a priori è a sua volta assolutamente continua rispetto ad una misura σ-finita ψ su $(\Theta, \mathcal{T})$, ovvero $\pi = \rho \cdot \psi$, dove ρ è la densità di π rispetto a ψ. In tali condizioni la legge a posteriori π_ω è anch'essa assolutamente continua rispetto a ψ e la sua densità vale

$$\frac{\mathrm{d}\pi_\omega}{\mathrm{d}\psi}(\theta) = \frac{\mathbf{L}(\theta, \omega)\, \rho(\theta)}{\int_\Theta \mathbf{L}(t, \omega)\, \rho(t)\ \mathrm{d}\psi(t)} \quad \mathbf{Q}\text{-q.c.}. \tag{13.11}$$

Gli Esempi che seguono illustrano i concetti introdotti finora.

ESEMPIO 13.2 (Bayes, 1763)**.** ▶

Un punto è posizionato "a caso" sul segmento $(0, 1)$; sia θ la sua ascissa. Di seguito, altri n punti vengono posti a caso sullo stesso segmento; sia X_i l'ascissa del punto i-esimo e sia S_n la v.a. che conta quanti degli n punti appartengono all'intervallo $(0, \theta)$, ovvero $S_n = \sum_{i=1}^n Y_i$, con $Y_i = \mathbb{I}_{(0,\theta)}(X_i)$, $i = 1, \ldots, n$.

Il modello statistico Bayesiano soggiacente può essere scelto nel seguente modo: $\Omega = \{0, \ldots, n\}$, $\mathcal{F} = \mathcal{P}(\Omega)$, $\Theta = (0, 1)$, $\mathcal{T} = \mathcal{B}(\Theta)$, e quale legge a priori su $(\Theta, \mathcal{T})$ è naturale prendere $\pi \sim \mathtt{U_c}(0, 1)$. Evidentemente, fissato $\theta \in \Theta$, si ha $S_n \sim \mathtt{Bi}(n, \theta)$, e perciò

$$\mathbf{P}_\theta\{S_n = k\} = \binom{n}{k} \theta^k\, (1 - \theta)^{n-k},$$

con $k = 0, \ldots, n$. Ora, il problema dello statistico Bayesiano è il seguente: avendo osservato $S_n = s$, cosa si può inferire su θ (che, ricordiamo, è una v.a.)? La densità della legge a posteriori π_s rispetto alla legge a priori π vale

$$\pi_s(\theta) = \frac{\binom{n}{s}\, \theta^s\, (1 - \theta)^{n-s}}{\int_0^1 \binom{n}{s}\, t^s\, (1 - t)^{n-s}\ \mathrm{d}t} \mathbb{I}_{(0,1)}(\theta) = \frac{\theta^s\, (1 - \theta)^{n-s}}{\mathrm{B}(s + 1, n - s + 1)} \mathbb{I}_{(0,1)}(\theta).$$

Dunque, in virtù del risultato $S_n = s$ dell'esperimento, si è passati dalla legge a priori $\pi \sim \mathtt{U_c}(0, 1)$, avente speranza pari a $1/2$, a quella a posteriori $\pi_s \sim \beta(s + 1, n - s + 1)$ sullo stesso spazio, avente speranza pari a $(s + 1)/(n + 2) \in (0, 1)$. ◀

NOTA 13.1.

Evidentemente, volendo stimare il parametro θ, la legge a posteriori π_s in generale fornisce più "informazioni" rispetto alla legge a priori π Uniforme sul segmento. Vedremo nel seguito come affrontare il problema della stima da un punto di vista Bayesiano. Si osservi che, nell'ambito della statistica classica, sarebbe $\hat{\theta} = \overline{Y}_n$, avente valore $s/n \in [0, 1]$. Evidentemente, per $s = 0$ si ottiene $\hat{\theta} = 0$ e per $s = n$ si ottiene $\hat{\theta} = 1$, due stime non consistenti con l'ipotesi $\theta \in (0, 1)$.

ESEMPIO 13.3 (Laplace, 1773)**.** ▶

Un'urna contiene n palline, delle quali si sa che $r = \theta n$ sono rosse (R) e le rimanenti $g = (1 - \theta)n$ sono gialle (G). Sapendo che l'urna contiene almeno una pallina rossa e una gialla, il problema dello statistico Bayesiano è il seguente: avendo estratto una pallina di colore X, cosa si può inferire su θ (che, ricordiamo, è una v.a.)?

Il modello statistico Bayesiano soggiacente può essere scelto nel seguente modo: $\Omega = \{\mathrm{G}(=0), \mathrm{R}(=1)\}$, $\mathcal{F} = \mathcal{P}(\Omega)$, $\Theta = \left\{\frac{1}{n}, \ldots, \frac{n-1}{n}\right\}$ essendo $\theta \neq 0, 1$ per ipotesi, $\mathcal{T} = \mathcal{P}(\Theta)$, e quale legge a priori π su $(\Theta, \mathcal{T})$ è naturale prendere quella Uniforme su Θ, cioè

$$\pi\left(\frac{k}{n}\right) = \frac{1}{n-1}, \quad k = 1, \ldots, n-1.$$

Evidentemente, fissato $\theta \in \Theta$, si ha $\mathbf{P}_\theta\{X = \mathrm{R}(=1)\} = \theta$, ovvero $X \sim \mathtt{Be}(\theta)$. Ora, se ad esempio si è effettivamente estratta una pallina rossa, la densità della legge a posteriori π_{R} rispetto alla legge a priori π vale

$$\pi_{\mathrm{R}}(\theta) = \frac{\theta \frac{1}{n-1}}{\sum_{i=1}^{n-1} \frac{i}{n} \frac{1}{n-1}} \mathbb{I}_\Theta(\theta) = \frac{2\theta}{n-1} \mathbb{I}_\Theta(\theta).$$

Dunque, in virtù del risultato $X = \mathrm{R}(=1)$ dell'esperimento, si è passati dalla legge a priori π Uniforme su Θ a quella a posteriori π_{R} sullo stesso spazio, avente speranza pari a $(2n-1)/(3n) \in [\frac{1}{2}, \frac{2}{3}]$, essendo $n > 1$. ——◀

NOTA 13.2.

Anche in questo caso, in generale la legge a posteriori π_{R} risulta essere più "informativa" rispetto alla legge a priori π. Si osservi che, nell'ambito della statistica classica, uno stimatore non distorto di θ è dato da $\hat{\theta} = \overline{X}_n = X$; tuttavia, essendo $X = 0$ o $X = 1$, la stima fornita non è consistente con l'ipotesi $\theta \neq 0, 1$.

13.2 Decisioni Bayesiane

La *Teoria delle decisioni* è alla base di (quasi) tutta la Statistica Matematica: in fondo, dopo aver osservato un campione, lo statistico prende una decisione e compie un'azione, ad esempio accettando o rifiutando un'ipotesi. Nel seguito indicheremo con $\mathfrak{A}$ l'insieme delle *azioni*, che supporremo dotato di una σ-algebra $\mathcal{A}$, e con $\mathfrak{D}$ l'insieme delle *decisioni*, che è strettamente legato a quello delle azioni.

DEFINIZIONE 13.5 (Decisione)**.** Una *decisione* $\delta \in \mathfrak{D}$ è una funzione (*Borel-*)misurabile $\delta\colon \mathbb{R}^n \to \mathfrak{A}$.

NOTA 13.3.

Più in generale, una decisione è una funzione $\delta\colon \Omega \to \mathfrak{A}$ che sia $\mathcal{F}$-misurabile.

Dunque una decisione non è altro che una *regola*, che specifica quale azione intraprendere in seguito all'osservazione: infatti, in base al risultato $\boldsymbol{x}$ dell'esperimento, lo statistico prende la decisione δ e compie l'azione $a = \delta(\boldsymbol{x})$. Ad esempio, sia $\mathfrak{A} = \{a_0, a_1\}$; attraverso la decisione δ si induce una partizione su $\mathbb{R}^n$ (o Ω), che risulta diviso nella regione C tale per cui $\delta(\boldsymbol{x}) = a_0$ se $\boldsymbol{x} \in C$, e nella regione $C^{\complement}$ tale per cui $\delta(\boldsymbol{x}) = a_1$ se $\boldsymbol{x} \in C^{\complement}$: evidentemente si tratta di un problema di verifica di ipotesi (deterministico). Ancora, sia $\mathfrak{A} = \Theta$; in questo caso si tratta di un problema di stima. Ogni decisione ha inevitabilmente delle conseguenze, le quali comportano dei costi; al fine di misurare le perdite si introduce una *funzione costo*.

DEFINIZIONE 13.6 (Funzione costo)**.** Una funzione non negativa $C\colon \Theta \times \mathfrak{A} \to \mathbb{R}_+$ tale che, per ogni $a \in \mathfrak{A}$, la funzione $\theta \mapsto C(\theta, a)$ sia $\mathcal{T}$-misurabile, è detta *funzione costo* (o *di perdita*).

Dunque, $C(\theta, a)$ rappresenta la perdita a cui lo statistico va incontro compiendo l'azione a quando il parametro è θ; in particolare, se si prende la decisione $\delta(\boldsymbol{X})$ — e si compie l'azione $a = \delta(\boldsymbol{X})$ — il costo è dato dalla v.a. $C(\theta, \delta(\boldsymbol{X}))$, ed è possibile introdurre il concetto di (funzione di) *rischio*.

DEFINIZIONE 13.7 (Rischio)**.** La quantità

$$\mathbf{R}(\theta, \delta) = \mathbf{E}_\theta(C(\theta, \delta(\boldsymbol{X}))) = \int_\Omega C(\theta, \delta(\omega)) \, \mathrm{d}\mathbf{P}_\theta(\omega) \tag{13.12}$$

è detta (funzione di) *rischio* della decisione δ.

Evidentemente, fissata la regola di decisione δ, $\mathbf{R}(\theta, \delta)$ è una funzione del parametro θ.

ESEMPIO 13.4. ►

Sia $\mathfrak{A} = \Theta \subseteq \mathbb{R}$ e sia $C(\theta, a) = |\theta - a|^2$. Allora per la decisione $\delta\colon \mathbb{R}^n \to \mathfrak{A} = \Theta$ vale

$$\mathbf{R}(\theta, \delta) = \mathbf{E}_\theta(C(\theta, \delta(\boldsymbol{X}))) = \mathbf{E}_\theta\Big(|\theta - \delta(\boldsymbol{X})|^2\Big),$$

che rappresenta il rischio usuale dello stimatore $\delta(\boldsymbol{X})$; in particolare, se $\delta(\boldsymbol{X})$ è corretto allora il rischio coincide con la varianza dello stimatore. ◄

La funzione di rischio è usata nell'ambito della *Teoria delle Decisioni* quale strumento per confrontare regole di decisione differenti: in particolare, date le decisioni δ_1 e δ_2, se $\mathbf{R}(\theta, \delta_1) < \mathbf{R}(\theta, \delta_2)$ per ogni $\theta \in \Theta$, allora si preferisce la decisione δ_1. Purtroppo, non sempre tale disuguaglianza è rispettata per ogni valore del parametro, anzi, tipicamente, le due funzioni di rischio si intersecano e perciò non risultano confrontabili. Tuttavia, nella statistica Bayesiana il parametro θ è una v.a., avente distribuzione pari alla legge a priori π, ciò che consente di introdurre il concetto di *rischio Bayesiano*, attraverso il quale è possibile procedere al confronto di regole di decisione differenti.

