Inferenza Statistica

a.a. 2024-2025

Fulvio De Santis, Stefania Gubbiotti, Luca Tardella

28 maggio 2025

Appunti dei corsi Inferenza statistica.
Dipartimento di Scienze Statistiche – Sapienza Università di Roma
A cura di F. De Santis, S. Gubbiotti, L. Tardella

Indice

Capitolo 1

Le basi dell'inferenza statistica

Il capitolo introduce gli elementi base dell'inferenza statistica. A partire da alcuni semplici esempi, viene formulato il problema generale dell'inferenza statistica e presentati i concetti di modello statistico e di campione. Illustriamo quindi i principali modelli per variabili aleatorie discrete e assolutamente continue, con particolare attenzione al caso di campioni casuali. Presentiamo inoltre alcune classi di modelli di particolare rilevanza teorica e applicativa: i modelli di posizione e scala e la classe dei modelli che costituiscono famiglie esponenziali. Il capitolo si chiude con la schematizzazione dei principali problemi e impostazioni dell'inferenza statistica.

1.1 Premessa

1.1.1 Esempi introduttivi

Le procedure dell'inferenza statistica si applicano quando i dati osservati relativi a un fenomeno di interesse sono un sottoinsieme di quelli potenzialmente osservabili. In questi casi si dispone di un *campione* di osservazioni, attraverso il quale si vogliono trarre informazioni generali sulla popolazione da cui i dati stessi provengono.

1.1 Esempio (Proiezioni Elettorali). Siano A e B le due liste in competizione in una elezione politica. Al momento della chiusura delle urne, siamo interessati a stabilire quale sia stata, nella popolazione degli elettori, la proporzione θ di coloro che hanno votato per la lista A e la proporzione $1-\theta$ di coloro che hanno votato per la lista B. Prima di effettuare lo spoglio completo delle schede elettorali, per avere una valutazione del risultato in tempi rapidi, è consuetudine *stimare* la quantità θ considerando un sottoinsieme delle schede e calcolando la frequenza relativa dei voti assegnati a ciascuna lista in questo *campione*. □

1.2 Esempio (Studi Biometrici). Supponiamo di voler stabilire l'altezza media incognita (che indichiamo con θ) delle studentesse immatricolate in un corso di laurea molto numeroso. Si considera un sottoinsieme della popolazione femminile del corso di laurea e su ciascuna unità selezionata si rileva il carattere di interesse. Sulla base dei dati così raccolti si vuole *stimare* (ad esempio attraverso la media aritmetica dei valori osservati) l'altezza media incognita θ di tutta la popolazione. □

1.3 Esempio (Studi di Affidabilità). Per valutare la durata di vita media incognita (θ) di componenti elettronici (ad esempio di un computer) prodotti da una fabbrica, un campione di pezzi viene messo in funzione e, per ciascun oggetto coinvolto nell'esperimento, viene registrato il tempo che intercorre fino alla sua "rottura". La media dei tempi di vita

del campione può essere impiegata per *stimare* il tempo di vita media θ dei componenti prodotti dalla fabbrica. □

1.4 Esempio (Prove Cliniche). Lo studio dell'efficacia di nuovi farmaci (ad esempio per la riduzione della concentrazione di colesteorlo nel sangue) viene effettuato con esperimenti chiamati *prove cliniche*. Per molteplici ragioni, non è in genere possibile somministrare un nuovo farmaco a tutti i potenziali soggetti che ne necessiterebbero e misurare la risposta media incognita (θ) al trattamento in tutta la popolazione. Il farmaco viene assegnato quindi a un numero limitato di pazienti, in ciascuno dei quali si misura l'effetto (riduzione di concentrazione di colesterolo nel singolo soggetto). Con l'insieme dei valori rilevati sui pazienti considerati, si cerca di *stimare* l'effetto medio incognito del farmaco (riduzione media di colesterolo) per l'intera popolazione. □

Questi semplici esempi illustrano alcune situazioni nelle quali, per studiare una caratteristica di un fenomeno di interesse (θ) in tutta una popolazione, viene effettuata una rilevazione parziale. Sulla base dei risultati di questi esperimenti si valuta (attraverso una *stima*) un aspetto non noto e di interesse del fenomeno in esame.

I principali motivi per cui si ricorre all'operazione di campionamento da una popolazione di interesse sono:

Costi. Il costo di un esperimento aumenta con il numero di unità considerate.
Tempi. Le indagini totali (censimenti) richiedono tempi lunghi.
Accuratezza. Le indagini esaustive di grandi dimensioni presentano un'elevata frequenza di errori di varia natura che possono inficiare la qualità dei dati e delle conclusioni.
Necessità. Nell'Esempio 1.3 le rilevazioni sono distruttive. In questi casi è attuabile solo lo studio campionario, data l'impossibilità pratica di includere nella sperimentazione tutte le unità di una popolazione.

1.1.2 Campionamento da popolazioni finite e da variabili aleatorie

Negli esempi considerati si distinguono due diverse situazioni. Nei primi due esempi la popolazione di riferimento (elettori di un paese e studentesse di un certo corso di laurea) è costituita da un numero finito di individui e preesiste all'osservazione che dà luogo ai dati campionari. Viene selezionata opportunamente una piccola parte di individui (*campione*) sui quali si rileva la caratteristica di interesse (la lista politica, nell'Esempio 1.1 e l'altezza nell'Esempio 1.2) che vogliamo stimare sulla base delle informazioni parziali che si ottengono dal campione. Questa caratteristica non nota di interesse prende nome di *parametro*. Si parla in questo caso di *campionamento da popolazioni finite*.
Negli Esempi 1.3 e 1.4, i dati sono generati da un *esperimento*, e non esisterebbero senza di questo. Nel primo caso l'esperimento consiste nel mettere in funzione le componenti di cui si vuole studiare la durata di vita; nel secondo consiste nel somministrare al paziente il farmaco oggetto di studio e nel rilevare la risposta del soggetto allo stesso. Il risultato dell'esperimento è di tipo aleatorio, ovvero non certo a priori e governato da una legge di probabilità. Nell'Esempio 1.3 la variabile aleatoria è la durata di vita delle componenti messe in funzione; nell'Esempio 1.4 la variabile aleatoria è la risposta al farmaco.
In generale, l'obiettivo di un esperimento è produrre dei dati con cui poter valutare gli aspetti non noti delle leggi di probabilità che governano i fenomeni di interesse. In questo contesto, che prende nome di *campionamento da variabili aleatorie* (v.a.) (o campionamento da popolazioni teoriche), il *parametro* è rappresentato dall'aspetto non noto della legge di probabilità. Nell'Esempio 1.3, il parametro incognito è il valore atteso della variabile aleatoria durata di vita; nell'Esempio 1.4 il parametro è il valore atteso della variabile aleatoria effetto del farmaco.

I metodi del campionamento e dell'inferenza per popolazioni finite sono oggetto della disciplina denominata *teoria dei campioni*. In questo testo ci occupiamo dei problemi di inferenza statistica nel caso di campionamento da variabili aleatorie.
La distinzione tra inferenza nel campionamento da popolazioni finite e da variabili casuali, pur avendo un suo preciso fondamento concettuale, può essere ricomposta in una accezione ampia di *esperimento" statistico*. Si può infatti pensare all'operazione di estrazione di unità tipica del campionamento da popolazioni finite come ad un esperimento, da intendere come "prova" il cui esito è aleatorio, e considerare i dati campionari ottenuti come generati dall'esperimento di estrazione.

1.1.3 Formulazione generale del problema

Per quanto discusso fin qui, possiamo dire che l'inferenza statistica è la disciplina che studia come utilizzare i *dati* per ricostruire il meccanismo aleatorio che li ha generati. Più precisamente, il problema dell'inferenza statistica nel campionamento da variabili aleatorie si può formulare nel modo seguente. Si considera un fenomeno aleatorio osservabile di interesse al quale associamo una variabile aleatoria, X, che può assumere valori nello spazio $\mathcal{X}$ (*spazio campionario*). Indichiamo con f_X la *legge di probabilità* di X. Se X è una v.a. discreta, f_X è una *funzione di massa di probabilità*; se X è una v.a. assolutamente continua, f_X è una *funzione di densità di probabilità*. Assumiamo inoltre di non conoscere esattamente quale sia la legge di probabilità di X ma di sapere solo che f_X appartiene ad una *famiglia* $\mathcal{F}$ di leggi di probabilità. Il problema è allora quello di individuare, tra gli elementi che costituiscono $\mathcal{F}$, la specifica legge f_X^* che regola il fenomeno aleatorio considerato. A tal fine si utilizzano i risultati osservati del fenomeno aleatorio in esame, detti *dati campionari* (o osservazioni campionarie), ovvero un numero limitato di *realizzazioni* della v.a. X. Negli esempi più comuni, $\mathcal{F}$ è una *famiglia parametrica* di leggi di probabilità:

$$\mathcal{F} = (f_X(\,\cdot\,;\theta),\ \theta \in \Theta)\,.$$

Ciò vuol dire che la legge di probabilità $f_X(\cdot;\theta)$ dipende da una quantità θ non nota, detta *parametro*, che può assumere valori nell'insieme Θ, detto *spazio parametrico*. Il parametro *indicizza* la legge di probabilità, nel senso che, al variare di θ, si ottengono gli elementi della famiglia $\mathcal{F}$: in corrispondenza di ciascun elemento di Θ si ottiene un unico membro della famiglia $\mathcal{F}$. In particolare ipotizziamo che tra gli elementi di Θ e i membri di $\mathcal{F}$ vi sia una corrispondenza biunivoca (uno-a-uno): si parla, in questo caso, di modello *identificato*. L'inferenza statistica si pone l'obiettivo di stabilire, sulla base delle osservazioni campionarie, quale elemento di $\mathcal{F}$ (ovvero quale valore θ^* del parametro θ) individua la legge di X. In altre parole, si assume di conoscere l'espressione esplicita della legge di probabilità della v.a. X che ha generato le osservazioni (ad esempio possiamo supporre che la funzione di densità sia quella di una v.a. normale, esponenziale, beta, oppure che la funzione di massa di probabilità sia quella di una v.a. bernoulliana o di Poisson), ma di non conoscere quale, tra gli elementi della famiglia, sia la legge della v.a. che ha generato i dati.

1.2 Il modello statistico di base

I tre elementi che caratterizzano dal punto di vista probabilistico una v.a. X, la cui legge di probabilità dipende da un parametro θ, sono:

- $\mathcal{X}$ insieme dei valori che la v.a. X può assumere;
- $f_X(\cdot;\theta)$ legge di probabilità;
- Θ *spazio parametrico*.

I tre elementi descritti si raccolgono nella terna

$$\{\mathcal{X}, \quad f_X(\cdot;\theta), \quad \theta \in \Theta\},$$

che chiamiamo *modello statistico di base* o *modello probabilistico* per la v.a. X. In generale, X può essere una variabile aleatoria multidimensionale e Θ uno spazio di dimensioni superiori ad uno (problemi *multiparametrici*). Tuttavia, nella maggior parte degli esempi che seguono in questo testo, X rappresenta una v.a. semplice e il parametro θ uno scalare (Θ è un sottoinsieme di $\mathbb{R}$), oppure un vettore $\theta = (\theta_1, \theta_2)$ di due elementi (Θ è allora un sottoinsieme di $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$).