DEFINIZIONE 13.8 (Rischio Bayesiano)**.** Si dice *rischio Bayesiano* della decisione δ, rispetto alla legge a priori π, il numero

$$\mathbf{B}(\pi, \delta) = \mathbf{E}_\pi(\mathbf{R}(\theta, \delta)) = \mathbf{E}_\pi(\mathbf{E}_\theta(C(\theta, \delta(\boldsymbol{X})))) = \int_\Theta \int_\Omega C(\theta, \delta(\omega))\ \mathrm{d}\mathbf{P}_\theta(\omega)\ \mathrm{d}\pi(\theta). \tag{13.13}$$

Dunque, il rischio Bayesiano riduce la funzione di rischio della decisione δ ad un singolo numero reale, tramite il quale è possibile operare confronti: chiaramente, tra due diverse decisioni è da preferire quella con rischio Bayesiano minore.

DEFINIZIONE 13.9 (Decisione Bayesiana)**.** Si chiama *decisione Bayesiana*, rispetto alla legge a priori π, la regola $\tilde{\delta}$ tale che

$$\mathbf{B}(\pi, \tilde{\delta}) = \inf_{\delta \in \mathfrak{D}} \mathbf{B}(\pi, \delta). \tag{13.14}$$

NOTA 13.4.

A priori non è detto che $\tilde{\delta}$ esista, né, in tal caso, che sia unica.

Il Teorema che segue è fondamentale.

TEOREMA 13.4 (Wald)**.** *Dato un modello statistico $\mathfrak{M} = \big(\Omega, \mathcal{F}, \{\mathbf{P}_\theta\}_{\theta \in \Theta}\big)$ dominato dalla misura σ-finita μ, sia π la legge a priori su $(\Theta, \mathcal{T})$, e si supponga che per ogni $\omega \in \Omega$ esista la legge a posteriori π_ω condizionata da ω. Si definisca la funzione*

$$r(\omega, a) = \int_\Theta C(\theta, a)\ \mathrm{d}\pi_\omega(\theta), \tag{13.15a}$$

con $\omega \in \Omega$ e $a \in \mathfrak{A}$, e si supponga che per ogni $\omega \in \Omega$ esista una azione $a_\omega \in \mathfrak{A}$ tale che

$$r(\omega, a_\omega) = \inf_{a \in \mathfrak{A}} r(\omega, a). \tag{13.15b}$$

Se la regola di decisione $\omega \mapsto \tilde{\delta}$ data da $\tilde{\delta} = a_\omega$ è $\mathcal{F}$-misurabile, allora per ogni decisione $\delta \in \mathfrak{D}$ vale la disuguaglianza

$$\mathbf{B}(\pi, \tilde{\delta}) \leq \mathbf{B}(\pi, \delta). \tag{13.16}$$

I due risultati che seguono mostrano come nei casi che usualmente si incontrano nelle applicazioni la decisione Bayesiana $\tilde{\delta}$ del *Teorema (13.4)* sia ammissibile.

TEOREMA 13.5. *Se lo spazio dei parametri è (al più) numerabile, ossia $\Theta = \cup_{n \in \mathbb{N}} \{\theta_n\}$, con $\pi(\{\theta_n\}) > 0$ per ogni indice n, allora la decisione Bayesiana $\tilde{\delta}$ è ammissibile.*

TEOREMA 13.6. *Sia Θ un aperto di $\mathbb{R}^d$ e la sua chiusura $\overline{\Theta}$ sia il supporto della legge a priori π, che si suppone assolutamente continua rispetto alla misura di Lebesgue d-dimensionale λ_d; si supponga inoltre che, per ogni regola di decisione δ, la funzione $\theta \mapsto \mathbf{R}(\theta, \delta)$ sia continua. Allora la decisione Bayesiana $\tilde{\delta}$ è ammissibile.*

13.3 Stima Bayesiana puntuale

In ambito Bayesiano la stima della funzione misurabile $\xi\colon \Theta \to \mathbb{R}$ è subordinata alla scelta sia della legge a priori π sia della funzione costo C. In particolare, le funzioni di perdita comunemente utilizzate nelle applicazioni sono:

1. $C(\theta, a) = |\xi(\theta) - a|^2$, ovvero la funzione di perdita *quadratica*;
2. $C(\theta, a) = |\xi(\theta) - a|$, ovvero la funzione di perdita *assoluta*.

Tuttavia, numerose altre scelte sono possibili, come ad esempio:

- $C(\theta, a) = \begin{cases} |\xi(\theta) - a|^2, & \text{se } a < \xi(\theta) \\ c\,|\xi(\theta) - a|^2, & \text{se } a \geq \xi(\theta) \end{cases}$, con $c > 1$;
- $C(\theta, a) = \frac{|\xi(\theta)-a|^2}{|\xi(\theta)|+1}$;
- $C(\theta, a) = \frac{|\xi(\theta)-a|^2}{|\xi(\theta)|}$ se $\xi(\theta) \neq 0$ per ogni $\theta \in \Theta$.
- $C(\theta, a) = \sqrt{\frac{|\xi(\theta)-a|}{|\xi(\theta)|+1}}$;

Nel seguito quale funzione costo C adotteremo la funzione di perdita quadratica; inoltre sarà $\mathfrak{A} = \mathbb{R}$.

PROPOSIZIONE 13.1. *Se ν è una probabilità su $(\Theta, \mathcal{T})$ si ha*

$$\int_\Theta \left|\xi(\theta) - \tilde{\delta}\right|^2 \,\mathrm{d}\nu(\theta) = \min_{a\in\mathbb{R}} \int_\Theta |\xi(\theta) - a|^2 \,\mathrm{d}\nu(\theta),$$

ove la decisione $\tilde{\delta}$ è data da

$$\tilde{\delta} = \mathbf{E}_\nu(\xi) = \int_\Theta \xi(\theta) \,\mathrm{d}\nu(\theta). \tag{13.17}$$

DIM. Usando un procedimento classico si vede facilmente che

$$\begin{aligned}
\int_\Theta C(\theta, a) \,\mathrm{d}\nu(\theta) &= \int_\Theta (\xi(\theta) - a)^2 \,\mathrm{d}\nu(\theta) \\
&= \mathbf{E}_\nu\left(\left(\xi(\theta) - \tilde{\delta} + \tilde{\delta} - a\right)^2\right) \\
&= \mathbf{E}_\nu\left(\left(\xi(\theta) - \tilde{\delta}\right)^2\right) + \left(\tilde{\delta} - a\right)^2 + 2\left(\tilde{\delta} - a\right)\mathbf{E}_\nu\left(\xi(\theta) - \tilde{\delta}\right) \\
&= \mathbf{E}_\nu\left(\left(\xi(\theta) - \tilde{\delta}\right)^2\right) + \left(\tilde{\delta} - a\right)^2 \\
&\geq \mathbf{E}_\nu\left(\left(\xi(\theta) - \tilde{\delta}\right)^2\right).
\end{aligned}$$

Evidentemente il segno di uguaglianza vale se, e solo se, $a = \tilde{\delta}$. □

Grazie alla *Proposizione (13.1)* è possibile individuare la decisione Bayesiana $\tilde{\delta}$, che evidentemente vale $\tilde{\delta} = \mathbf{E}_{\pi_\omega}(\xi)$ in base alle *equazioni (13.15)*. Per le ipotesi fatte, nel caso presente le decisioni coincidono con gli stimatori, ed è dunque possibile calcolare lo stimatore Bayesiano della funzione ξ.

PROPOSIZIONE 13.2. *Se la funzione costo C è quella di perdita quadratica, lo stimatore Bayesiano della funzione ξ è dato da*

$$T(\omega) = \mathbf{E}_{\pi_\omega}(\xi) = \int_\Theta \xi(\theta) \,\mathrm{d}\pi_\omega(\theta) = \frac{\int_\Theta \xi(\theta)\,\mathbf{L}(\theta, \omega) \,\mathrm{d}\pi(\theta)}{\int_\Theta \mathbf{L}(t, \omega) \,\mathrm{d}\pi(t)}. \tag{13.18}$$

NOTA 13.5.

In generale è necessario accertarsi che $T(\omega)$ sia un valore "ammissibile". Ad esempio, se $\xi(\theta) = \theta$ — ovvero se si cerca di stimare il parametro θ — deve essere $T(\omega) \in \Theta$.

Alla luce del *Teorema (13.2)* e dell'*equazione (13.8)* si ha

$$T(\omega) = \mathbf{E}_{\mathbf{Q}}\left(\xi \mid \tilde{\mathcal{F}}\right)(\omega) \quad \mathbf{Q}\text{-q.c..}$$

OSSERVAZIONE 13.3.

Procedendo come nel caso della *Proposizione (13.2)* è possibile dimostrare che, se la funzione costo C è quella di perdita assoluta, lo stimatore Bayesiano della funzione $\xi(\theta) = \theta$ è dato dalla *mediana* della legge a posteriori π_ω.

Il Teorema che segue mostra che gli stimatori Bayesiani sono in genere *non corretti*.

TEOREMA 13.7. *Sia T uno stimatore Bayesiano non distorto di quadrato sommabile della funzione limitata e misurabile $\xi\colon \Theta \to \mathbb{R}$. Allora*

$$\int_{\Theta\times\Omega} |T(\omega) - \xi(\theta)|^2 \, \mathrm{d}\mathbf{Q} = 0.$$

OSSERVAZIONE 13.4.

Si noti che

$$\int_{\Theta\times\Omega} |T(\omega) - \xi(\theta)|^2 \, \mathrm{d}\mathbf{Q} = \int_{\Theta} \mathrm{d}\pi(\theta) \int_{\Omega} |T(\omega) - \xi(\theta)|^2 \, \mathrm{d}\mathbf{P}_\theta(\omega)$$

e, quindi, se T fosse corretto, allora per π-quasi ogni θ l'insieme $\{\omega \in \Omega\colon T(\omega) \neq \xi(\theta)\}$ sarebbe trascurabile rispetto alla misura $\mathbf{P}_\theta$. Pertanto, se le probabilità $\mathbf{P}_\theta$ sono *equivalenti* (cioè se per ogni coppia $\theta_1, \theta_2 \in \Theta$ si ha $\mathbf{P}_{\theta_1} \ll \mathbf{P}_{\theta_2}$ e $\mathbf{P}_{\theta_2} \ll \mathbf{P}_{\theta_1}$, ovvero in un modello regolare), uno stimatore Bayesiano non può essere corretto, tranne nel caso (banale) in cui ξ sia costante. Ciò accade, ad esempio, nei due casi importanti appena visti:

- se lo spazio dei parametri Θ è (al più) numerabile, ossia se $\Theta = \cup_{n\in\mathbb{N}} \{\theta_n\}$, con $\pi(\{\theta_n\}) > 0$ per ogni indice n;
- se lo spazio dei parametri Θ è un aperto di $\mathbb{R}^d$, la sua chiusura $\overline{\Theta}$ è il supporto della legge a priori π e la funzione ξ è continua.