Vengono ora illustrate le caratteristiche dei modelli statistici per le principali variabili aleatorie, distinguendo tra v.a. discrete e v.a. assolutamente continue.

1.2.1 Modelli statistici parametrici per v.a. discrete

Le v.a. discrete assumono un numero finito o al più numerabile di valori, che sono sono quindi in corrispondenza biunivoca con l'insieme dei numeri naturali, o con un suo sottoinsieme. La funzione $f_X(\cdot;\theta)$ indica in questo caso la *funzione di massa di probabilità* della v.a. X. Nel caso di un numero k finito di valori[1], che indichiamo con $x_1, x_2, \ldots, x_k$, la *distribuzione di probabilità* di X è data da

x_1	x_2	$\ldots$	x_k
$f_X(x_1;\theta)$	$f_X(x_2;\theta)$	$\ldots$	$f_X(x_k;\theta)$

dove

$$f_X(x_i;\theta) = \mathbb{P}(X = x_i;\theta), \qquad i = 1, 2, \ldots, k,$$

è la probabilità che X sia uguale a x_i, che dipende dal parametro incognito θ. Per la v.a. X con distribuzione di probabilità data dalla tabella di cui sopra, *valore atteso* e *varianza*[2] sono definiti rispettivamente da:

$$\mathbb{E}_\theta[X] = \sum_{i=1}^{k} x_i f_X(x_i;\theta), \qquad \mathbb{V}_\theta[X] = \sum_{i=1}^{k} (x_i - \mathbb{E}_\theta[X])^2 f_X(x_i;\theta).$$

La notazione utilizzata rende esplicita la dipendenza del valore atteso e della varianza della v.a. X dal parametro incognito θ.

1.5 Esempio (Modello bernoulliano). Per la v.a. discreta X *bernoulliana* di parametro θ, i tre elementi caratterizzanti sono

- $\mathcal{X} = \{0, 1\}$;
- $f_X(x;\theta) = \mathbb{P}(X = x;\theta) = \theta^x (1-\theta)^{1-x}, \qquad x \in \mathcal{X}$;
- $\Theta = [0, 1]$.

Osserviamo che con la notazione $f_X(x;\theta) = \theta^x(1-\theta)^{1-x}$, $x \in \mathcal{X} = \{0, 1\}$ intendiamo dire che

$$f_X(x;\theta) = \begin{cases} \theta^x (1-\theta)^{1-x} & x \in \{0, 1\} \\ 0 & x \notin \{0, 1\} \end{cases}$$

ovvero che

$$f_X(x;\theta) = \theta^x (1-\theta)^{1-x} I_{\{0,1\}}(x),$$

[1] L'estensione al caso di una infinità numerabile di modalità è immediata.

[2] Nel caso di v.a. discrete che possono assumere una infinità numerabile di valori, come la v.a. di Poisson, valore atteso e varianza esistono se le corrispondenti serie che li definiscono sono convergenti.

dove $I_{\{0,1\}}(x) = I_{\mathcal{X}}(x)$ denota la funzione indicatrice[3] dell'insieme $\mathcal{X}$, supporto della v.a. X. Il modello statistico di base è quindi individuato dalla terna

$$\left\{\mathcal{X} = \{0,1\}, \quad f_X(x;\theta) = \theta^x(1-\theta)^{1-x} I_{\{0,1\}}(x), \quad \theta \in \Theta = [0,1]\right\}.$$

Il valore atteso e la varianza della v.a. bernoulliana sono:

$$\mathbb{E}_\theta[X] = \theta, \qquad \mathbb{V}_\theta[X] = \theta(1-\theta).$$

Nel seguito indicheremo che una v.a. discreta X ha distribuzione bernoulliana di parametro θ con la notazione: $X|\theta \sim \text{Ber}(\theta)$. Notazioni simili vengono utilizzate anche per le altre variabili aleatorie. □

1.6 Esempio (Modello binomiale). Se $X_1, \ldots, X_k$ sono k v.a. bernoulliane indipendenti di parametro θ, la v.a.

$$X = \sum_{i=1}^{k} X_i$$

prende il nome di v.a. *binomiale* di parametro θ relativa a k prove – sinteticamente: $X|\theta \sim \text{Bin}(k,\theta)$ –. La v.a. binomiale emerge pertanto dal cosiddetto *schema delle prove ripetute* e rappresenta il numero aleatorio di successi che si possono avere in k prove bernoulliane indipendenti in cui la probabilità di successo in ciascuna prova è costante e pari a θ. Per questa v.a. discreta i tre elementi caratterizzanti sono

- $\mathcal{X} = \{0,1,2\ldots,k\}$;
- $f_X(x;\theta) = \mathbb{P}(X = x;\theta) = \binom{k}{x}\theta^x(1-\theta)^{k-x}, \qquad x \in \mathcal{X}$;
- $\Theta = [0,1]$.

Il modello statistico di base è individuato dalla terna:

$$\left\{\mathcal{X} = \{0,1,2,\ldots,k\}, \quad f_X(x;\theta) = \binom{k}{x}\theta^x(1-\theta)^{k-x} I_{\{0,\ldots,k\}}(x), \quad \theta \in \Theta = [0,1]\right\}.$$

Il valore atteso e la varianza della v.a. binomiale di parametro θ sono:

$$\mathbb{E}_\theta[X] = k\,\theta, \qquad \mathbb{V}_\theta[X] = k\,\theta(1-\theta).$$

□

1.7 Osservazione. Per $k = 1$, la v.a. binomiale di parametro θ corrisponde a una bernoulliana di parametro θ. Si osservi Inoltre che, nei casi in cui non è noto neanche il valore k, il modello binomiale presenta due parametri incogniti. □

1.8 Esempio (Modello geometrico). Per la v.a. discreta X *geometrica* di parametro θ – sinteticamente: $X|\theta \sim \text{Geom}(\theta)$ – i tre elementi caratterizzanti sono

- $\mathcal{X} = \{0,1,2,\ldots\} = \overline{\mathbb{N}} = \mathbb{N} \cup \{0\}$;
- $f_X(x;\theta) = (1-\theta)^x\ \theta, \qquad x \in \mathcal{X}$;
- $\Theta = [0,1]$,

[3]Definiamo *funzione indicatrice* di un insieme A la seguente funzione:

$$I_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \in A^C \end{cases}.$$

dove $\mathbb{N} = \{1, 2, \ldots\}$ è l'insieme dei numeri naturali. Il modello statistico di base è quindi individuato dalla terna:

$$\left\{\mathcal{X} = \overline{\mathbb{N}}, \quad f_X(x;\theta) = \mathbb{P}\left(X = x;\theta\right) = (1-\theta)^x \theta I_{\overline{\mathbb{N}}}(x), \quad \theta \in \Theta = [0,1]\right\}.$$

Il valore atteso e la varianza della v.a. geometrica sono:

$$\mathbb{E}_\theta[X] = \frac{1-\theta}{\theta}, \qquad \mathbb{V}_\theta[X] = \frac{1-\theta}{\theta^2}.$$

□

Anche la v.a. geometrica, come la binomiale, può essere derivata dallo schema delle prove ripetute. Se $X_1, X_2, \ldots$ sono una successione di v.a. bernoulliane indipendenti di parametro θ, la v.a. geometrica X conta il numero totale di insuccessi (il numero di volte in cui le v.a. X_i assumono il valore zero) prima che si abbia un successo. Pertanto

$$\mathbb{P}(X = k;\theta) = \mathbb{P}(X_1 = 0, \ldots, X_k = 0, X_{k+1} = 1;\theta) = (1-\theta)^k\ \theta.$$

1.9 Esempio (Modello di Poisson). Per la v.a. discreta X di *Poisson* di parametro θ – sinteticamente: $X|\theta \sim \text{Pois}(\theta)$ – i tre elementi caratterizzanti sono

- $\mathcal{X} = \{0, 1, 2, \ldots\} = \overline{\mathbb{N}}$;
- $f_X(x;\theta) = \mathbb{P}\left(X = x;\theta\right) = \dfrac{e^{-\theta}\ \theta^x}{x!}, \qquad x \in \mathcal{X}$;
- $\Theta = \mathbb{R}^+$.

Il modello statistico di base è quindi individuato dalla terna:

$$\left\{\mathcal{X} = \overline{\mathbb{N}}, \quad f_X(x;\theta) = \mathbb{P}\left(X = x;\theta\right) = \frac{e^{-\theta}\theta^x}{x!} I_{\overline{\mathbb{N}}}(x), \quad \theta \in \Theta = \mathbb{R}^+\right\}.$$

Si tratta di una v.a. che può assumere una infinità numerabile di valori. Il valore atteso e la varianza della v.a. di Poisson coincidono e sono:

$$\mathbb{E}_\theta[X] = \mathbb{V}_\theta[X] = \theta.$$

La v.a. di Poisson viene comunemente usata per descrivere il numero aleatorio di eventi di interesse in un'unità spazio-temporale, come per esempio: il numero di chiamate ad un centralino in un dato riferimento temporale, il numero di soggetti infetti in una certa area geografica, il numero di clienti che acquistano un certo prodotto in un negozio. Storicamente la v.a. di Poisson è stata utilizzata come modello di distribuzione per eventi rari. □

1.2.2 Modelli statistici parametrici per v.a. assolutamente continue

Le v.a. assolutamente continue assumono una infinità non numerabile di valori. Gli elementi di $\mathcal{X}$ sono in corrispondenza biunivoca con i numeri reali, $\mathbb{R}$, o con un un loro sottoinsieme. In questo caso, la legge di probabilità $f_X(\cdot;\theta)$ è la *funzione di densità di probabilità* [4] della

[4] Per la definizione formale di funzione di densità di probabilità si veda, ad esempio, Dall'Aglio G. (2003). *Calcolo delle Probabilità*, Zanichelli (pag. 93). Ricordiamo che $f : \mathbb{R} \to \mathbb{R}^+$ è una funzione di densità se: $f(x) \geq 0,\ x \in \mathbb{R}$ e $\int_{\mathbb{R}} f(x)dx = 1$. Ricordiamo inoltre che il *supporto* di una v.a. è costituito dall'insieme $\mathcal{X} = \{x \in \mathbb{R} :\ f(x) > 0\}$.

v.a. X. Valore atteso e varianza di X sono definiti dai seguenti integrali, che assumiamo esistere (finiti):

$$\mathbb{E}_\theta[X] = \int_{\mathbb{R}} x f_X(x;\theta)dx, \qquad \mathbb{V}_\theta[X] = \int_{\mathbb{R}} (x - \mathbb{E}_\theta[X])^2 f_X(x;\theta)dx.$$

1.10 Esempio (Modello esponenziale negativo). Per la v.a. assolutamente continua X *esponenziale negativa* di parametro θ – sinteticamente: $X|\theta \sim \text{EN}(\theta)$ – i tre elementi caratterizzanti sono

- $\mathcal{X} = \mathbb{R}^+$;
- $f_X(x;\theta) = \theta e^{-\theta x}, \qquad x \in \mathcal{X}$;
- $\Theta = \mathbb{R}^+$.

il modello statistico di base è quindi individuato dalla terna

$$\left\{\mathcal{X} = \mathbb{R}^+, \quad f_X(x;\theta) = \theta e^{-\theta x} I_{\mathbb{R}^+}(x), \quad \theta \in \Theta = \mathbb{R}^+\right\}.$$

Il valore atteso e la varianza della v.a. esponenziale negativa sono rispettivamente uguali a:

$$\mathbb{E}_\theta[X] = \frac{1}{\theta}, \qquad \mathbb{V}_\theta[X] = \frac{1}{\theta^2}.$$

La v.a. esponenziale negativa viene utilizzata per descrivere il tempo di attesa entro il quale si verifica un evento di interesse e trova applicazione nell'analisi di affidabilità e negli studi di sopravvivenza. □

1.11 Esempio (Modello esponenziale). Una semplice variazione di parametrizzazione nel modello precedente dà luogo al modello esponenziale, che si ottiene sostituendo il parametro θ con il suo inverso $1/\theta$. Per la v.a. assolutamente continua X *esponenziale* di parametro θ – sinteticamente: $X|\theta \sim \text{Esp}(\theta)$ – il modello statistico di base è quindi individuato dalla terna:

$$\left\{\mathcal{X} = \mathbb{R}^+, \quad f_X(x;\theta) = \frac{1}{\theta} e^{-x/\theta} I_{\mathbb{R}^+}(x), \quad \theta \in \Theta = \mathbb{R}^+\right\}.$$

Il valore atteso e la varianza della v.a. esponenziale sono rispettivamente uguali a:

$$\mathbb{E}_\theta[X] = \theta, \qquad \mathbb{V}_\theta[X] = \theta^2.$$

La Figura 1.1 riporta il grafico della funzione di densità della v.a. in esame per alcuni valori del parametro.

□

1.12 Osservazione. Come vedremo meglio nel Paragrafo 1.2.4, nel caso del modello esponenziale θ è un parametro di scala (o `scale`, in inglese) mentre nel caso del modello esponenziale negativo θ è detto parametro di tasso (`rate` in inglese). □

1.13 Esempio (Modello normale). Per la v.a. assolutamente continua X *normale* di parametro vettoriale $\boldsymbol{\theta} = (\theta_1, \theta_2)$ – notazione sintetica: $X|\boldsymbol{\theta} \sim \text{N}(\theta_1, \theta_2)$ – i tre elementi caratterizzanti sono

- $\mathcal{X} = \mathbb{R}$;
- $f_X(x;\boldsymbol{\theta}) = \dfrac{1}{\sqrt{2\pi\theta_2}} \exp\left\{-\dfrac{1}{2\theta_2}(x-\theta_1)^2\right\}, \quad x \in \mathbb{R}$;
- $\Theta = \Theta_1 \times \Theta_2, \quad \Theta_1 = \mathbb{R}, \quad \Theta_2 = \mathbb{R}^+$.

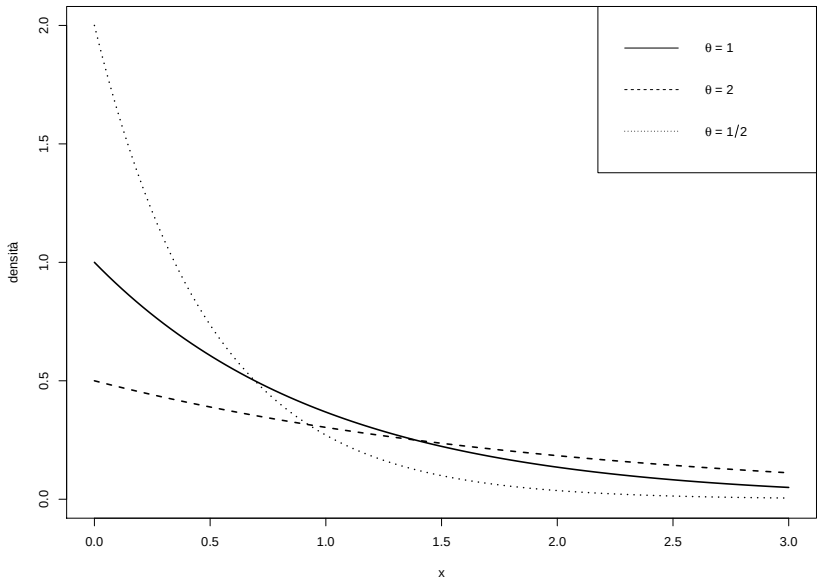

Figura 1.1: *Grafico della funzione di densità di una v.a. esponenziale (per alcuni valori del parametro di scala θ).*

Il modello statistico di base è quindi individuato dalla terna:

$$\left\{\mathcal{X} = \mathbb{R}, \quad f_X(x;\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\theta_2}}\ \exp\left\{-\frac{1}{2\theta_2}(x-\theta_1)^2\right\} I_{\mathbb{R}}(x), \quad \boldsymbol{\theta} = (\theta_1, \theta_2) \in \mathbb{R}\times\mathbb{R}^+\right\}.$$

Il valore atteso e la varianza della v.a. normale sono:

$$\mathbb{E}_\theta[X] = \theta_1, \qquad \mathbb{V}_\theta[X] = \theta_2.$$

La v.a. normale viene utilizzata in moltissimi contesti. Ad esempio, può essere sfruttata per descrivere misure biomediche, quali la pressione sanguigna o durata di una gravidanza, oppure le misure ripetute di una stessa grandezza fisica. □

Dal modello generale della v.a. normale si ottengono due casi particolari. Il primo si ha quando è noto il paramero θ_2 e incognito il valore atteso θ_1; il secondo, viceversa, quando è noto il valore atteso θ_1 e incognita la varianza, θ_2.

1.14 Esempio (Modello normale con varianza nota). Assumiamo ora che la varianza di X sia nota e pari a σ_0^2. In questo caso il solo parametro incognito è $\theta_1 = \theta$ e il modello statistico di base è individuato dalla terna:

$$\left\{\mathcal{X} = \mathbb{R}, \quad f_X(x;\theta) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left\{-\frac{1}{2\sigma_0^2}(x-\theta)^2\right\} I_{\mathbb{R}}(x), \quad \theta \in \Theta = \mathbb{R}\right\}.$$

La Figura (1.2) riporta il grafico della funzione di densità della v.a. in esame per alcuni valori del parametro θ (valore atteso della v.a. X). □

1.15 Esempio (Modello normale con valore atteso noto). Se invece il valore atteso di X è noto e pari a μ_0, allora il solo parametro incognito è $\theta_2 = \theta$ e il modello statistico di

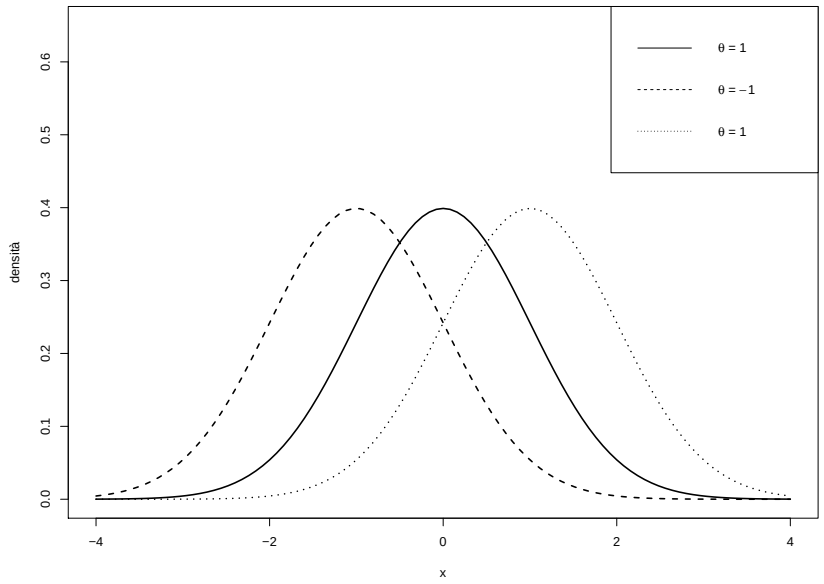

Figura 1.2: *Grafico della funzione di densità di una v.a. normale per alcuni valori del valore atteso θ e con varianza $\sigma_0^2 = 1$.*

base è individuato dalla terna:

$$\left\{\mathcal{X} = \mathbb{R}, \quad f_X(x;\theta) = \frac{1}{\sqrt{2\pi\theta}} \exp\left\{-\frac{1}{2\theta}(x-\mu_0)^2\right\} I_{\mathbb{R}}(x), \quad \theta \in \Theta = \mathbb{R}^+\right\}.$$

La Figura (1.3) riporta il grafico della funzione di densità della v.a. in esame per alcuni valori del parametro θ (varianza della v.a. X). □

1.16 Esempio (Modello gamma). Per la v.a. assolutamente continua X con distribuzione *gamma* di parametro $\boldsymbol{\theta} = (\theta_1, \theta_2)$ – sinteticamente: $X|\boldsymbol{\theta} \sim \text{Gamma}(\theta_1, \theta_2)$ – i tre elementi caratterizzanti sono[5]:

- $\mathcal{X} = \mathbb{R}^+$;
- $f_X(x;\boldsymbol{\theta}) = \dfrac{\theta_2^{\theta_1}}{\Gamma(\theta_1)} x^{\theta_1 - 1} e^{-\theta_2 x}, \qquad x \in \mathbb{R}^+$;
- $\Theta = \mathbb{R}^+ \times \mathbb{R}^+$.