Se dal punto di vista della correttezza gli stimatori Bayesiani non offrono prestazioni particolari, tuttavia essi mostrano proprietà interessanti per quanto concerne il loro rischio, come mostra il Teorema che segue.

TEOREMA 13.8. *Se il modello statistico $\mathfrak{M} = \left(\Omega, \mathcal{F}, \{\mathbf{P}_\theta\}_{\theta\in\Theta}\right)$, dominato dalla misura σ-finita μ, ammette una statistica sufficiente S, allora lo stimatore Bayesiano T della funzione ξ dato dall'equazione (13.18) è funzione di S, ovvero esiste una funzione Boreliana $\varphi\colon \mathbb{R} \to \mathbb{R}$ tale che $T = \varphi \circ S$.*

DIM. Per il *Criterio di fattorizzazione* la verosimiglianza del modello si può scrivere nella forma

$$\mathbf{L}_\theta(\omega) = h(\omega)\, \varphi_\theta(S(\omega)),$$

valida per ogni $\theta \in \Theta$ e per quasi (rispetto a μ) ogni $\omega \in \Omega$. Dunque l'espressione per T diventa

$$T(\omega) = \frac{\int_\Theta \xi(\theta)\, \varphi_\theta(S(\omega)) \, \mathrm{d}\pi(\theta)}{\int_\Theta \varphi_t(S(\omega)) \, \mathrm{d}\pi(t)},$$

dalla quale è evidente che T dipende da ω solo attraverso S. □

Gli Esempi che seguono illustrano la costruzione di alcuni stimatori Bayesiani.

ESEMPIO 13.5 (Legge di Bernoulli). ►

Si consideri il campione $\mathbf{X}^n_{\mathrm{Be}(\theta)}$, $\theta \in \Theta = (0, 1)$. La densità di ogni v.a. X_i del campione è

$$f_\theta(x_i) = \theta^{x_i}\, (1-\theta)^{1-x_i}\, \mathbb{I}_{\{0,1\}}(x_i);$$

pertanto, ponendo $S_n = \sum_{i=1}^n x_i$, la verosimiglianza del campione è

$$\mathbf{L}_\theta(\boldsymbol{x}) = \theta^{S_n}\,(1-\theta)^{n-S_n}.$$

Ora, si prenda come legge a priori $\pi \sim \beta(a,b)$, con $a,b>0$ noti, avente densità

$$\frac{\mathrm{d}\pi(\theta)}{\mathrm{d}\theta} = \frac{\theta^{a-1}\,(1-\theta)^{b-1}}{\mathrm{B}(a,b)}\,\mathbb{I}_{(0,1)}(\theta).$$

Se $S_n = s$, la legge a posteriori π_s risulta individuata da

$$\frac{\mathrm{d}\pi_s(\theta)}{\mathrm{d}\theta} = \frac{\theta^{s+a-1}\,(1-\theta)^{n-s+b-1}}{\int_0^1 t^{s+a-1}\,(1-t)^{n-s+b-1}\,\mathrm{d}t}\,\mathbb{I}_{(0,1)}(\theta) = \frac{\theta^{s+a-1}\,(1-\theta)^{n-s+b-1}}{\mathrm{B}(s+a,n-s+b)}\,\mathbb{I}_{(0,1)}(\theta),$$

ovvero una legge $\beta(s+a,n-s+b)$. Dunque lo stimatore Bayesiano di $\xi(\theta)=\theta$ è

$$\begin{aligned} T(\boldsymbol{x}) &= \int_0^1 \theta\;\mathrm{d}\pi_s(\theta) \\ &= \frac{\mathrm{B}(s+a+1,n-s+b)}{\mathrm{B}(s+a,n-s+b)} \\ &= \frac{s+a}{n+a+b}. \end{aligned}$$

In termini delle v.a. del campione lo stimatore Bayesiano T si scrive

$$T(\boldsymbol{X}) = \frac{\sum_{i=1}^n X_i + a}{n+a+b} \tag{13.19}$$

che, evidentemente, è distorto, essendo $a,b>0$. Si noti che in questo caso lo SMV di θ è lo stimatore corretto $\hat\theta = \overline{X}_n$, che si potrebbe ottenere ponendo $a=b=0$ nella definizione di T: ciò non è possibile, in quanto π non sarebbe una legge di probabilità. ◀

OSSERVAZIONE 13.5.

La stima Bayesiana del parametro θ avviene attraverso un "aggiornamento" della legge a priori π indotto dall'osservazione campionaria s: infatti si passa dalla legge $\pi \sim \beta(a,b)$ alla legge $\pi_s \sim \beta(s+a,n-s+b)$. Ora, *a priori* (ovvero prima di aver osservato il campione), la stima di θ quale speranza della legge π sarebbe stata $\frac{a}{a+b}$; *a posteriori* (ovvero dopo aver osservato il campione) tale stima diventa $\frac{s+a}{n+a+b}$, che "corregge" l'assunzione iniziale. In effetti, l'approccio Bayesiano fornisce uno stimatore T che risulta essere una combinazione lineare, opportunamente pesata, della stima *a priori* e di quella fornita dallo stimatore classico (corretto) $\overline{X}_n$: infatti

$$T = \frac{a+b}{n+a+b}\,\frac{a}{a+b} + \frac{n}{n+a+b}\,\overline{X}_n.$$

Si noti come, al crescere di n, $T \to \overline{X}_n$. Evidentemente T non può fornire stime di θ pari ai valori estremali $\theta = 0,1$. Si osservi inoltre come T sia funzione della statistica sufficiente $\sum_{i=1}^n X_i$ e come la varianza della legge π, pari a $\frac{ab}{(1+a+b)(a+b)^2}$, sia maggiore di quella della legge π_s, pari a $\frac{(n-s+b)(s+a)}{(1+a+b+n)(a+b+n)^2}$, che tende a zero per $n \to \infty$.

ESEMPIO 13.6 (Legge di Poisson)**.** ►

Si consideri il campione $\mathbf{X}^n_{\mathrm{P}(\theta)}$, $\theta \in \Theta = \mathbb{R}_+$. La densità di ogni v.a. X_i del campione è

$$f_\theta(x_i) = e^{-\theta}\,\frac{\theta^{x_i}}{x_i!}\,\mathbb{I}_{\mathbb{N}}(x_i);$$

pertanto, ponendo $S_n = \sum_{i=1}^n x_i$, la verosimiglianza del campione è

$$\mathbf{L}_\theta(\boldsymbol{x}) = e^{-n\theta}\,\frac{\theta^{S_n}}{\prod_{i=1}^n x_i!}.$$

Ora, si prenda come legge a priori $\pi \sim \Gamma(\alpha, \lambda)$, con $\alpha, \lambda > 0$ noti, avente densità

$$\frac{\mathrm{d}\pi(\theta)}{\mathrm{d}\theta} = \frac{\lambda^\alpha}{\Gamma(\alpha)}\, \theta^{\alpha-1}\, e^{-\lambda\theta}\, \mathbb{I}_{\mathbb{R}_+}(\theta).$$

Se $S_n = s$, la legge a posteriori π_s risulta individuata da

$$\frac{\mathrm{d}\pi_s(\theta)}{\mathrm{d}\theta} = \frac{\theta^{s+\alpha-1}\, e^{-(\lambda+n)\theta}}{\int_0^\infty t^{s+\alpha-1}\, e^{-(\lambda+n)t}\, \mathrm{d}t}\, \mathbb{I}_{\mathbb{R}_+}(\theta),$$

ovvero una legge $\Gamma(s+\alpha, \lambda+n)$. Dunque lo stimatore Bayesiano di $\xi(\theta) = \theta$ è

$$\begin{aligned} T(\boldsymbol{x}) &= \int_0^\infty \theta\ \mathrm{d}\pi_s(\theta) \\ &= \frac{\int_0^\infty \theta^{s+\alpha}\, e^{-(\lambda+n)\theta}\ \mathrm{d}\theta}{\int_0^\infty t^{s+\alpha-1}\, e^{-(\lambda+n)t}\ \mathrm{d}t} \\ &= \frac{s+\alpha}{\lambda+n}. \end{aligned}$$

In termini delle v.a. del campione lo stimatore Bayesiano T si scrive

$$T(\boldsymbol{X}) = \frac{\sum_{i=1}^n X_i + \alpha}{\lambda+n} \tag{13.20}$$

che, evidentemente, è distorto, essendo $\alpha, \lambda > 0$. Si noti che in questo caso lo SMV di θ è lo stimatore corretto $\hat{\theta} = \overline{X}_n$, che si potrebbe ottenere ponendo $\alpha = \lambda = 0$ nella definizione di T: ciò non è possibile, in quanto π non sarebbe una legge di probabilità. ◀

OSSERVAZIONE 13.6.

Anche in questo caso l'approccio Bayesiano alla stima del parametro θ consiste nell'"aggiornamento" della legge a priori π indotto dall'osservazione campionaria s: infatti si passa dalla legge $\pi \sim \Gamma(\alpha, \lambda)$ alla legge $\pi_s \sim \Gamma(s+\alpha, \lambda+n)$. Ancora una volta si passa dalla stima *a priori* $\frac{\alpha}{\lambda}$ a quella *a posteriori* $\frac{s+\alpha}{\lambda+n}$, e lo stimatore Bayesiano T risulta essere una combinazione lineare, opportunamente pesata, della stima *a priori* e di quella fornita dallo stimatore classico (corretto) $\overline{X}_n$: infatti

$$T = \frac{\lambda}{\lambda+n}\, \frac{\alpha}{\lambda} + \frac{n}{\lambda+n}\, \overline{X}_n.$$

Si noti come, al crescere di n, $T \to \overline{X}_n$. Evidentemente T non può fornire stime di θ pari al valore estremale $\theta = 0$. Si osservi inoltre come T sia funzione della statistica sufficiente $\sum_{i=1}^n X_i$ e come la varianza della legge π, pari a $\frac{\alpha}{\lambda^2}$, sia maggiore di quella della legge π_s, pari a $\frac{s+\alpha}{(\lambda+n)^2}$, che tende a zero per $n \to \infty$.

ESEMPIO 13.7 (Legge Normale). ►

Si consideri il campione $\mathbf{X}^n_{\mathtt{N}(\theta,1)}$, $\theta \in \Theta = \mathbb{R}$. La densità di ogni v.a. X_i del campione è

$$f_\theta(x_i) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}(x_i-\theta)^2}, \quad x_i \in \mathbb{R};$$

pertanto, ponendo $S_1 = \sum_{i=1}^n x_i$ e $S_2 = \sum_{i=1}^n x_i^2$, la verosimiglianza del campione è

$$\mathbf{L}_\theta(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}}\, e^{-\frac{1}{2}S_2 + \theta S_1 - \frac{n}{2}\theta^2}.$$

Ora, si prenda come legge a priori $\pi \sim \mathtt{N}\big(0, \sigma^2\big)$, con $\sigma > 0$ noto. Dunque, la legge congiunta del vettore $(\boldsymbol{X}, \theta)$ è Normale multivariata, e perciò la legge a posteriori $\pi_{\boldsymbol{x}}$ (ovvero la legge condizionale di θ dato il campione) risulta essere ancora Normale. In particolare, dopo calcoli tediosi, si ha

$$\pi_{\boldsymbol{x}} \sim \mathtt{N}\left(\frac{n\sigma^2}{1+n\sigma^2}\, \overline{x}_n,\, \frac{\sigma^2}{1+n\sigma^2}\right),$$

sicché lo stimatore Bayesiano di $\xi(\theta) = \theta$, ovvero la speranza della legge $\pi_{\boldsymbol{x}}$, è

$$T(\boldsymbol{X}) = \frac{n\sigma^2}{1 + n\sigma^2}\, \overline{X}_n, \tag{13.21}$$

che, evidentemente, è distorto. Si noti che in questo caso lo SMV di θ è lo stimatore corretto $\hat{\theta} = \overline{X}_n$. ◀

OSSERVAZIONE 13.7.

Anche in questo caso l'approccio Bayesiano alla stima del parametro θ consiste nell'"aggiornamento" della legge a priori π indotto dall'osservazione campionaria $\boldsymbol{x}$: infatti si passa dalla legge $\pi \sim \mathtt{N}\big(0, \sigma^2\big)$ alla legge $\pi_{\boldsymbol{x}} \sim \mathtt{N}\Big(\frac{n\sigma^2}{1+n\sigma^2}\, \overline{x}_n, \frac{\sigma^2}{1+n\sigma^2}\Big)$. Ancora una volta si passa dalla stima *a priori* pari a 0 a quella *a posteriori* pari a $\frac{n\sigma^2}{1+n\sigma^2}\, \overline{x}_n$, e lo stimatore Bayesiano T risulta essere una combinazione lineare della stima *a priori* e di quella fornita dallo stimatore classico (corretto) $\overline{X}_n$. Si noti come, al crescere di n, $T \to \overline{X}_n$. Si osservi inoltre come T sia funzione della statistica sufficiente $\sum_{i=1}^n X_i$ e come la varianza della legge π, pari a σ^2, sia maggiore di quella della legge $\pi_{\boldsymbol{x}}$, pari a $\frac{\sigma^2}{1+n\sigma^2}$, che tende a zero per $n \to \infty$. Infine, poco sarebbe cambiato (se non la complessità dei calcoli) se la legge del campione fosse stata Normale con varianza (nota) diversa da 1, o se la legge a priori fosse stata Normale con speranza (nota) diversa da 0.

Negli Esempi appena visti non si è utilizzata come legge a priori π quella uniforme; in effetti, in alcuni casi le informazioni a disposizione dello sperimentatore portano a ritenere maggiormente consigliabile l'adozione di una legge differente, tipicamente più "informativa". Inoltre, l'uso della legge Uniforme è del tutto impossibile ogni qual volta la chiusura $\overline{\Theta}$ dello spazio dei parametri Θ in $\mathbb{R}^d$ non sia compatta. In tutti questi casi è necessario ricorrere a leggi a priori che dipendano da uno o più parametri. In particolare, talora è possibile (e conveniente) utilizzare leggi a priori cosiddette *coniugate*, una strategia che, sebbene possa non riflettere in maniera adeguata l'informazione disponibile allo statistico prima dell'esperimento, è molto diffusa, essendo giustificata da ragioni tecniche di semplificazione dei calcoli.

DEFINIZIONE 13.10 (Legge coniugata)**.** Una famiglia $\mathfrak{H}$ di leggi di probabilità su $(\Theta, \mathcal{T})$ si dice *coniugata* se, per ogni legge a priori $\pi \in \mathfrak{H}$ e per ogni $\omega \in \Omega$, la legge a posteriori π_ω appartiene anch'essa a $\mathfrak{H}$.

Dunque, utilizzando le leggi coniugate, nel passaggio dalla legge a priori a quella a posteriori si ha un "aggiornamento" solamente dei parametri che individuano la legge all'interno della famiglia $\mathfrak{H}$. Il Teorema che segue mette in luce, ancora una volta, l'importanza del modello esponenziale.

TEOREMA 13.9. *Dato un campione $X_1, \ldots, X_n$ avente legge appartenente alla famiglia esponenziale con densità, rispetto alla misura di Lebesgue, uguale a*

$$f_\theta(x) = s(x)\, e^{\theta x - \chi(\theta)},$$

le leggi $\pi_{\alpha,\beta}$ aventi densità

$$\pi_{\alpha,\beta}(\theta) = c(\alpha, \beta)\, e^{\beta\theta - \alpha\chi(\theta)} \tag{13.22}$$

costituiscono una famiglia coniugata.

DIM. La verosimiglianza del campione è, per $\boldsymbol{x} \in \mathbb{R}^n$,

$$\mathbf{L}_\theta(\boldsymbol{x}) = \prod_{i=1}^n s(x_i)\, e^{\theta \sum_{i=1}^n x_i - n\chi(\theta)}.$$

Pertanto, se la legge a priori è data dall'*equazione (13.22)*, la legge a posteriori ha densità proporzionale a

$$e^{\theta\left(\beta + \sum_{i=1}^n x_i\right) - (n+\alpha)\chi(\theta)},$$

che è ancora dello stesso tipo: i parametri iniziali α e β sono stati sostituiti, rispettivamente, da

$$\alpha \to \alpha + n \quad \text{e} \quad \beta \to \beta + \sum_{i=1}^n x_i. \tag{13.23}$$

□

NOTA 13.6.

In base al *Teorema (13.9)* l'"aggiornamento" del parametro α è funzione dell'ampiezza n del campione, mentre quello del parametro β è funzione dell'osservazione $\boldsymbol{x}$.

Evidentemente negli *Esempi (13.5)–(13.7)* si sono utilizzate leggi coniugate; in particolare, si noti come negli ultimi due $\overline{\Theta}$ non sia compatto. Gli Esempi che seguono illustrano altri casi importanti.

ESEMPIO 13.8 (Legge Esponenziale). ►

Si consideri il campione $\mathbf{X}^n_{\mathsf{e}(\theta)}$, $\theta \in \Theta = \mathbb{R}_+$. La densità di ogni v.a. X_i del campione è

$$f_\theta(x_i) = \theta\, e^{-\theta x_i}\, \mathbb{I}_{\mathbb{R}_+}(x_i);$$

pertanto, ponendo $S_n = \sum_{i=1}^n x_i$, la verosimiglianza del campione è

$$\mathbf{L}_\theta(\boldsymbol{x}) = \theta^n\, e^{-\theta S_n}.$$

Ora, si prenda come legge a priori $\pi \sim \mathsf{e}(\alpha) \sim \Gamma(1, \alpha)$, con $\alpha > 0$ noto. Se $S_n = s$, la legge a posteriori π_s (rispetto alla misura di Lebesgue λ) ha densità

$$\frac{\mathrm{d}\pi_s(\theta)}{\mathrm{d}\theta} = \frac{\theta^n\, e^{-\theta s}\, \alpha\, e^{-\alpha\theta}}{\alpha \int_0^\infty t^n\, e^{-t(\alpha+s)}\, \mathrm{d}t}\, \mathbb{I}_{\mathbb{R}_+}(\theta) = \frac{(\alpha+s)^{n+1}}{n!}\, \theta^n\, e^{-\theta(\alpha+s)}\, \mathbb{I}_{\mathbb{R}_+}(\theta),$$

ovvero una legge $\Gamma(n+1, \alpha+s)$. Dunque lo stimatore Bayesiano di $\xi(\theta) = \theta$ è

$$\begin{aligned} T(\boldsymbol{x}) &= \int_0^\infty \theta\ \mathrm{d}\pi_s(\theta) \\ &= \frac{n+1}{\alpha+s} \\ &= \frac{1+\frac{1}{n}}{\frac{\alpha}{n}+\overline{x}_n}. \end{aligned}$$

In termini delle v.a. del campione lo stimatore Bayesiano T si scrive

$$T(\boldsymbol{X}) = \frac{n+1}{\alpha + \sum_{i=1}^n X_i} = \frac{1+\frac{1}{n}}{\frac{\alpha}{n}+\overline{X}_n}. \tag{13.24}$$

Si noti che in questo caso lo SMV di θ è $\hat{\theta} = 1/\overline{X}_n$. ◄

OSSERVAZIONE 13.8.

Anche in questo caso l'approccio Bayesiano alla stima del parametro θ consiste nell'"aggiornamento" della legge a priori π indotto dall'osservazione campionaria s: infatti si passa dalla legge $\pi \sim \Gamma(1, \alpha)$ alla legge $\pi_s \sim \Gamma(n+1, \alpha+s)$, e la stima *a priori* $\frac{1}{\alpha}$ diventa quella *a posteriori* $\frac{n+1}{\alpha+s}$. Evidentemente $T \to 1/\overline{X}_n$ al crescere di n. Si osservi inoltre come T sia funzione della statistica sufficiente $\sum_{i=1}^n X_i$ e come la varianza della legge π, pari a $\frac{1}{\alpha^2}$, sia maggiore di quella della legge π_s, pari a $\frac{n+1}{(\alpha+s)^2}$, che tende a zero per $n \to \infty$.

Nell'Esempio appena visto è possibile adottare una strategia alternativa alla stima del parametro θ che, sebbene difficile da giustificare pienamente, porta a risultati del tutto accettabili. Si scelga come "legge" a priori π la misura di Lebesgue λ su $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$, che *non* è una misura di probabilità sullo spazio indicato. A questa legge corrisponde lo stimatore

$$T' = \frac{\int_0^\infty \theta^{n+1}\, e^{-\theta S_n}\ \mathrm{d}\theta}{\int_0^\infty \theta^n\, e^{-\theta S_n}\ \mathrm{d}\theta} = \frac{n+1}{\sum_{i=1}^n X_i}. \tag{13.25}$$

Dunque, pur avendo utilizzato una "legge" a priori che, in termini rigorosi, è illegittima e non ammissibile, si perviene ad un risultato del tutto credibile.

ESEMPIO 13.9 (Legge Geometrica). ►

Si consideri il campione $\mathbf{X}^n_{\mathsf{g}(\theta)}$, $\theta \in \Theta = (0, 1)$. La densità di ogni v.a. X_i del campione è

$$f_\theta(x_i) = \theta\, (1-\theta)^{x_i - 1}, \quad x_i = 1, 2, \ldots;$$

pertanto, ponendo $S_n = \sum_{i=1}^n x_i$, la verosimiglianza del campione è

$$\mathbf{L}_\theta(\boldsymbol{x}) = \theta^n (1-\theta)^{S_n - n}.$$

Ora, si prenda come legge a priori $\pi \sim \mathtt{U_c}(0,1)$. Se $S_n = s$, la legge a posteriori π_s (rispetto alla misura di Lebesgue λ) risulta individuata da

$$\frac{\mathrm{d}\pi_s(\theta)}{\mathrm{d}\theta} = \frac{\theta^n (1-\theta)^{s-n}}{\int_0^1 t^n (1-t)^{s-n}\ \mathrm{d}t}\, \mathbb{I}_{(0,1)}(\theta) = \frac{\theta^n (1-\theta)^{s-n}}{\mathrm{B}(n+1, s-n+1)}\, \mathbb{I}_{(0,1)}(\theta),$$

ovvero una legge $\beta(n+1, s-n+1)$. Dunque lo stimatore Bayesiano di $\xi(\theta) = \theta$ è

$$\begin{aligned} T(\boldsymbol{x}) &= \int_0^1 \theta\ \mathrm{d}\pi_s(\theta) \\ &= \frac{\int_0^1 \theta^{n+1} (1-\theta)^{s-n}\ \mathrm{d}\theta}{\mathrm{B}(n+1, s-n+1)} \\ &= \frac{\mathrm{B}(n+2, s-n+1)}{\mathrm{B}(n+1, s-n+1)} \\ &= \frac{n+1}{s+2}. \end{aligned}$$

In termini delle v.a. del campione lo stimatore Bayesiano T si scrive

$$T(\boldsymbol{X}) = \frac{n+1}{\sum_{i=1}^n X_i + 2}. \tag{13.26}$$

Si noti che in questo caso lo SMV di θ è $\hat{\theta} = 1/\overline{X}_n$. ◀

OSSERVAZIONE 13.9.