Il modello statistico di base è quindi individuato dalla terna:

$$\left\{\mathcal{X} = \mathbb{R}^+, \quad f_X(x;\boldsymbol{\theta}) = \frac{\theta_2^{\theta_1}}{\Gamma(\theta_1)} x^{\theta_1 - 1} e^{-\theta_2 x} I_{\mathbb{R}^+}(x), \quad \boldsymbol{\theta} = (\theta_1, \theta_2) \in \Theta = \mathbb{R}^+ \times \mathbb{R}^+ \right\}.$$

[5]Si ricordi che la funzione gamma $\Gamma: \mathbb{R}^+ \mapsto \mathbb{R}^+$ è definita come segue: $\Gamma(y) = \int_0^{+\infty} t^{y-1} e^{-t} dt, \ y > 0$. Tale funzione soddisfa la relazione ricorsiva $\Gamma(y+1) = y\Gamma(y)$. Nel caso in cui si consideri un numero intero n, si ha: $\Gamma(n) = (n-1)!$

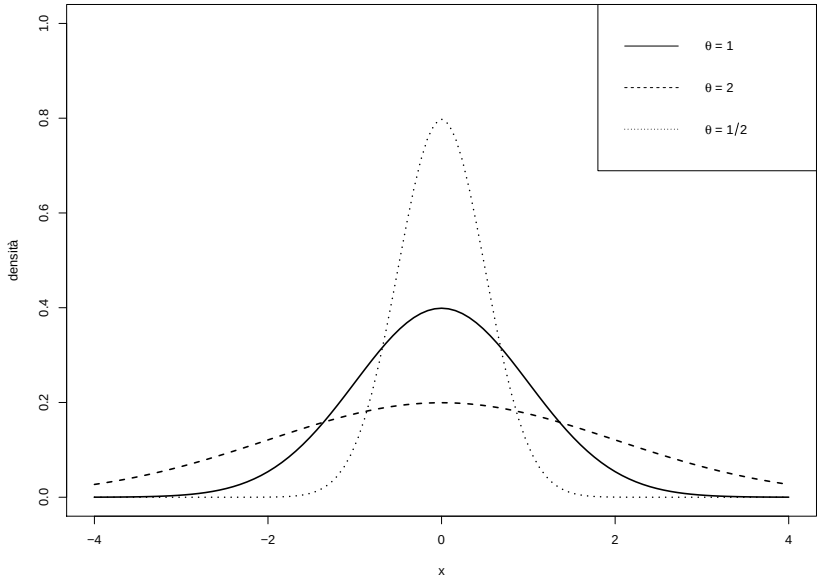

Figura 1.3: *Grafico della funzione di densità di una v.a. normale per alcuni valori della varianza θ e con valore atteso $\mu_0 = 0$.*

l valore atteso e la varianza della v.a. gamma sono:

$$\mathbb{E}_\theta[X] = \frac{\theta_1}{\theta_2}, \qquad \mathbb{V}_\theta[X] = \frac{\theta_1}{\theta_2^2}.$$

La Figura (1.4) riporta il grafico della funzione di densità della v.a. in esame per alcuni valori dei parametri θ_1 e θ_2. □

1.17 Osservazione.

1. La distribuzione gamma viene comunemente utilizzata per rappresentare il tempo aleatorio entro il quale si verifica un evento di interesse. Analogamente alla distribuzione esponenziale, trova molte applicazione negli studi di sopravvivenza.

2. La distribuzione Gamma$(1, \theta)$ coincide con la v.a. EN(θ).

3. Se $X_1, \ldots, X_k$ sono k v.a. Gamma$(\theta_{1,i}, \theta_2)$ indipendenti e non somiglianti, ciascuna con parametro $\theta_{1,i}$, la v.a. $\sum_{i=1}^k X_i \sim$ Gamma$(\sum_{i=1}^k \theta_{1,i}, \theta_2)$. Se le k v.a. sono anche somiglianti ($\theta_{1,i} = \theta_1$, $i = 1, \ldots, k$), allora $\sum_{i=1}^k X_i \sim$ Gamma$(k\theta_1, \theta_2)$. Per la dimostrazione si veda il Cap. 2.

4. Se $X_1, \ldots, X_k$ sono k v.a. EN(θ) indipendenti, la v.a. $\sum_{i=1}^k X_i \sim$ Gamma(k, θ).

5. A volte viene usata una parametrizzazione diversa da quella riportata. In particolare il parametro θ_2 (qui è un parametro di `rate`) è sostituito dal suo inverso. In questo caso la funzione di densità della v.a. Gamma(θ_1, θ_2) risulta essere:

$$f_X(x; \boldsymbol{\theta}) = \frac{1}{\Gamma(\theta_1)\, \theta_2^{\theta_1}} x^{\theta_1 - 1} e^{-\frac{x}{\theta_2}}, \qquad x > 0,$$

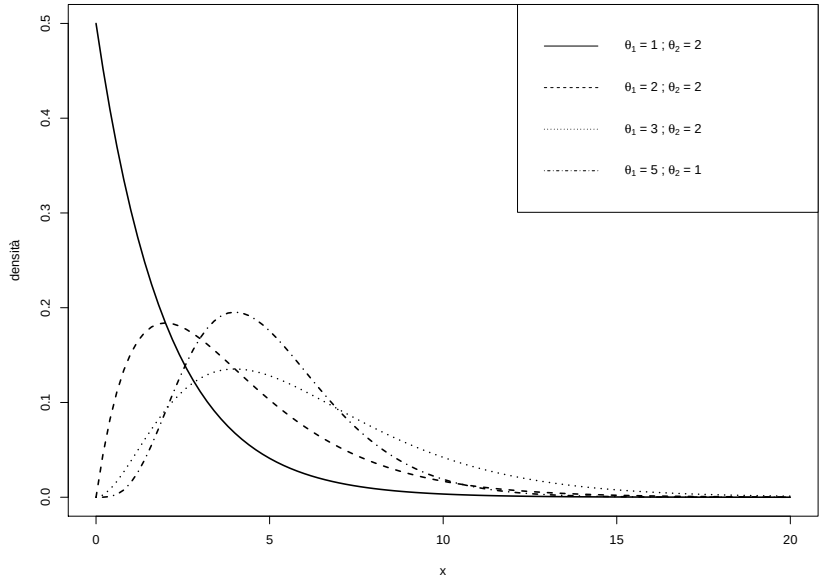

Figura 1.4: *Grafico della funzione di densità di una v.a. gamma per alcuni valori dei parametri θ_1 e θ_2 (con θ_2 parametro di scala).*

e il parametro θ_2 assume il ruolo di *parametro di scala* (in inglese `scale`, vedi paragrafo 1.2.4). Inoltre

$$\mathbb{E}_\theta[X] = \theta_1\theta_2, \qquad \mathbb{V}_\theta[X] = \theta_1\theta_2^2.$$

6. Se si utilizza la parametrizzazione appena introdotta e se si pone $\theta_1 = \frac{\nu}{2}$, $\nu > 0$, e $\theta_2 = 2$, si ottiene la funzione di densità della v.a. Chi quadrato di parametro $\nu > 0$:

$$f_X(x;\nu) = \frac{1}{\Gamma(\frac{\nu}{2})2^{\nu/2}}\, x^{(\nu/2)-1}\, e^{-x/2} I_{\mathbb{R}^+}(x).$$

Per il parametro ν si usa la denominazione di *gradi di libertà* e inoltre si ha:

$$E_\nu[X] = \nu, \qquad V_\nu[X] = 2\nu.$$

7. Nel seguito, quando necessario, indicheremo in modo esplicito se il secondo parametro della v.a. gamma rappresenta un parametro `rate` o `scale`, usando le seguenti notazioni: $\text{Gamma}(\theta_1, \texttt{rate} = \theta_2)$ e $\text{Gamma}(\theta_1, \texttt{scale} = \theta_2)$. Di conseguenza potremo scrivere che $\text{EN}(\theta) = \text{Gamma}(1, \texttt{rate} = \theta) = \text{Gamma}(1, \texttt{scale} = 1/\theta)$ e $\text{Esp}(\theta) = \text{Gamma}(1, \texttt{scale} = \theta) = \text{Gamma}(1, \texttt{rate} = 1/\theta)$.

8. Nel seguito useremo le seguenti proprietà della v.a. Gamma, che si verificano facilmente (si veda il Cap. 2). Se $a > 0$, allora

$$X \sim \text{Gamma}(\theta_1, \texttt{scale} = \theta_2) \quad \Longrightarrow \quad aX \sim \text{Gamma}(\theta_1, \texttt{scale} = a\theta_2).$$

e

$$X \sim \text{Gamma}(\theta_1, \texttt{rate} = \theta_2) \quad \Longrightarrow \quad aX \sim \text{Gamma}(\theta_1, \texttt{rate} = \theta_2/a).$$

9. Integrale gamma. Sfruttando la parametrizzazione `rate` abbiamo

$$\int_0^\infty x^{\theta_1-1}e^{-\theta_2 x}dx = \frac{\Gamma(\theta_1)}{\theta_2^{\theta_1}}.$$

Con la parametrizzazione `scale` otteniamo

$$\int_0^\infty x^{\theta_1-1}e^{-\frac{x}{\theta_2}}dx = \Gamma(\theta_1)\theta_2^{\theta_1}.$$

□

1.18 Esempio (Modello di Cauchy). Una v.a. assolutamente continua X ha distribuzione di *Cauchy* di parametro $\boldsymbol{\theta} = (\theta_1, \theta_2)$, con $\theta_1 \in \mathbb{R}$ e $\theta_2 \in \mathbb{R}^+$ – sinteticamente $X|\boldsymbol{\theta} \sim \text{Cau}(\theta_1, \theta_2)$ – se i tre elementi caratterizzanti del modello sono

- $\mathcal{X} = \mathbb{R}$,
- $f_X(x;\boldsymbol{\theta}) = \dfrac{1}{\pi\theta_2\left[1+\left(\frac{x-\theta_1}{\theta_2}\right)^2\right]}, \qquad x \in \mathbb{R};$
- $\Theta = \left\{(\theta_1, \theta_2) \in \mathbb{R} \times \mathbb{R}^+\right\}.$

Il modello statistico di base è quindi individuato dalla terna:

$$\left\{\mathcal{X} = \mathbb{R},\ f_X(x;\boldsymbol{\theta}) = \frac{1}{\pi\theta_2\left[1+\left(\frac{x-\theta_1}{\theta_2}\right)^2\right]} I_{\mathbb{R}}(x),\ \boldsymbol{\theta} = (\theta_1, \theta_2) \in \Theta = \mathbb{R} \times \mathbb{R}^+\right\}.$$

Il valore atteso e la varianza della v.a. di Cauchy non esistono. Il parametro θ_1, coincidente con la moda e la mediana della v.a. di Cauchy, è un parametro di *posizione*. Il secondo parametro è invece un parametro di *scala*, che controlla il grado di dispersione della densità rispetto al parametro θ_1. La v.a. di Cauchy è un caso speciale della distribuzione t di Student, che verrà esaminata più avanti (vedi Paragrafo 3.4.2 del Cap. 3). La Figura (1.5) riporta il grafico della funzione di densità della v.a. in esame per alcuni valori dei parametri θ_1 e θ_2. Si noti che il parametro θ_1 controlla la posizione del punto di massimo della funzione di densità, mentre il parametro θ_2 controlla il grado di dispersione della densità intorno al suo punto di massimo. □

1.19 Esempio (Modello beta). Una v.a. assolutamente continua X ha distribuzione *beta* di parametro $\boldsymbol{\theta} = (\theta_1, \theta_2)$, con $\theta_1, \theta_2 \in \mathbb{R}^+$ – sinteticamente: $X|\boldsymbol{\theta} \sim \text{Beta}(\theta_1, \theta_2)$ – se i tre elementi caratterizzanti del modello sono

- $\mathcal{X} = [0, 1]$;
- $f_X(x;\boldsymbol{\theta}) = \dfrac{1}{B(\theta_1, \theta_2)} x^{\theta_1-1}(1-x)^{\theta_2-1}, \quad x \in \mathcal{X},$
- $\Theta = \left\{(\theta_1, \theta_2) \in \mathbb{R}^+ \times \mathbb{R}^+\right\}.$

dove

$$B(\theta_1, \theta_2) = \frac{\Gamma(\theta_1)\Gamma(\theta_2)}{\Gamma(\theta_1+\theta_2)}.$$