Anche in questo caso l'approccio Bayesiano alla stima del parametro θ consiste nell'"aggiornamento" della legge a priori π indotto dall'osservazione campionaria s: infatti si passa dalla legge $\pi \sim \mathtt{U_c}(0,1)$, assai poco ricca di "informazione", alla legge $\pi_s \sim \beta(n+1, s-n+1)$, di certo più informativa, e la stima *a priori* $\frac{1}{2}$ diventa quella *a posteriori* $\frac{n+1}{s+2}$. Essendo

$$\frac{1}{T(\boldsymbol{X})} = \frac{\sum_{i=1}^n X_i + 2}{n+1} = \frac{n}{n+1}\, \overline{X}_n + \frac{2}{n+1},$$

si ha che $T \to 1/\overline{X}_n$ al crescere di n. Evidentemente T non può fornire stime di θ pari ai valori estremali $\theta = 0, 1$. Si osservi inoltre come T sia funzione della statistica sufficiente $\sum_{i=1}^n X_i$ e come la varianza della legge π, pari a $\frac{1}{12}$, sia maggiore di quella della legge π_s, pari a $\frac{(n+1)(s-n+1)}{(s+3)(s+2)^2}$, che tende a zero per $n \to \infty$.

ESEMPIO 13.10 (Controllo di qualità)**.** ▶

Nell'ambito delle procedure di *controllo di qualità* una fabbrica verifica il 10% dei prodotti che produce. In un lotto di 100 pezzi ne sono esaminati 10, e 3 di questi risultano difettosi. Si vuole stimare la percentuale θ di pezzi non conformi.

Evidentemente siamo in presenza di un campione Bernoulliano di legge $\mathtt{Be}(\theta)$, $\theta \in [0,1]$, ed è naturale supporre che sia $\Omega = \{0, 1, \ldots, 10\}$ e $\mathcal{F} = \mathcal{P}(\Omega)$; si può inoltre assumere che sia $\Theta = [0,1]$ e $\mathcal{T} = \mathcal{B}(\Theta)$. In mancanza di altre informazioni si può scegliere quale legge a priori $\pi \sim \mathtt{U_c}(0,1)$. Dunque la verosimiglianza del campione è

$$\mathbf{L}_\theta(\boldsymbol{x}) = \theta^3 (1-\theta)^7.$$

La legge a posteriori π_3 (rispetto alla misura di Lebesgue λ) risulta individuata da

$$\frac{\mathrm{d}\pi_3(\theta)}{\mathrm{d}\theta} = \frac{\theta^3 (1-\theta)^7}{\int_0^1 t^3 (1-t)^7\ \mathrm{d}t}\, \mathbb{I}_{(0,1)}(\theta) = \frac{\theta^3 (1-\theta)^7}{\mathrm{B}(4,8)}\, \mathbb{I}_{(0,1)}(\theta),$$

ovvero una legge $\beta(4,8)$. Dunque lo stimatore Bayesiano di $\xi(\theta) = \theta$ è

$$T(3) = \int_0^1 \theta\ \mathrm{d}\pi_3(\theta) = \frac{\int_0^1 \theta^4 (1-\theta)^7\ \mathrm{d}\theta}{\mathrm{B}(4,8)} = \frac{\mathrm{B}(5,8)}{\mathrm{B}(4,8)}.$$

che assume il valore $T(3) = \frac{1}{3}$. Tuttavia, l'ipotesi che il parametro θ sia distribuito uniformemente è poco credibile: infatti ciò significherebbe che percentuali "piccole" (e quindi ragionevoli) di pezzi difettosi sarebbero probabili quanto percentuali elevate, una situazione che sicuramente porterebbe al fallimento dell'impresa. Basandosi sulle conoscenze riguardanti i macchinari, l'abilità e le competenze delle maestranze, e altre informazioni simili, la fabbrica sceglierà una legge a priori differente.

Ad esempio si può scegliere $\pi' \sim \beta(2, 10)$, che ha moda $\tilde{\theta} = \frac{1}{10}$ e speranza $\overline{\theta} = \frac{1}{6}$: ciò manifesta da parte della fabbrica una certa fiducia nella bontà dei propri manufatti, di certo maggiore di quella espressa dalla legge Uniforme, che ha media pari a $\frac{1}{2}$. Dunque la nuova legge a posteriori π'_3 risulta individuata da

$$\frac{\mathrm{d}\pi'_3(\theta)}{\mathrm{d}\theta} = \frac{\theta^4 (1-\theta)^{16}}{\int_0^1 t^4 (1-t)^{16} \, \mathrm{d}t} \mathbb{I}_{(0,1)}(\theta) = \frac{\theta^4 (1-\theta)^{16}}{\mathrm{B}(5, 17)} \mathbb{I}_{(0,1)}(\theta),$$

ovvero una legge $\beta(5, 17)$, che ha moda $\tilde{\theta} = \frac{2}{10}$ e speranza $\overline{\theta} \approx 0.23$. Dunque la nuova stima Bayesiano di θ è

$$T'(3) = \int_0^1 \theta \, \mathrm{d}\pi'_3(\theta) = \frac{\int_0^1 \theta^5 (1-\theta)^{16} \, \mathrm{d}\theta}{\mathrm{B}(5, 17)} = \frac{\mathrm{B}(6, 17)}{\mathrm{B}(5, 17)} \approx 0.23.$$

Si noti che in questo caso lo SMV di θ è $\hat{\theta} = \overline{X}_n$, che assume il valore $\overline{x}_n = \frac{3}{10}$. ◀

Appendice A

Quantili della legge Normale

z_α	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965

Tabella A.1: quantili della legge Normale $\mathtt{N}(0, 1)$.

Appendice B

Quantili della legge t-Student

	α						
n	**0.75**	**0.90**	**0.95**	**0.975**	**0.99**	**0.995**	**0.9995**
1	1.00000	3.07770	6.31380	12.70600	31.82100	63.65700	636.62000
2	0.81650	1.88560	2.92000	4.30270	6.96460	9.92480	31.59900
3	0.76489	1.63770	2.35340	3.18240	4.54070	5.84090	12.92400
4	0.74070	1.53320	2.13180	2.77640	3.74690	4.60410	8.61030
5	0.72669	1.47590	2.01500	2.57060	3.36490	4.03210	6.86880
6	0.71756	1.43980	1.94320	2.44690	3.14270	3.70740	5.95880
7	0.71114	1.41490	1.89460	2.36460	2.99800	3.49950	5.40790
8	0.70639	1.39680	1.85950	2.30600	2.89650	3.35540	5.04130
9	0.70272	1.38300	1.83310	2.26220	2.82140	3.24980	4.78090
10	0.69981	1.37220	1.81250	2.22810	2.76380	3.16930	4.58690
11	0.69745	1.36340	1.79590	2.20100	2.71810	3.10580	4.43700
12	0.69548	1.35620	1.78230	2.17880	2.68100	3.05450	4.31780
13	0.69383	1.35020	1.77090	2.16040	2.65030	3.01230	4.22080
14	0.69242	1.34500	1.76130	2.14480	2.62450	2.97680	4.14050
15	0.69120	1.34060	1.75310	2.13140	2.60250	2.94670	4.07280
16	0.69013	1.33680	1.74590	2.11990	2.58350	2.92080	4.01500
17	0.68920	1.33340	1.73960	2.10980	2.56690	2.89820	3.96510
18	0.68836	1.33040	1.73410	2.10090	2.55240	2.87840	3.92160
19	0.68762	1.32770	1.72910	2.09300	2.53950	2.86090	3.88340
20	0.68695	1.32530	1.72470	2.08600	2.52800	2.84530	3.84950
21	0.68635	1.32320	1.72070	2.07960	2.51760	2.83140	3.81930
22	0.68581	1.32120	1.71710	2.07390	2.50830	2.81880	3.79210
23	0.68531	1.31950	1.71390	2.06870	2.49990	2.80730	3.76760
24	0.68485	1.31780	1.71090	2.06390	2.49220	2.79690	3.74540
25	0.68443	1.31630	1.70810	2.05950	2.48510	2.78740	3.72510
26	0.68404	1.31500	1.70560	2.05550	2.47860	2.77870	3.70660
27	0.68368	1.31370	1.70330	2.05180	2.47270	2.77070	3.68960
28	0.68335	1.31250	1.70110	2.04840	2.46710	2.76330	3.67390
29	0.68304	1.31140	1.69910	2.04520	2.46200	2.75640	3.65940
30	0.68276	1.31040	1.69730	2.04230	2.45730	2.75000	3.64600
40	0.68067	1.30310	1.68390	2.02110	2.42330	2.70450	3.55100
60	0.67860	1.29580	1.67060	2.00030	2.39010	2.66030	3.46020
80	0.67757	1.29220	1.66410	1.99010	2.37390	2.63870	3.41630
100	0.67695	1.29010	1.66020	1.98400	2.36420	2.62590	3.39050
120	0.67654	1.28860	1.65770	1.97990	2.35780	2.61740	3.37350

Tabella B.1: quantili della legge t-Student $\mathtt{t}(n)$.