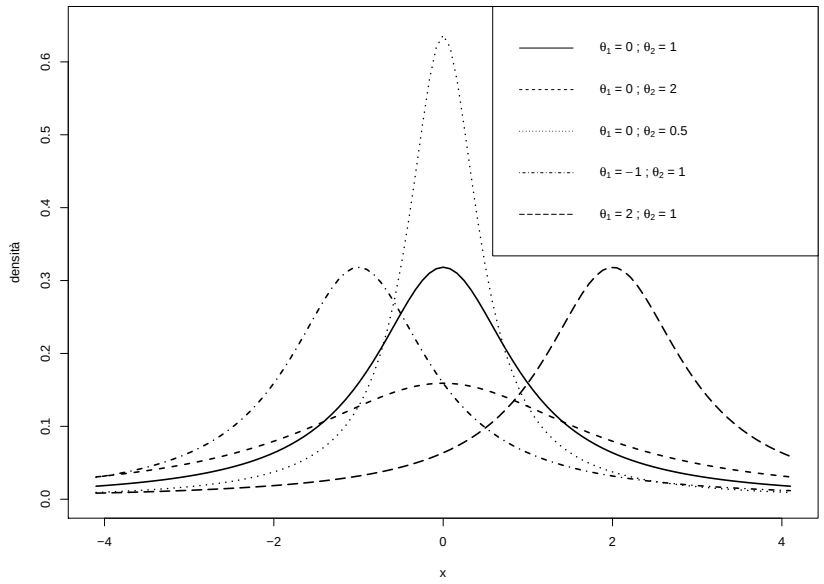

Figura 1.5: *Grafico della funzione di densità di una v.a. Cauchy per alcuni valori dei parametri* θ_1 *e* θ_2*.*

Il modello statistico di base è quindi individuato dalla terna:

$$\left\{\mathcal{X} = [0,1],\ f_X(x;\boldsymbol{\theta}) = \frac{1}{B(\theta_1,\theta_2)} x^{\theta_1-1}(1-x)^{\theta_2-1} I_{[0,1]}(x),\ \boldsymbol{\theta} = (\theta_1,\theta_2) \in \Theta\mathbb{R}^+ \times \mathbb{R}^+\right\}.$$

Il valore atteso e la varianza della v.a. beta sono:

$$\mathbb{E}_\theta[X] = \frac{\theta_1}{\theta_1+\theta_2}, \qquad \mathbb{V}_\theta[X] = \frac{\theta_2\theta_1}{(\theta_1+\theta_2)^2(\theta_1+\theta_2+1)}.$$

La Figura (1.6) riporta il grafico della funzione di densità della v.a. in esame per alcuni valori dei parametri θ_1 e θ_2. □

1.2.3 Modelli con supporto dipendente dal parametro

Fin qui abbiamo considerato degli esempi in cui il supporto della v.a. X non dipende dal parametro incognito. Consideriamo ora un esempio in cui invece l'insieme dei valori in cui $f_X(\cdot;\theta) > 0$ dipende da θ. In questo caso indichiamo la dipendenza del supporto dal parametro con la notazione $\mathcal{X}_\theta$. Vedremo nel prosieguo che la presenza de parametro nel supporto del modello di base determina la necessità di attenzione particolare nella elaborazione di tutte le procedure inferenziali.

1.20 Esempio (Modello uniforme). Per la v.a. assolutamente continua X *uniforme* di parametro $\boldsymbol{\theta} = (\theta_1,\theta_2)$, con $\theta_1,\theta_2 \in \mathbb{R}$ e $\theta_1 < \theta_2$ – sinteticamente: $X|\boldsymbol{\theta} \sim \text{Unif}(\theta_1,\theta_2)$ – i tre

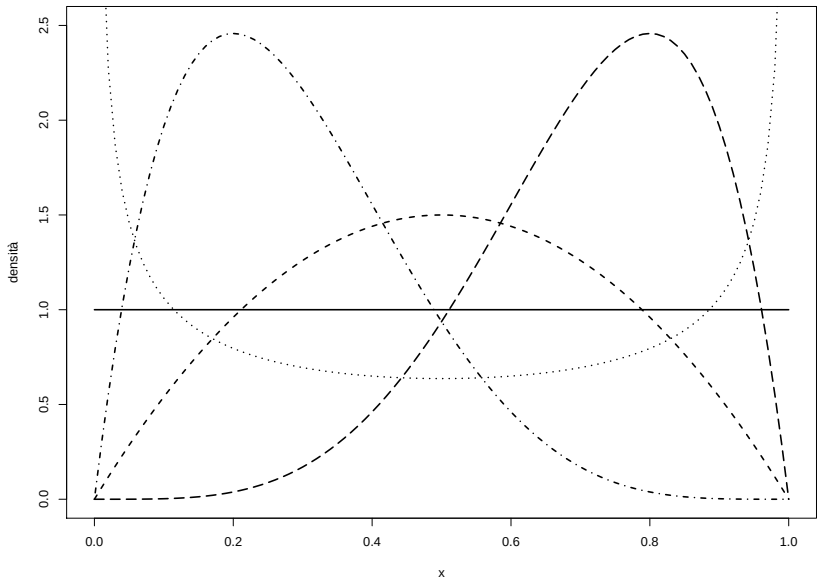

Figura 1.6: *Grafico della funzione di densità di una v.a. beta per alcuni valori dei parametri θ_1 e θ_2: densità uniforme ($\theta_1 = \theta_2 = 1$), con moda in $x = 1/2$ ($\theta_1 = \theta_2 = 2$), con moda in $x = 1/5$ ($\theta_1 = 2, \theta_2 = 5$), con moda in $x = 4/5$ ($\theta_1 = 5, \theta_2 = 2$) e con punto di minimo in $x = 1/2$ ($\theta_1 = 1/2, \theta_2 = 1/2$).*

elementi caratterizzanti del modello sono

- $\mathcal{X}_\theta = [\theta_1, \theta_2]$;
- $f_X(x; \boldsymbol{\theta}) = \dfrac{1}{\theta_2 - \theta_1}, \qquad x \in \mathcal{X}_\theta$;
- $\Theta = \{(\theta_1, \theta_2) \in \mathbb{R} \times \mathbb{R} : \theta_1 < \theta_2\}$.

In questo caso

$$f_X(x; \boldsymbol{\theta}) = \begin{cases} \frac{1}{\theta_2 - \theta_1} & x \in [\theta_1, \theta_2] \\ 0 & x \notin [\theta_1, \theta_2] \end{cases} = \frac{1}{\theta_2 - \theta_1} I_{[\theta_1, \theta_2]}(x)$$

e il supporto della v.a. dipende dal parametro incognito θ. Il modello statistico di base è quindi individuato dalla terna:

$$\left\{ \mathcal{X}_{\boldsymbol{\theta}} = [\theta_1, \theta_2],\ f_X(x; \boldsymbol{\theta}) = \frac{1}{\theta_2 - \theta_1}\ I_{[\theta_1, \theta_2]}(x),\ \boldsymbol{\theta} \in \Theta = \{(\theta_1, \theta_2) \in \mathbb{R} \times \mathbb{R},\ \theta_1 < \theta_2\} \right\}.$$

Il valore atteso e la varianza della v.a. uniforme nell'intervallo $[\theta_1, \theta_2]$ sono:

$$\mathbb{E}_\theta[X] = \frac{\theta_2 + \theta_1}{2}, \qquad \mathbb{V}_\theta[X] = \frac{(\theta_2 - \theta_1)^2}{12}.$$

□

1.2.4 Famiglie posizione-scala

Consideriamo tre diversi metodi per costruire famiglie di distribuzioni per v.a. assolutamente continue, utili dal punto di vista applicativo e dotate di interessanti proprietà. Le tre famiglie prendono il nome di *famiglie di posizione*, *famiglie di scala* e *famiglie di posizione-scala*. Le prime due famiglie sono casi particolari della terza. Con piccolo abuso rispetto alla convenzione usata nel resto del testo, indicheremo le funzioni di densità dei membri delle tre famiglie con i simboli g_p, g_s e g_{ps}.
L'elemento da cui partire per la costruzione di queste famiglie parametriche è una funzione di densità $f(\cdot)$ (in genere, ma non necessariamente, non dipendente da parametri incogniti), denominata *funzione di densità standard*. Un esempio di densità standard è $f(x) = (2\pi)^{-1/2}\exp\{-x^2/2\},\ x \in \mathbb{R}$, la densità della v.a. $\mathrm{N}(0,1)$. Partendo da una funzione di densità standard, con opportune trasformazioni è possibile definire tre distinti modelli statistici di base, utilizzando il seguente risultato.

1.21 Teorema. Sia $f(x)$ una funzione di densità di probabilità con supporto $\mathbb{R}$. Date due costanti $\mu \in \mathbb{R}$ e $\sigma \in \mathbb{R}^+$, la funzione

$$g(x;\mu,\sigma) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$$

è una funzione di densità di probabilità.

Dimostrazione. Per verificare che g è una funzione di densità è sufficiente mostrare che

$$g(x;\mu,\sigma) \geq 0 \qquad \forall x \in \mathbb{R}$$

e che

$$\int_{-\infty}^{+\infty} g(x;\mu,\sigma)dx = 1.$$

La positività di g discende da quella di f per ogni valore reale x e da quella di σ. Inoltre, considerando la sostituzione $y = (x-\mu)/\sigma$ (da cui $dx = \sigma dy$) si ha che

$$\int_{-\infty}^{+\infty} g(x;\mu,\sigma)dx = \int_{-\infty}^{+\infty} f(y)dy = 1.$$

□

A questo punto possiamo definire le famiglie di posizione, di scala e di posizione-scala.

1.22 Definizione (Famiglia di posizione). Sia $f(x)$ una qualsiasi funzione di densità con supporto $\mathbb{R}$. La famiglia di funzioni di densità $(g_p(\cdot;\mu),\ \mu \in \mathbb{R})$, definita ponendo

$$g_p(x;\mu) = f(x-\mu), \qquad \mu \in \mathbb{R}$$

ottenuta al variare di $\mu \in \mathbb{R}$ è denominata *famiglia di posizione con funzione di densità standard* f. Il parametro μ è denominato *parametro di posizione* della famiglia. □

Il parametro di posizione ha l'effetto di traslare la funzione di densità standard, mentre la forma e il grado di dispersione della stessa funzione restano inalterati. Molte delle funzioni di densità di uso comune sono famiglie di posizione.

1.23 Esempio (Famiglia di posizione normale). Consideriamo la funzione di densità standard

$$f(x) = \frac{1}{\sqrt{2\pi}}\exp\{-\frac{x^2}{2}\}I_{\mathbb{R}}(x).$$

La corrispondente famiglia di posizione è l'insieme delle densità delle v.a. $\mathrm{N}(\mu, 1)$, $\mu \in \mathbb{R}$:

$$g_p(x;\mu) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x-\mu)^2\right\} I_{\mathbb{R}}(x-\mu) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x-\mu)^2\right\} I_{\mathbb{R}}(x), \quad \mu \in \mathbb{R}$$

dove la seconda uguaglianza si ottiene osservando che $I_{\mathbb{R}}(x-\mu) = I_{\mathbb{R}}(x)$. In questo caso il parametro di posizione coincide con il valore atteso μ della v.a. La Fig. 1.2 mostra il grafico della densità normale per tre valori del parametro di posizione (qui indicato con θ). □

1.24 Esempio (Famiglia di posizione di Cauchy). Se consideriamo la funzione di densità standard di Cauchy

$$f(x) = \frac{1}{\pi} \frac{1}{1+x^2} I_{\mathbb{R}}(x),$$

osservando che, come nel precedente esempio, $I_{\mathbb{R}}(x-\mu) = I_{\mathbb{R}}(x)$, la corrispondente famiglia di posizione di parametro μ risulta essere

$$g_p(x;\mu) = \frac{1}{\pi} \frac{1}{1+(x-\mu)^2} I_{\mathbb{R}}(x), \quad \mu \in \mathbb{R},$$

ovvero la famiglia delle densità di Cauchy con parametri $(\mu, 1)$. In questo caso il parametro di posizione μ non coincide con il valore atteso della v.a. di Cauchy (per tale v.a. non esistono nè valore atteso nè varianza). □

Supporto limitato. Nel teorema e nella definizione di famiglia di posizione abbiamo assunto che il supporto della densità standard sia $\mathbb{R}$, che è anche il supporto di g_p, dal momento che $I_{\mathbb{R}}(x-\mu) = I_{\mathbb{R}}(x)$. Le cose cambiano quando il supporto delle variabili considerate è un sottoinsieme proprio di $\mathbb{R}$. Nei casi più comuni il supporto è un intervallo reale $[a, b]$ oppure una semiretta. In queste situazioni anche la funzione indicatrice subisce delle modifiche. In generale si ha che

$$I_{[a,b]}(x-\mu) = I_{[a+\mu, b+\mu]}(x).$$

I casi per $b \to -\infty$ e $a \to \infty$ si ottengono di conseguenza.

1.25 Esempio (Famiglia di posizione esponenziale). Consideriamo la densità standard $f(x) = e^{-x} I_{\mathbb{R}^+}(x)$. Poichè $I_{[0,\infty)}(x-\mu) = I_{[\mu,\infty)}(x)$ abbiamo che la famiglia di posizione esponenziale è definita dalle densità

$$g_p(x;\mu) = e^{-(x-\mu)} I_{[0,\infty)}(x-\mu) = e^{-(x-\mu)} I_{[\mu,\infty)}(x), \qquad \mu \in \mathbb{R}.$$

□

1.26 Definizione (Famiglia di scala). Sia $f(x)$ una qualsiasi funzione di densità con supporto $\mathbb{R}$. La famiglia di funzioni di densità $(g_s(\cdot;\sigma),\ \sigma \in \mathbb{R}^+)$, definita ponendo

$$g_s(x;\sigma) = \frac{1}{\sigma} f\left(\frac{x}{\sigma}\right), \qquad \sigma \in \mathbb{R}^+$$

ottenuta al variare di $\sigma \in \mathbb{R}^+$ è denominata *famiglia di scala con funzione di densità standard* f. Il parametro σ è denominato *parametro di scala* della famiglia. □

Il parametro di scala regola, in generale, il grado di dispersione della funzione f intorno a un valore di riferimento. Se si considera, ad esempio, una densità standard simmetrica rispetto a un valore x_0, punto di massimo unico di f interno a $\mathcal{X}$, possiamo verificare che,

introducendo il parametro di scala, la densità g_s resta simmetrica rispetto a x_0 ma valori di $\sigma > 1$ determinano un "appiattimento" della densità, mentre valori di $\sigma < 1$ determinano un "addensamento" di g_s intorno a x_0. La Fig. 1.3 mostra il grafico della densità normale per tre valori del parametro di scala (qui indicato con θ).

1.27 Esempio (Famiglia di scala di Cauchy). Se consideriamo la funzione di densità standard di Cauchy (vedi Es. 1.18), osservando che, $\forall \sigma > 0$, $I_{\mathbb{R}}(x/\sigma) = I_{\mathbb{R}}(x)$, abbiamo che la famiglia di scala è definita dalle densità

$$g_s(x;\sigma) = \frac{1}{\pi\sigma} \frac{1}{\left[1 + \left(\frac{x}{\sigma}\right)^2\right]} I_{\mathbb{R}}(x), \qquad \sigma \in \mathbb{R}^+.$$

□

Supporto limitato. Anche in questo caso, se il supporto della densità standard è limitato, è necessario fare attenzione alla funzione indicatrice del supporto della densità standard nel determinare l'espressione di g_s. In generale, se il supporto della densità standard è l'intervallo $[a, b]$, si ha che

$$I_{[a,b]}\left(\frac{x}{\sigma}\right) = I_{[\sigma a, \sigma b]}(x).$$

I casi per $b \to -\infty$ e $a \to \infty$ si ottengono di conseguenza.

1.28 Esempio (Famiglia di scala esponenziale). Consideriamo la densità standard esponenziale $f(x) = e^{-x} I_{[0,\infty)}(x)$. La famiglia di scala esponenziale risulta essere definita dalle densità

$$g_s(x;\sigma) = \frac{1}{\sigma}\, e^{-\frac{x}{\sigma}} I_{[0,\infty)}\left(\frac{x}{\sigma}\right) = \frac{1}{\sigma}\, e^{-\frac{x}{\sigma}} I_{[0,\infty)}(x), \quad \sigma \in \mathbb{R}^+.$$

Si ottiene così l'usuale funzione di densità della v.a. esponenziale di parametro (di scala) σ. La Fig. 1.1 mostra il grafico della densità della v.a. esponenziale per tre valori del parametro di scala (qui indicato con θ). □

1.29 Definizione (Famiglia di posizione-scala). Sia $f(x)$ una qualsiasi funzione di densità (funzione di densità standard) con supporto $\mathbb{R}$. Date due costanti $\mu \in \mathbb{R}$ e $\sigma \in \mathbb{R}^+$, la famiglia di funzioni di densità

$$g_{ps}(x;\mu,\sigma) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right), \quad \mu \in \mathbb{R}, \quad \sigma \in \mathbb{R}^+,$$

è denominata *famiglia di posizione-scala con funzione di densità standard* f. I parametri μ e σ sono denominati rispettivamente *parametro di posizione* e *parametro di scala* della famiglia. □

In questo caso, la funzione di densità g_{ps} risulta traslata rispetto a f e con un diverso grado di dispersione, a seconda del valore di σ. Le funzioni di densità delle v.a. $\mathrm{N}(\mu, \sigma^2)$ e $\mathrm{Cau}(\mu, \sigma)$ sono esempi di famiglie di posizione-scala.

Supporto limitato. Se il supporto della densità standard è l'intervallo $[a, b]$, nel determinare l'espressione di f_{ps} si tenga conto del fatto che

$$I_{[a,b]}\left(\frac{x-\mu}{\sigma}\right) = I_{[\mu+\sigma a, \mu+\sigma b]}(x).$$

I casi per $b \to -\infty$ e $a \to \infty$ si ottengono di conseguenza.

1.30 Esempio (Famiglia posizione-scala uniforme). Consideriamo come densità standard quella della v.a. uniforme in $[-1, 1]$, ovvero $f(x) = \frac{1}{2} I_{[-1,1]}(x)$. In questo caso con semplici passaggi si verifica che

$$g_{ps}(x) = \frac{1}{2\sigma} I_{[\mu-\sigma,\mu+\sigma]}(x), \quad \mu \in \mathbb{R},\ \sigma \in \mathbb{R}^+.$$

□

1.31 Osservazione.

i) Le famiglie di posizione e le famiglie di scala si possono ottenere da quella di posizione-scala ponendo, rispettivamente, $\sigma = 1$ e $\mu = 0$.

ii) Si può verificare che, se f è una funzione di densità simmetrica rispetto a zero, ovvero se $x = 0$ è mediana di $f(x)$, allora μ è la mediana delle densità di $g_p(x; \mu)$ e $g_{ps}(x; \mu)$.

iii) Dalle definizioni date, si evince che è possibile definire una famiglia di posizione-scala partendo da densità standard a loro volta già membri di una famiglia di posizione-scala. Se ad esempio la densità standard è quella di una v.a. $\mathrm{N}(a, b)$, per due specifici valori $a \in \mathbb{R}$ e $b \in \mathbb{R}^+$, la corrispondente famiglia ha densità

$$g_{ps}(x; \mu, \sigma)(x) = \frac{1}{\sigma\sqrt{2\pi b}} \exp -\frac{1}{2} \frac{(x - a - \mu)^2}{b\sigma^2}, \qquad \mu \in \mathbb{R},\ \sigma \in \mathbb{R}^+.$$

Tuttavia, l'arbitrarietà dei valori di μ e σ fa sì che le famiglie di posizione-scala ottenute considerando come densità standard quelle delle v.a. $\mathrm{N}(0, 1)$ e $\mathrm{N}(a, b)$ siano le stesse.

□

I seguenti due risultati mettono in relazione le v.a. le cui funzioni di densità sono membri di una famiglia posizione-scala (che indichiamo con Y) con la v.a. (indicata con Z) associata alla funzione di densità standard della stessa famiglia. Per la dimostrazione si rimanda a Casella-Berger (2002, pp. 120-121).

1.32 Teorema. Sia $f(\cdot)$ una qualsiasi funzione di densità, μ un qualsiasi numero reale e σ un qualsiasi valore reale positivo. La v.a. Y ha funzione di densità $1/\sigma f((y - \mu)/\sigma)$ se e solo se esiste una v.a. Z con funzione di densità $f(\cdot)$ e tale che $Y = \sigma Z + \mu$. □

Il teorema precedente dice sostanzialmente che la v.a. Y, la cui densità è un membro della famiglia di posizione-scala con parametri (μ, σ), si ottiene dalla v.a. Z la cui densità coincide con la densità standard, moltiplicandola per σ (trasformazione di scala di Z) e aggiungendo μ (traslazione di Z).
Il teorema che segue mette in relazione il valore atteso e la varianza (supposto che esistano) delle v.a. Y aventi densità che appartengono a una famiglia di locazione-scala con il valore atteso e la varianza della v.a. corrispondente alla densità standard con la quale la famiglia viene costruita.