Appendice C

Quantili della legge Chi-quadro (I)

	α						
n	**0.005**	**0.01**	**0.025**	**0.05**	**0.1**	**0.25**	**0.5**
1	0.00004	0.00016	0.00098	0.00393	0.01579	0.10153	0.45494
2	0.01003	0.02010	0.05064	0.10259	0.21072	0.57536	1.38630
3	0.07172	0.11483	0.21580	0.35185	0.58437	1.21250	2.36600
4	0.20699	0.29711	0.48442	0.71072	1.06360	1.92260	3.35670
5	0.41174	0.55430	0.83121	1.14550	1.61030	2.67460	4.35150
6	0.67573	0.87209	1.23730	1.63540	2.20410	3.45460	5.34810
7	0.98926	1.23900	1.68990	2.16730	2.83310	4.25490	6.34580
8	1.34440	1.64650	2.17970	2.73260	3.48950	5.07060	7.34410
9	1.73490	2.08790	2.70040	3.32510	4.16820	5.89880	8.34280
10	2.15590	2.55820	3.24700	3.94030	4.86520	6.73720	9.34180
11	2.60320	3.05350	3.81570	4.57480	5.57780	7.58410	10.34100
12	3.07380	3.57060	4.40380	5.22600	6.30380	8.43840	11.34000
13	3.56500	4.10690	5.00880	5.89190	7.04150	9.29910	12.34000
14	4.07470	4.66040	5.62870	6.57060	7.78950	10.16500	13.33900
15	4.60090	5.22930	6.26210	7.26090	8.54680	11.03700	14.33900
16	5.14220	5.81220	6.90770	7.96160	9.31220	11.91200	15.33800
17	5.69720	6.40780	7.56420	8.67180	10.08500	12.79200	16.33800
18	6.26480	7.01490	8.23070	9.39050	10.86500	13.67500	17.33800
19	6.84400	7.63270	8.90650	10.11700	11.65100	14.56200	18.33800
20	7.43380	8.26040	9.59080	10.85100	12.44300	15.45200	19.33700
21	8.03370	8.89720	10.28300	11.59100	13.24000	16.34400	20.33700
22	8.64270	9.54250	10.98200	12.33800	14.04100	17.24000	21.33700
23	9.26040	10.19600	11.68900	13.09100	14.84800	18.13700	22.33700
24	9.88620	10.85600	12.40100	13.84800	15.65900	19.03700	23.33700
25	10.52000	11.52400	13.12000	14.61100	16.47300	19.93900	24.33700
26	11.16000	12.19800	13.84400	15.37900	17.29200	20.84300	25.33600
27	11.80800	12.87900	14.57300	16.15100	18.11400	21.74900	26.33600
28	12.46100	13.56500	15.30800	16.92800	18.93900	22.65700	27.33600
29	13.12100	14.25600	16.04700	17.70800	19.76800	23.56700	28.33600
30	13.78700	14.95300	16.79100	18.49300	20.59900	24.47800	29.33600
31	14.45800	15.65500	17.53900	19.28100	21.43400	25.39000	30.33600
32	15.13400	16.36200	18.29100	20.07200	22.27100	26.30400	31.33600
33	15.81500	17.07400	19.04700	20.86700	23.11000	27.21900	32.33600

Tabella C.1: quantili della legge Chi-quadro $\chi^2(n)$ (I).

Appendice D

Quantili della legge Chi-quadro (II)

	α						
n	**0.75**	**0.9**	**0.95**	**0.975**	**0.99**	**0.995**	**0.9995**
1	1.32330	2.70550	3.84150	5.02390	6.63490	7.87940	12.11600
2	2.77260	4.60520	5.99150	7.37780	9.21030	10.59700	15.20200
3	4.10830	6.25140	7.81470	9.34840	11.34500	12.83800	17.73000
4	5.38530	7.77940	9.48770	11.14300	13.27700	14.86000	19.99700
5	6.62570	9.23640	11.07000	12.83300	15.08600	16.75000	22.10500
6	7.84080	10.64500	12.59200	14.44900	16.81200	18.54800	24.10300
7	9.03710	12.01700	14.06700	16.01300	18.47500	20.27800	26.01800
8	10.21900	13.36200	15.50700	17.53500	20.09000	21.95500	27.86800
9	11.38900	14.68400	16.91900	19.02300	21.66600	23.58900	29.66600
10	12.54900	15.98700	18.30700	20.48300	23.20900	25.18800	31.42000
11	13.70100	17.27500	19.67500	21.92000	24.72500	26.75700	33.13700
12	14.84500	18.54900	21.02600	23.33700	26.21700	28.30000	34.82100
13	15.98400	19.81200	22.36200	24.73600	27.68800	29.81900	36.47800
14	17.11700	21.06400	23.68500	26.11900	29.14100	31.31900	38.10900
15	18.24500	22.30700	24.99600	27.48800	30.57800	32.80100	39.71900
16	19.36900	23.54200	26.29600	28.84500	32.00000	34.26700	41.30800
17	20.48900	24.76900	27.58700	30.19100	33.40900	35.71800	42.87900
18	21.60500	25.98900	28.86900	31.52600	34.80500	37.15600	44.43400
19	22.71800	27.20400	30.14400	32.85200	36.19100	38.58200	45.97300
20	23.82800	28.41200	31.41000	34.17000	37.56600	39.99700	47.49800
21	24.93500	29.61500	32.67100	35.47900	38.93200	41.40100	49.01100
22	26.03900	30.81300	33.92400	36.78100	40.28900	42.79600	50.51100
23	27.14100	32.00700	35.17200	38.07600	41.63800	44.18100	52.00000
24	28.24100	33.19600	36.41500	39.36400	42.98000	45.55900	53.47900
25	29.33900	34.38200	37.65200	40.64600	44.31400	46.92800	54.94700
26	30.43500	35.56300	38.88500	41.92300	45.64200	48.29000	56.40700
27	31.52800	36.74100	40.11300	43.19500	46.96300	49.64500	57.85800
28	32.62000	37.91600	41.33700	44.46100	48.27800	50.99300	59.30000
29	33.71100	39.08700	42.55700	45.72200	49.58800	52.33600	60.73500
30	34.80000	40.25600	43.77300	46.97900	50.89200	53.67200	62.16200
31	35.88700	41.42200	44.98500	48.23200	52.19100	55.00300	63.58200
32	36.97300	42.58500	46.19400	49.48000	53.48600	56.32800	64.99500
33	38.05800	43.74500	47.40000	50.72500	54.77600	57.64800	66.40300

Tabella D.1: quantili della legge Chi-quadro $\chi^2(n)$ (II).

Appendice E

Quantili della legge Fisher-Snedecor (I)

	m										
n	**1**	**3**	**5**	**8**	**10**	**12**	**15**	**20**	**30**	**50**	**100**
1	161.45	215.71	230.16	238.88	241.88	243.91	245.95	248.01	250.10	251.77	253.04
2	18.51	19.16	19.30	19.37	19.40	19.41	19.43	19.45	19.46	19.48	19.49
3	10.13	9.28	9.01	8.85	8.79	8.74	8.70	8.66	8.62	8.58	8.55
4	7.71	6.59	6.26	6.04	5.96	5.91	5.86	5.80	5.75	5.70	5.66
5	6.61	5.41	5.05	4.82	4.74	4.68	4.62	4.56	4.50	4.44	4.41
6	5.99	4.76	4.39	4.15	4.06	4.00	3.94	3.87	3.81	3.75	3.71
7	5.59	4.35	3.97	3.73	3.64	3.57	3.51	3.44	3.38	3.32	3.27
8	5.32	4.07	3.69	3.44	3.35	3.28	3.22	3.15	3.08	3.02	2.97
9	5.12	3.86	3.48	3.23	3.14	3.07	3.01	2.94	2.86	2.80	2.76
10	4.96	3.71	3.33	3.07	2.98	2.91	2.85	2.77	2.70	2.64	2.59
11	4.84	3.59	3.20	2.95	2.85	2.79	2.72	2.65	2.57	2.51	2.46
12	4.75	3.49	3.11	2.85	2.75	2.69	2.62	2.54	2.47	2.40	2.35
13	4.67	3.41	3.03	2.77	2.67	2.60	2.53	2.46	2.38	2.31	2.26
14	4.60	3.34	2.96	2.70	2.60	2.53	2.46	2.39	2.31	2.24	2.19
15	4.54	3.29	2.90	2.64	2.54	2.48	2.40	2.33	2.25	2.18	2.12
16	4.49	3.24	2.85	2.59	2.49	2.42	2.35	2.28	2.19	2.12	2.07
17	4.45	3.20	2.81	2.55	2.45	2.38	2.31	2.23	2.15	2.08	2.02
18	4.41	3.16	2.77	2.51	2.41	2.34	2.27	2.19	2.11	2.04	1.98
19	4.38	3.13	2.74	2.48	2.38	2.31	2.23	2.16	2.07	2.00	1.94
20	4.35	3.10	2.71	2.45	2.35	2.28	2.20	2.12	2.04	1.97	1.91
21	4.32	3.07	2.68	2.42	2.32	2.25	2.18	2.10	2.01	1.94	1.88
22	4.30	3.05	2.66	2.40	2.30	2.23	2.15	2.07	1.98	1.91	1.85
23	4.28	3.03	2.64	2.37	2.27	2.20	2.13	2.05	1.96	1.88	1.82
24	4.26	3.01	2.62	2.36	2.25	2.18	2.11	2.03	1.94	1.86	1.80
25	4.24	2.99	2.60	2.34	2.24	2.16	2.09	2.01	1.92	1.84	1.78
26	4.23	2.98	2.59	2.32	2.22	2.15	2.07	1.99	1.90	1.82	1.76
27	4.21	2.96	2.57	2.31	2.20	2.13	2.06	1.97	1.88	1.81	1.74
28	4.20	2.95	2.56	2.29	2.19	2.12	2.04	1.96	1.87	1.79	1.73
29	4.18	2.93	2.55	2.28	2.18	2.10	2.03	1.94	1.85	1.77	1.71
30	4.17	2.92	2.53	2.27	2.16	2.09	2.01	1.93	1.84	1.76	1.70

Tabella E.1: quantili di ordine 0.95 della legge Fisher-Snedecor $\mathrm{F}(m, n)$.

Appendice F

Quantili della legge Fisher-Snedecor (II)

	m										
n	**1**	**3**	**5**	**8**	**10**	**12**	**15**	**20**	**30**	**50**	**100**
1	4052.2	5403.4	5763.6	5981.1	6055.8	6106.3	6157.3	6208.7	6260.6	6302.5	6334.1
2	98.50	99.17	99.30	99.37	99.40	99.42	99.43	99.45	99.47	99.48	99.49
3	34.12	29.46	28.24	27.49	27.23	27.05	26.87	26.69	26.50	26.35	26.24
4	21.20	16.69	15.52	14.80	14.55	14.37	14.20	14.02	13.84	13.69	13.58
5	16.26	12.06	10.97	10.29	10.05	9.89	9.72	9.55	9.38	9.24	9.13
6	13.75	9.78	8.75	8.10	7.87	7.72	7.56	7.40	7.23	7.09	6.99
7	12.25	8.45	7.46	6.84	6.62	6.47	6.31	6.16	5.99	5.86	5.75
8	11.26	7.59	6.63	6.03	5.81	5.67	5.52	5.36	5.20	5.07	4.96
9	10.56	6.99	6.06	5.47	5.26	5.11	4.96	4.81	4.65	4.52	4.41
10	10.04	6.55	5.64	5.06	4.85	4.71	4.56	4.41	4.25	4.12	4.01
11	9.65	6.22	5.32	4.74	4.54	4.40	4.25	4.10	3.94	3.81	3.71
12	9.33	5.95	5.06	4.50	4.30	4.16	4.01	3.86	3.70	3.57	3.47
13	9.07	5.74	4.86	4.30	4.10	3.96	3.82	3.66	3.51	3.38	3.27
14	8.86	5.56	4.69	4.14	3.94	3.80	3.66	3.51	3.35	3.22	3.11
15	8.68	5.42	4.56	4.00	3.80	3.67	3.52	3.37	3.21	3.08	2.98
16	8.53	5.29	4.44	3.89	3.69	3.55	3.41	3.26	3.10	2.97	2.86
17	8.40	5.18	4.34	3.79	3.59	3.46	3.31	3.16	3.00	2.87	2.76
18	8.29	5.09	4.25	3.71	3.51	3.37	3.23	3.08	2.92	2.78	2.68
19	8.18	5.01	4.17	3.63	3.43	3.30	3.15	3.00	2.84	2.71	2.60
20	8.10	4.94	4.10	3.56	3.37	3.23	3.09	2.94	2.78	2.64	2.54
21	8.02	4.87	4.04	3.51	3.31	3.17	3.03	2.88	2.72	2.58	2.48
22	7.95	4.82	3.99	3.45	3.26	3.12	2.98	2.83	2.67	2.53	2.42
23	7.88	4.76	3.94	3.41	3.21	3.07	2.93	2.78	2.62	2.48	2.37
24	7.82	4.72	3.90	3.36	3.17	3.03	2.89	2.74	2.58	2.44	2.33
25	7.77	4.68	3.85	3.32	3.13	2.99	2.85	2.70	2.54	2.40	2.29
26	7.72	4.64	3.82	3.29	3.09	2.96	2.81	2.66	2.50	2.36	2.25
27	7.68	4.60	3.78	3.26	3.06	2.93	2.78	2.63	2.47	2.33	2.22
28	7.64	4.57	3.75	3.23	3.03	2.90	2.75	2.60	2.44	2.30	2.19
29	7.60	4.54	3.73	3.20	3.00	2.87	2.73	2.57	2.41	2.27	2.16
30	7.56	4.51	3.70	3.17	2.98	2.84	2.70	2.55	2.39	2.25	2.13

Tabella F.1: quantili di ordine 0.99 della legge Fisher-Snedecor $\mathtt{F}(m, n)$.

Appendice G

Risultati di Algebra Lineare

In questa Appendice illustriamo brevemente alcuni risultati di algebra lineare utili ai fini del Corso. Sia $\boldsymbol{A}$ una matrice $n \times k$ *iniettiva*, ovvero tale per cui $\boldsymbol{Ax} \neq \boldsymbol{Ay}$ quando i vettori $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^k$ siano distinti. Con abuso di notazione si confonderà la matrice $\boldsymbol{A}$ con l'operatore lineare $\boldsymbol{A}\colon \mathbb{R}^k \to \mathbb{R}^n$ che essa rappresenta. Si decomponga $\mathbb{R}^n$ nella somma diretta

$$\mathbb{R}^n = \left(\boldsymbol{A}\mathbb{R}^k\right) \oplus \left(\boldsymbol{A}\mathbb{R}^k\right)^{\perp}$$

e sia π la proiezione ortogonale di $\mathbb{R}^n$ sul sottospazio $\left(\boldsymbol{A}\mathbb{R}^k\right)$ definita dalla matrice $\boldsymbol{P}$, mentre sia $\pi^{\perp}$ la proiezione ortogonale di $\mathbb{R}^n$ sul sottospazio $\left(\boldsymbol{A}\mathbb{R}^k\right)^{\perp}$, ortogonale a $\left(\boldsymbol{A}\mathbb{R}^k\right)$, definita dalla matrice $\boldsymbol{P}^{\perp}$. Chiaramente si ha $\boldsymbol{P}^{\perp} = \boldsymbol{I}_n - \boldsymbol{P}$. Valgono i seguenti risultati.

PROPOSIZIONE G.1. *Sia $\boldsymbol{y} \in \mathbb{R}^n$ e $\boldsymbol{x} \in \mathbb{R}^k$. Allora*

$$\|\boldsymbol{y} - \boldsymbol{Ax}\|^2 = \|\boldsymbol{y}\|^2 + \|\boldsymbol{Ax}\|^2 - 2\langle \boldsymbol{Py}, \boldsymbol{Ax}\rangle .$$

DIM. Poiché $\boldsymbol{y} = \boldsymbol{Py} + \boldsymbol{P}^{\perp}\boldsymbol{y}$ si ha

$$\begin{aligned}
\|\boldsymbol{y} - \boldsymbol{Ax}\|^2 &= \langle \boldsymbol{y} - \boldsymbol{Ax}, \boldsymbol{y} - \boldsymbol{Ax}\rangle \\
&= \|\boldsymbol{y}\|^2 + \|\boldsymbol{Ax}\|^2 - 2\langle \boldsymbol{y}, \boldsymbol{Ax}\rangle \\
&= \|\boldsymbol{y}\|^2 + \|\boldsymbol{Ax}\|^2 - 2\langle \boldsymbol{Py}, \boldsymbol{Ax}\rangle - 2\left\langle \boldsymbol{P}^{\perp}\boldsymbol{y}, \boldsymbol{Ax}\right\rangle \\
&= \|\boldsymbol{y}\|^2 + \|\boldsymbol{Ax}\|^2 - 2\langle \boldsymbol{Py}, \boldsymbol{Ax}\rangle
\end{aligned}$$

da cui la tesi. □

PROPOSIZIONE G.2. *Ogni proiezione ortogonale π in $\mathbb{R}^n$ è simmetrica, ovvero $\pi = \pi^{\intercal}$.*

DIM. Per ogni coppia $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ si ha

$$\langle \pi\boldsymbol{x}, \boldsymbol{y}\rangle = \langle \pi\boldsymbol{x}, \pi\boldsymbol{y}\rangle + \left\langle \pi\boldsymbol{x}, \pi^{\perp}\boldsymbol{y}\right\rangle = \langle \pi\boldsymbol{x}, \pi\boldsymbol{y}\rangle$$

e

$$\langle \boldsymbol{x}, \pi\boldsymbol{y}\rangle = \langle \pi\boldsymbol{x}, \pi\boldsymbol{y}\rangle + \left\langle \pi^{\perp}\boldsymbol{x}, \pi\boldsymbol{y}\right\rangle = \langle \pi\boldsymbol{x}, \pi\boldsymbol{y}\rangle .$$

Perciò è $\langle \pi\boldsymbol{x}, \boldsymbol{y}\rangle = \langle \boldsymbol{x}, \pi\boldsymbol{y}\rangle$, vale a dire $\pi = \pi^{\intercal}$. □

PROPOSIZIONE G.3. *Sia $\boldsymbol{A}$ una matrice $n \times k$ iniettiva e sia $\boldsymbol{y} \in \mathbb{R}^n$. Allora $\|\boldsymbol{y} - \boldsymbol{Ax}\|^2$, con $\boldsymbol{x} \in \mathbb{R}^k$, raggiunge il minimo in $\boldsymbol{x} = \boldsymbol{Uy}$, ove $\boldsymbol{U}$ è la matrice $k \times n$ data da*

$$\boldsymbol{U} = \left(\boldsymbol{A}^{\intercal}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\intercal};$$

dunque

$$\min_{\boldsymbol{x} \in \mathbb{R}^k} \|\boldsymbol{y} - \boldsymbol{Ax}\|^2 = \|\boldsymbol{y} - \boldsymbol{AUy}\|^2 . \tag{G.1}$$

DIM. Sia $\boldsymbol{P}$ la proiezione ortogonale di $\mathbb{R}^n$ su $\left(\boldsymbol{A}\mathbb{R}^k\right)$. Allora

$$\begin{aligned}\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|^2 &= \left\langle \boldsymbol{P}\boldsymbol{y}+\boldsymbol{P}^\perp\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}, \boldsymbol{P}\boldsymbol{y}+\boldsymbol{P}^\perp\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\right\rangle \\ &= \left\|\boldsymbol{P}^\perp\boldsymbol{y}\right\|^2 + \|\boldsymbol{P}\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|^2 + 2\left\langle \boldsymbol{P}^\perp\boldsymbol{y}, \boldsymbol{P}\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\right\rangle \\ &= \left\|\boldsymbol{P}^\perp\boldsymbol{y}\right\|^2 + \|\boldsymbol{P}\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|^2 .\end{aligned}$$

Da quest'ultima relazione si evince che il minimo di $\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|^2$, al variare di $\boldsymbol{x}$ in $\mathbb{R}^k$, si ha quando

$$\boldsymbol{P}\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}.$$

Posto $\boldsymbol{x} = \boldsymbol{U}\boldsymbol{y}$ si ha $\boldsymbol{P} = \boldsymbol{A}\boldsymbol{U}$, una relazione che basta a determinare $\boldsymbol{U}$. Infatti, dalla definizione di $\boldsymbol{P}$ quale proiezione ortogonale su $(\boldsymbol{A}\mathbb{R}^k)$, scende che $\boldsymbol{P}\boldsymbol{A} = \boldsymbol{A}$, onde

$$\boldsymbol{A}^\intercal = \boldsymbol{A}^\intercal\boldsymbol{P}^\intercal = \boldsymbol{A}^\intercal\boldsymbol{P} = \boldsymbol{A}^\intercal\boldsymbol{A}\boldsymbol{U}.$$

La matrice $\boldsymbol{A}^\intercal\boldsymbol{A}$ è quadrata, di dimensione $k\times k$, ed è invertibile: infatti $\langle \boldsymbol{A}^\intercal\boldsymbol{A}\boldsymbol{z}, \boldsymbol{z}\rangle = \langle \boldsymbol{A}\boldsymbol{z}, \boldsymbol{A}\boldsymbol{z}\rangle \neq 0$ se $\boldsymbol{z}\neq\boldsymbol{0}$, in virtù della iniettività di $\boldsymbol{A}$. Perciò $\boldsymbol{U} = (\boldsymbol{A}^\intercal\boldsymbol{A})^{-1}\boldsymbol{A}^\intercal$. □

Il risultato della *Proposizione (G.3)* è importante e merita di essere approfondito. Alla discussione è opportuno premettere il seguente risultato.

PROPOSIZIONE G.4. *Sia E un sottospazio di $\mathbb{R}^n$ e sia π la proiezione ortogonale di $\mathbb{R}^n$ su E. Allora, se $\boldsymbol{y}\in\mathbb{R}^n$,*

$$\boldsymbol{z} = \pi\,\boldsymbol{y} \tag{G.2}$$

è il vettore di E che si trova a distanza minima da $\boldsymbol{y}$.

DIM. Sia $\boldsymbol{z}$ un generico vettore in E e si scriva $\boldsymbol{y} = \boldsymbol{y}_1 + \boldsymbol{y}_2$, con $\boldsymbol{y}_1\in E$ e $\boldsymbol{y}_2\in E^\perp$. Allora

$$\begin{aligned}\|\boldsymbol{z}-\boldsymbol{y}\|^2 &= \|(\boldsymbol{z}-\boldsymbol{y}_1)-\boldsymbol{y}_2\|^2 \\ &= \langle(\boldsymbol{z}-\boldsymbol{y}_1)-\boldsymbol{y}_2, (\boldsymbol{z}-\boldsymbol{y}_1)-\boldsymbol{y}_2\rangle \\ &= \|\boldsymbol{z}-\boldsymbol{y}_1\|^2 + \|\boldsymbol{y}_2\|^2 - 2\langle\boldsymbol{z}-\boldsymbol{y}_1, \boldsymbol{y}_2\rangle \\ &= \|\boldsymbol{z}-\boldsymbol{y}_1\|^2 + \|\boldsymbol{y}_2\|^2 .\end{aligned}$$

Da quest'ultima relazione si evince che $\|\boldsymbol{z}-\boldsymbol{y}\|^2$ è minima quando $\boldsymbol{z}=\boldsymbol{y}_1$, cioè la tesi. □

Ora, si scomponga la matrice $\boldsymbol{A}$ nei suoi vettori colonna $\boldsymbol{a}_1,\ldots,\boldsymbol{a}_k$; evidentemente questi sono linearmente indipendenti, essendo $\boldsymbol{A}$ iniettiva. Se $\boldsymbol{x}$ varia in $\mathbb{R}^k$, allora il vettore

$$\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x} = \boldsymbol{a}_1x_1 + \cdots + \boldsymbol{a}_kx_k$$

varia in tutto il sottospazio $E = (\boldsymbol{A}\mathbb{R}^k)$ generato dai vettori $\boldsymbol{a}_1,\ldots,\boldsymbol{a}_k$. Se $\boldsymbol{y}\in\mathbb{R}^n$, dalla *Proposizione (G.4)* scende che il vettore di E avente distanza minima da $\boldsymbol{y}$ è dato dalla proiezione ortogonale di $\boldsymbol{y}$ su E. D'altro canto, dalla *Proposizione (G.3)* sappiamo che tale minimo si verifica in corrispondenza di $\boldsymbol{x} = \boldsymbol{U}\boldsymbol{y}$; dunque, $\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}\boldsymbol{U}\boldsymbol{y} = \boldsymbol{P}\boldsymbol{y}$ è la proiezione ortogonale di $\boldsymbol{y}$ su E.