1.33 Teorema. Sia Z una v.a. con funzione di densità f. Supponiamo che esistano $\mathbb{E}(Z)$ e $\mathbb{V}(Z)$. Se Y è una v.a. con funzione di densità $1/\sigma f((y - \mu)/\sigma)$, allora

$$\mathbb{E}(Y) = \sigma\mathbb{E}(Z) + \mu, \qquad \mathbb{V}(Y) = \sigma^2\mathbb{V}(Z).$$

Se $\mathbb{E}(Z) = 0$ e $\mathbb{V}(Z) = 1$, allora $\mathbb{E}(Y) = \mu$ e $\mathbb{V}(Y) = \sigma^2$. □

Come osservato in precedenza, nulla vieta che la v.a. Z abbia valore atteso diverso da zero e varianza diversa da uno. Tuttavia, senza perdere in generalità, si può scegliere come densità standard quella con $\mathbb{E}(Z) = 0$ e $\mathbb{V}(Z) = 1$. In questo caso i parametri di posizione e scala coincidono con il valore atteso e la radice quadrata della varianza della v.a. Y.

1.3 Il modello statistico

1.3.1 Esperimento e modello statistico

Si è detto (vedi Prg. 1.1.3) che l'idea alla base dell'inferenza statistica è che i dati disponibili per un un fenomeno di interesse siano stati generati da un "meccanismo" aleatorio, governato cioè da una legge di probabilità. Ciascuna osservazione x viene considerata, in questo contesto, come la *realizzazione* di una v.a. X (si noti che, come usuale, utilizziamo le lettere maiuscole per indicare le v.a. e le corrispondenti lettere minuscole per indicarne le realizzazioni). Abbiamo anche detto che l'obiettivo dell'inferenza statistica è l'individuazione del meccanismo aleatorio (ovvero la legge di probabilità) che governa la v.a. X, partendo dai dati a disposizione. Nei problemi di inferenza *parametrica* ipotizziamo che la legge di probabilità di X appartenga alla famiglia $\mathcal{F} = (f(\cdot;\theta),\ \theta \in \Theta)$ e che l'incertezza consista nel non sapere quale, tra tutte le leggi in $\mathcal{F}$, sia quella che regola il manifestarsi di X. In pratica, desideriamo individuare il valore di θ, che indichiamo con θ^*, al quale corrisponde il membro $f(\cdot;\theta^*)$ di $\mathcal{F}$ che presiede al meccanismo aleatorio che ha generato i dati.
L'individuazione di θ^* avviene quindi attraverso l'uso di *dati campionari*. L'idea è quella di effettuare n prove (n lanci della moneta, n lanci del dado, la somministrazione del farmaco a n soggetti, la rilevazione di un carattere su n individui estratti da una popolazione...) e di utilizzare i risultati osservati per l'operazione di inferenza su θ. Prima di effettuare l'esperimento, i risultati delle n prove sono assimilabili a n variabili aleatorie $X_1, X_2, \ldots, X_n$, in cui la generica X_i rappresenta la v.a. "risultato della prova i-esima". Il vettore aleatorio

$$\mathbf{X}_n = (X_1, X_2, \ldots, X_i, \ldots, X_n)$$

costituisce il *campione*, di *numerosità* (o *ampiezza*) pari a n. Le v.a. X_i sono denominate *componenti* del campione $\mathbf{X}_n$. L'insieme delle possibili realizzazioni del campione $\mathbf{X}_n$, denominato *spazio campionario* e indicato con $\mathcal{X}^n$, è costituito da tutte le possibili n-ple

$$\mathbf{x}_n = (x_1, x_2, \ldots, x_n)$$

che possiamo osservare. Una realizzazione $\mathbf{x}_n$ di $\mathbf{X}_n$ verrà chiamata *campione osservato* (o *realizzato*). In pratica quindi, ciascun valore x_i rappresenta il valore osservato della v.a. X_i, relativa alla prova i-esima. Poichè la legge di probabilità di ciascuna v.a. X_i, $f_{X_i}(\cdot;\theta)$, dipende dal parametro incognito θ, anche la legge di probabilità della v.a. multipla $X_1, \ldots, X_n$, che indichiamo con

$$f_{X_1 \cdots X_n}(\cdot, \ldots, \cdot; \theta)$$

dipende dallo stesso parametro incognito θ. Per semplicità, nel seguito indicheremo la legge $f_{X_1 \cdots X_n}$ anche con le notazioni $f_{\mathbf{X}_n}$ e f_n e con

$$f_n(\mathbf{x}_n;\theta) = f_{\mathbf{X}_n}(\mathbf{x}_n;\theta) = f_n(x_1, \ldots, x_n;\theta) = f_{X_1 \cdots X_n}(x_1, \ldots, x_n;\theta)$$

il valore della funzione di densità congiunta della v.a. $X_1, \ldots, X_n$ in corrispondenza del campione $\mathbf{x}_n$. Al solito, f_n indica la funzione di densità di $\mathbf{X}_n$, quando le v.a. X_i di $\mathbf{X}_n$ sono assolutamente continue; oppure la funzione di massa di probabilità, nel caso di v.a. discrete[6].
Così come abbiamo fatto per la v.a. di base X, possiamo ora definire, con riferimento al campione $\mathbf{X}_n$ e alla sua legge di probabilità, il *modello statistico*. Questo è costituito dalla terna:

$$\{\mathcal{X}^n,\ f_n(\cdot;\theta),\ \theta \in \Theta\}\,.$$

[6]Nel caso di v.a. assolutamente continue $f_n : \mathcal{X}^n \to \mathbb{R}^+$, mentre per le v.a. discrete $f_n : \mathcal{X}^n \to [0,1]$.

Useremo anche la notazione

$$\{\mathcal{X}^n,\ \mathcal{F}_n\},$$

dove

$$\mathcal{F}_n = (f_n(\cdot;\theta),\ \theta \in \Theta)$$

rappresenta la famiglia parametrica di leggi di probabilità della v.a. multipla $X_1, \ldots, X_n$.

1.3.2 Campioni casuali

Una situazione sperimentale particolare, ma molto rilevante, si ha quando si assume che le n prove vengano effettuate tutte nelle stesse condizioni e indipendentemente l'una dall'altra. In questo modo possiamo considerare le variabili aleatorie $X_1, \ldots, X_n$ indipendenti tra loro e con stessa distribuzione di probabilità. Si assume cioè che gli esiti delle varie prove non influenzino l'esito delle altre e che, in pratica, le n prove possano essere considerate delle repliche di una stessa prova, il cui risultato è una realizzazione della stessa v.a. di base, X, che ha legge di probabilità f_X. Queste due assunzioni vengono formalizzate con le ipotesi che le v.a. $X_1, \ldots, X_n$ siano *indipendenti* e *somiglianti* (o *identicamente distribuite*, i.i.d.). In questo caso il campione $(X_1, \ldots, X_n)$ prende il nome di *campione casuale*, mentre $(x_1, \ldots, x_n)$ indica il *campione casuale osservato*. L'assunzione di somiglianza implica che le leggi di probabilità $f_{X_i}(\cdot;\theta)$ delle v.a. componenti il campione casuale siano tutte uguali tra loro, e uguali alla legge $f_X(\cdot;\theta)$ della v.a. di base, X:

$$f_{X_1}(\cdot;\theta) = f_{X_2}(\cdot;\theta) = \ldots = f_{X_n}(\cdot;\theta) = f_X(\cdot;\theta).$$

Si noti che il parametro delle leggi di probabilità delle n v.a. è, in virtù dell'ipotesi di uguale distribuzione, sempre lo stesso. Per l'assunzione di indipendenza delle v.a. $X_1, \ldots, X_n$ possiamo scrivere che la probabilità (nel caso discreto) o la densità di probabilità (caso continuo) di osservare un generico campione $(x_1, \ldots, x_n)$, elemento di $\mathcal{X}^n$, risulta uguale al prodotto delle singole funzioni di massa di probabilità o di densità di probabilità delle n v.a. aleatorie, nei punti $x_1, \ldots, x_n$:

$$f_n(\mathbf{x}_n;\theta) = f_{X_1 \ldots X_n}(x_1, \ldots, x_n;\theta) = \prod_{i=1}^{n} f_{X_i}(x_i;\theta) = \prod_{i=1}^{n} f_X(x_i;\theta). \tag{1.1}$$

La seconda uguaglianza è la conseguenza dell'ipotesi di indipendenza delle n v.a.; la terza della loro identica distribuzione. Si noti che, nell'ultima produttoria, l'indice i agisce unicamente sui valori x_i in cui la legge $f_X(\cdot;\theta)$ viene valutata. Dall'ipotesi di indipendenza e identica distribuzione discende quindi una notevole semplificazione del problema. Infatti, la legge di probabilità della v.a. *multipla* n-dimensionale $\mathbf{X}_n$, (definita in $\mathbb{R}^n$ o in $\overline{\mathbb{N}}^n$ e che assume valori in $\mathbb{R}^+$) può essere scritta come prodotto della legge di probabilità della v.a. *semplice* di base, X (definita in $\mathbb{R}$ o in $\overline{\mathbb{N}}$ e a valori in $\mathbb{R}^+$), valutata negli n punti $x_1, \ldots, x_n$ di un campione.

1.3.3 Campione casuale, popolazioni virtuali e finite

Lo schema sperimentale che è alla base della definizione di campione casuale è talvolta denominato *campionamento da popolazioni infinite*[7] . Per chiarire, supponiamo di considerare una popolazione costituita da un numero "praticamente" infinito di unità e di poter ottenere i dati campionari $X_1, \ldots, X_n$ estraendo le unità in modo sequenziale, rilevando il carattere X sulle unità estratte. Assumiamo inoltre che nella prima prova si ottenga per X_1 la realizzazione x_1. Il fatto di considerare una popolazione infinita fa sì che tale risultato non

[7] Vedi, ad esempio, Casella-Berger (2002), Cap. 5, pag. 209.