PROPOSIZIONE G.5. *Se $\mathbb{R}^n = E\oplus E^\perp$, ove E è un sottospazio di $\mathbb{R}^n$ di dimensione k, e se π è la proiezione ortogonale di $\mathbb{R}^n$ su E, si ha*

$$\mathrm{Tr}(\pi) = k \quad e \quad \|\pi\|^2 = k.$$

DIM. Dimostreremo solo la prima eguaglianza. La proiezione ortogonale π è idempotente, cioè $\pi^2=\pi$, e dunque gli autovalori di π sono tutti uguali a 1 o a 0: infatti, dalla relazione $\pi\boldsymbol{x} = \lambda\boldsymbol{x}$, con $\boldsymbol{x}\neq\boldsymbol{0}_n$, scende

$$\lambda\boldsymbol{x} = \pi\boldsymbol{x} = \pi^2\boldsymbol{x} = \pi(\pi\boldsymbol{x}) = \lambda\pi\boldsymbol{x} = \lambda^2\boldsymbol{x},$$

sicché $\lambda=\lambda^2$, cioè $\lambda=0$ oppure $\lambda=1$. La molteplicità dell'autovalore 1 è pari alla dimensione del sottospazio E sul quale π proietta. La matrice $\boldsymbol{P}$ corrispondente a π è congruente ad una matrice diagonale $\boldsymbol{D}$ avente k degli elementi diagonali uguali a 1 e i rimanenti $n-k$ uguali a 0, ovvero esiste una matrice ortogonale $\boldsymbol{Q}$, con $\boldsymbol{Q}^{-1} = \boldsymbol{Q}^\intercal$, tale che

$$\boldsymbol{P} = \boldsymbol{Q}^\intercal\boldsymbol{D}\boldsymbol{Q} = \boldsymbol{Q}^{-1}\boldsymbol{D}\boldsymbol{Q}.$$

Ora, poiché matrici congruenti hanno gli stessi autovalori, e quindi la stessa traccia, si ha $\mathrm{Tr}(\boldsymbol{P}) = \mathrm{Tr}(\boldsymbol{D}) = k$. □

Appendice H

Complementi sul p-value

In questa Appendice illustriamo brevemente la nozione di "p-value", che gioca un ruolo fondamentale in Statistica Matematica. Quando si esegue una verifica di ipotesi, si fissa a priori il livello α del test, che concorre ad individuare la regione critica di rifiuto dell'ipotesi $\mathcal{H}_0$. Nella pratica generalmente si usano i valori $\alpha = 0.10, 0.05, 0.01$. Una volta raccolto il campione si esegue il test: se il valore della statistica test utilizzata appartiene alla regione critica, allora si rifiuta l'ipotesi $\mathcal{H}_0$ e si dice che "*il test è significativo all'*α%".

Al fine di mettere in condizione altri ricercatori che non hanno eseguito il test di poter valutare "come / per quanto" si sia ritenuto di rifiutare o accettare l'ipotesi $\mathcal{H}_0$, è possibile fornire una indicazione più analitica circa il risultato del test stesso. A tale scopo tradizionalmente si usa il "*p-value*", definito come *il minimo livello di significatività* α *del test per il quale si rifiuterebbe l'ipotesi nulla*. Dunque non stupisce che il p-value venga anche detto *livello di significatività osservato*, e sia talora indicato con α_{oss}.

In generale, nella pratica si adotta la seguente strategia:

- se $\alpha_{\text{oss}} > 0.05$ si accetta l'ipotesi $\mathcal{H}_0$;
- se $0.01 < \alpha_{\text{oss}} < 0.05$ si è in presenza di una situazione di indecidibilità, con una certa propensione al rifiuto dell'ipotesi $\mathcal{H}_0$;
- se $\alpha_{\text{oss}} < 0.01$ si rifiuta l'ipotesi $\mathcal{H}_0$.

Quale illustrazione si consideri lo studio dei dati di acqua alta a Venezia dell'*Esempio (11.1)*, analizzati tramite il modello Normale condizionale nella *Sezione 11.1.3*. La *Tabella (11.3)* mostra, rispettivamente, i valori t_{θ_0} e t_{θ_1} delle statistiche T_{θ_0} e T_{θ_1}, e i p-value relativi alla verifica delle ipotesi $\mathcal{H}_0 \colon \theta_0 = 0$ e $\mathcal{H}_0 \colon \theta_1 = 0$. Come già discusso nel testo, in entrambi i casi l'ipotesi nulla può essere decisamente rifiutata.

Un modo alternativo di vedere il problema è il seguente. Si consideri ad esempio la statistica T_{θ_1}, che assume il valore $t^* = t_{\theta_1} \approx 3.0210$. Ora, a quale livello "critico" di significatività α^* tale valore corrisponde? O, meglio, a quale quantile "critico" $t_{n-2,1-\alpha^*/2}$ (con $n - 2 = 48$ nel caso in esame) tale valore corrisponde? Grazie alle Tavole della legge t-Student è immediato calcolare $\alpha^* \approx 0.0040309$ e $1 - \alpha^*/2 \approx 0.99798455$. Dunque, l'ipotesi $\mathcal{H}_0$ potrebbe essere accettata solo a livelli non inferiori al 99.798455%, decisamente superiori a quelli tradizionalmente utilizzati nella pratica. In altri termini, scelto comunque $\alpha > \alpha^*$ (ad esempio, i valori usuali $\alpha = 0.10, 0.05, 0.01$), si avrebbe $1 - \alpha/2 < 1 - \alpha^*/2$ e $t^* > t_{n-2,1-\alpha/2}$, come mostrato nella tabella che segue

$\boldsymbol{\alpha}$	$\boldsymbol{1 - \alpha/2}$	$\boldsymbol{t_{48,1-\alpha/2}}$
0.10	0.95	1.67722
0.05	0.975	2.01063
0.01	0.995	2.68220
0.0040309	0.99798455	3.0210

e quindi si dovrebbe rifiutare l'ipotesi nulla al livello α.

Bibliografia

[1] P. Baldi. *Calcolo delle probabilità e statistica*. McGraw-Hill, second edition, 1998.

[2] R. Bartoszyński and M. Niewiadomska-Bugaj. *Probability and statistical inference*. J. Wiley & Sons, 1996.

[3] J. Beirlant, J. L. Teugels, and P. Vynckier. *Practical analysis of extreme values*. Leuven University Press, Leuven, 1996.

[4] P. Billingsley. *Probability and measure*. J. Wiley & Sons, third edition, 1995.

[5] W. M. Bolstad. *Introduction to Bayesian Statistics*. J. Wiley & Sons, 2004.

[6] G. Casella and R.L. Berger. *Statistical inference*. Duxbury Press, 1990.

[7] E. Castillo. *Extreme value theory in engineering*. Academic Press, 1988.

[8] S. Coles. *An introduction to statistical modeling of extreme values*. Springer, 2001.

[9] H. A. David. *Order statistics*. J. Wiley & Sons, 1970.

[10] M.H. DeGroot. *Probability and statistics*. Addison-Wesley, 1989.

[11] M. Falk, J. Hüsler, and R. Reiss. *Laws of small numbers: extremes and rare events*. Birkhäuser, Basel, 1994.

[12] G. R. Grimmett and D. R. Stirzaker. *Probability and random processes*. Oxford University Press, second edition, 1992.

[13] T.P. Hutchinson and C.D. Lai. *Continuous bivariate distributions, emphasising applications*. Rumsby Scientific Publishing, 1990.

[14] H. Joe. Multivariate concordance. *J. Multivar. Anal.*, 35:12–30, 1990.

[15] H. Joe. *Multivariate models and dependence concepts*. Chapman & Hall, 1997.

[16] K.-R. Koch. *Introduction to Bayesian Statistics*. Springer, second edition, 2007.

[17] N.T. Kottegoda and R. Rosso. *Probability, statistics, and reliability for civil and environmental engineers*. McGraw-Hill, 1997.

[18] S. Kotz and S. Nadarajah. *Extreme value distributions: theory and applications*. Imperial College Press, 2000.

[19] P. M. Lee. *Bayesian Statistics: An Introduction*. Wiley, third edition, 2004.

[20] R.B. Nelsen. Nonparametric measures of multivariate association. In L. Rüschendorf, B. Schweizer, and M.D. Taylor, editors, *Distributions with Fixed Marginals and Related Topics*. Institute of Mathematical Statistics, Hayward, CA, 1996.

[21] R.B. Nelsen. *An introduction to copulas*. Springer-Verlag, second edition, 2006.

[22] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. *Numerical recipes in Fortran*. Cambridge University Press, 1989.

[23] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. *Numerical recipes in C*. Cambridge University Press, second edition, 1992.

[24] V.K. Rohatgi. *An introduction to probability theory and mathematical statistics*. J. Wiley & Sons, 1976.

[25] F.J. Rohlf and R.R. Sokal. *Statistical Tables*. W.H. Freeman & Company, third edition, 1995.

[26] G. Salvadori, C. De Michele, N.T. Kottegoda, and R. Rosso. *Extremes in nature. An approach using copulas*, volume 56 of *Water Science and Technology Library*. Springer, 2007.

[27] R. Saunders and P. Laud. The multidimensional Kolmogorov goodness-of-fit test. *Biometrika*, 67(1):237, 1980.

[28] M. J. Schervish. *Theory of statistics*. Springer-Verlag, 1995.

[29] F. Schmid and R. Schmidt. Multivariate conditional versions of Spearman's rho and related measures of tail dependence. *J. Multivar. Anal.*, 98(6):1123–1140, 2007.

[30] B. Schweizer and A. Sklar. *Probabilistic metric spaces*. North-Holland, 1983.

[31] A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231, 1959.

[32] R.L. Smith. Extreme Value Theory based on the r largest annual events. *Journal of Hydrology*, 86(1/2):27–43, 1986.

[33] S. Weisberg. *Applied linear regression*. J. Wiley & Sons, second edition, 1985.

[34] D. Williams. *Probability with Martingales*. Cambridge University Press, 1991.

[35] J. Yan and I. Kojadinovic. *Copula: Multivariate dependence with copulas. R package version 0.8-12*. http://cran.r-project.org, 2009.