influenzi l'esito della seconda prova: la distribuzione di probabilità di X_2 non è modificata dal fatto di sapere che nella prima prova si è osservato il valore x_1. Ciò avviene indipendentemente dall'avere o meno rimosso l'unità estratta alla prima prova dell'esperimento[8]. In questo caso, quindi, lo schema concettuale del campione casuale (indipendenza e identica distribuzione delle v.a. componenti) è compatibile con il campionamento (con o senza rimpiazzo) da popolazioni costituite da un numero infinito di unità statistiche.
Nel caso di campionamento da popolazioni finite, invece, lo schema del campione casuale come definito nella (1.1) è compatibile con lo schema di estrazione con ripetizione ma non con quello senza ripetizione. Siano $\{a_1, \ldots, a_N\}$ i valori - che qui per semplicità espositiva assumiamo distinti - della variabile X associati alle $N < +\infty$ unità di una popolazione e siano $X_1, \ldots, X_n$ i valori (aleatori) di X associati alle n estrazioni delle unità della popolazione. Nel caso di estrazione con ripetizione, si estraggono sequenzialmente n unità dalla popolazione e si riportano i valori $x_1, \ldots, x_n$ osservati di X, sotto l'ipotesi che, ad ogni estrazione, ciascuna unità della popolazione possa essere nuovamente estratta. In questo caso, le v.a. $X_1, \ldots, X_n$ sono tra loro indipendenti e con stessa distribuzione. Infatti, ad esempio, $\mathbb{P}(X_2 = a_i | X_1 = a_j) = \mathbb{P}(X_2 = a_i) = 1/N$, ovvero, il fatto che alla prima prova sia stata estratta la j-esima unità della popolazione, ovvero il valore a_j, non altera la distribuzione della v.a. X_2 (indipendenza). Inoltre, per ciascuna X_i e ciascuna a_h, $\mathbb{P}(X_i = a_h)$ è costante, e pari a $1/N$: vi è cioè identica distribuzione delle n v.a. $X_1, \ldots, X_n$.
Al contrario, se si considera lo schema di estrazione senza ripetizione, una unità estratta ad una prova non può più essere osservata nelle estrazioni successive. Ad esempio, se consideriamo due elementi distinti della popolazione a cui corrispondono i valori a_i e a_j della variabile, si ha che $\mathbb{P}(X_2 = a_j | X_1 = a_j) = 0$, mentre $\mathbb{P}(X_2 = a_j | X_1 = a_k) = 1/(N-1)$. Pertanto, quanto avviene alla prima estrazione (ovvero l'aver estratto a_j o meno) influenza la probabilità dei valori che può assumere la variabile X_2, associata alla seconda estrazione. Ciò vuol dire che X_1 e X_2 non sono tra loro indipendenti e quindi che, in generale, $\mathbf{X}_n = (X_1, \ldots, X_n)$ non è un campione casuale.
Si noti che, nel caso di estrazione senza ripetizione, le v.a. $X_i,\ i = 1, \ldots, n$, pur non essendo indipendenti tra loro, sono identicamente distribuite. Infatti, per ogni possibile valore $x \in \{a_1, \ldots, a_N\}$ si ha che:

$$\begin{aligned} \mathbb{P}(X_2 = x) &= \sum_{j=1}^{N} \mathbb{P}(X_2 = x | X_1 = a_j) \times \mathbb{P}(X_1 = a_j) \\ &= (N-1)\frac{1}{N-1} \times \frac{1}{N} = \frac{1}{N}. \end{aligned}$$

Infatti, per il valore dell'indice j per il quale $a_j = x$, si ha che $\mathbb{P}(X_2 = x | X_1 = a_j) = 0$; per gli altri $N-1$ valori di a_j si ha invece che $\mathbb{P}(X_2 = x | X_1 = a_j) = 1/(N-1)$ e $\mathbb{P}(X_1 = a_j) = 1/N$. In modo simile si può anche provare che la distribuzione marginale trovata è la stessa per tutte le v.a. $X_i,\ i = 1, \ldots, n$.

In sintesi, nello schema rappresentato dal campione casuale, possiamo in pratica pensare ad un generico insieme di dati osservati in tre modi diversi:

- come al risultato di un numero limitato di prove che, se ripetute all'infinito nelle stesse condizioni e sotto l'ipotesi di indipendenza, darebbero luogo a una popolazione in cui il fenomeno di interesse ha distribuzione rappresentata dalla legge $f_X(x; \theta)$;
- come ai valori di un carattere X rilevati su un numero finito di unità estratte (con o senza ripetizione) da una popolazione con un numero infinito di unità, in cui il carattere ha distribuzione rappresentata dalla legge $f_X(x; \theta)$;

[8] In teoria dei campioni si parla di campionamento con o senza ripetizione (o rimpiazzo) a seconda che le unità statistiche estratte possano o meno essere incluse nel campione più di una volta.

- come ai valori di un carattere X rilevati sulle unità estratte con ripetizione da una popolazione finita di unità.

Nel primo caso (si pensi all'esempio relativo alle misurazioni ripetute di una grandezza incognita) la popolazione è quindi virtuale. Nel secondo esempio (si pensi al caso di un carattere rilevabile sulla "popolazione" - virtualmente infinita - dei globuli rossi di un individuo) la popolazione è reale ma di numerosità trattata come se fosse infinita. Nel terzo caso (si pensi ad esempio ad una popolazione studentesca) la popolazione è reale e finita, ma lo schema di estrazione determina l'indipendenza e l'identica distribuzione delle variabili $X_1, \ldots, X_n$.

1.3.4 Modelli statistici per campioni casuali

Con riferimento al campione casuale $\mathbf{X}_n = (X_1, \ldots, X_n)$ e alla sua legge di probabilità, il *modello statistico* è rappresentato dalla terna

$$\left\{ \mathcal{X}^n,\ f_n(\mathbf{x}_n\ ;\theta) = \prod_{i=1}^{n}\ f_X(x_i;\theta),\ \theta \in \Theta \right\}. \tag{1.2}$$

Il modello statistico (1.2), relativo a campioni casuali, fornisce non solo le informazioni essenziali relative alla distribuzione di probabilità delle v.a. $X_1, \ldots, X_n$, ma ci informa anche dello schema sperimentale che dà luogo ai dati osservati, ovvero del fatto che le v.a. sono indipendenti e identicamente distribuite. Ciò avviene in generale: il modello statistico racchiude sia le informazioni sulla distribuzione di probabilità delle v.a. coinvolte sia le informazioni sulla procedura sperimentale che produce le osservazioni.

Possiamo ora descrivere i modelli statistici per campioni casuali relativi alle principali variabili aleatorie introdotte in precedenza. Per semplicità di notazione, quando consideremero modelli per v.a. con supporto che non dipende dal parametro, nelle espressioni delle funzioni $f_n(\mathbf{x}_n;\theta)$ verrà omessa la funzione indicatrice dell'insieme $\mathcal{X}^n$.

1.34 Esempio (Modello bernoulliano). Supponiamo che $\mathbf{X}_n = (X_1, \ldots, X_n)$ sia un campione casuale bernoulliano di parametro incognito θ. Questo significa che le v.a. $X_1, \ldots, X_n$ sono tra loro indipendenti e che ciascuna v.a. X_i ha la stessa distribuzione della v.a. di base bernoulliana X con parametro incognito θ. Sinteticamente:

$$X_1, \ldots, X_n|\theta \quad \text{i.i.d.}, \qquad\qquad X_i|\theta \sim \text{Ber}(\theta),\ i = 1, \ldots, n.$$

Vogliamo ora individuare gli elementi che costituiscono il modello statistico bernoulliano. Dato che per ciascuna delle v.a. di base lo spazio $\mathcal{X}$ dei possibili risultati è l'insieme $\{0,1\}$, ogni campione bernoulliano di n elementi potenzialmente osservabile è costituito da n numeri che possono essere, ciascuno, uno 0 o un 1. Lo spazio $\mathcal{X}^n$ è quindi costituito da tutte le n-ple distinte di 0 e di 1.

$$\mathcal{X}^n = \underbrace{\{0,1\} \times \{0,1\} \times \ldots \times \{0,1\}}_{n\ volte} = \{0,1\}^n.$$

Ricordando che, per la v.a. bernoulliana di parametro θ,

$$f_X(x;\theta) = \mathbb{P}\left(X = x;\theta\right) = \theta^x(1-\theta)^{1-x}, \qquad x = 0, 1,$$

dalla (1.1) si ottiene:

$$\begin{aligned} f_n(x_1, \ldots, x_n;\theta) = f_n(\mathbf{x}_n;\theta) &= \prod_{i=1}^{n} f_X(x_i;\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} \\ &= \theta^{\sum_{i=1}^n x_i}\ (1-\theta)^{n-\sum_{i=1}^n x_i} = \theta^{n\bar{x}_n}\ (1-\theta)^{n-n\bar{x}_n}, \end{aligned}$$

dove l'ultima uguaglianza discende dal fatto che $\overline{x}_n = \sum_{i=1}^n x_i/n$.
Il modello statistico per un campione casuale bernoulliano è quindi individuato da:

$$\left\{\mathcal{X}^n = \{0,1\}^n, \quad f_n(\mathbf{x}_n;\theta) = \theta^{n\overline{x}_n}\ (1-\theta)^{n-n\overline{x}_n}, \quad \theta \in \Theta = [0,1]\right\}.$$

Supponiamo, ad esempio, di voler calcolare la probabilità di ottenere, con $n = 4$ lanci indipendenti di una stessa moneta, 3 teste (T) e 1 croce (C), nell'ordine T,T,C,T. Indichiamo l'esito T con 1 e l'esito C con 0. Siamo quindi interessati a calcolare la probabilità del campione $\mathbf{x}_n = (1,1,0,1)$. Per le ipotesi di indipendenza e di identica distribuzione, poiché $\sum_{i=1}^4 x_i = 3$ e $n = 4$, dalla (1.1) si ha che:

$$\begin{aligned} f_4(1,1,0,1;\theta) &= f_X(1;\theta) \times f_X(1;\theta) \times f_X(0;\theta) \times f_X(1;\theta) = \\ &= \theta^{\sum_{i=1}^4 x_i}(1-\theta)^{4-\sum_{i=1}^4 x_i} = \theta^3(1-\theta). \end{aligned}$$

La Tabella che segue riporta, nelle prime 4 colonne, i valori delle osservazioni possibili relative a tutti i campioni dello spazio campionario, $\mathcal{X}^4 = \{0,1\}^4$ e nella quinta colonna la probabilità dei singoli campioni. Si noti che la probabilità di ogni campione osservabile $\mathbf{x}_n$ dipende dal parametro incognito θ. La distribuzione di probabilità riportata costituisce la *distribuzione campionaria* per un esperimento bernoulliano di dimensione $n = 4$. Come vedremo in seguito, da questo legame tra dati e parametro deriva la possibilità di utilizzare le osservazioni campionarie per ottenere informazioni su θ.

$\mathcal{X}^4$				
x_1	x_2	x_3	x_4	$f_4(\mathbf{x}_n;\theta)$
0	0	0	0	$(1-\theta)^4$
0	0	0	1	$\theta(1-\theta)^3$
0	0	1	0	$\theta(1-\theta)^3$
0	1	0	0	$\theta(1-\theta)^3$
1	0	0	0	$\theta(1-\theta)^3$
0	0	1	1	$\theta^2(1-\theta)^2$
0	1	0	1	$\theta^2(1-\theta)^2$
0	1	1	0	$\theta^2(1-\theta)^2$
1	0	0	1	$\theta^2(1-\theta)^2$
1	0	1	0	$\theta^2(1-\theta)^2$
1	1	0	0	$\theta^2(1-\theta)^2$
0	1	1	1	$\theta^3(1-\theta)$
1	0	1	1	$\theta^3(1-\theta)$
1	1	0	1	$\theta^3(1-\theta)$
1	1	1	0	$\theta^3(1-\theta)$
1	1	1	1	θ^4

□

1.35 Esempio (Modello binomiale). Supponiamo che $X_1,\ldots,X_n$ sia un campione casuale binomiale di parametro incognito θ e relativo a k prove:

$$X_1,\ldots,X_n|\theta \quad \text{i.i.d.}, \qquad\qquad X_i|\theta \sim \text{Bin}(k,\theta).$$

Dato che per la v.a. di base, lo spazio $\mathcal{X}$ dei possibili risultati è l'insieme $\{0,1,\ldots,k\}$, ogni campione binomiale di n elementi potenzialmente osservabile è costituito da n numeri, ciascuno dei quali può essere un numero da 0 a k. Lo spazio $\mathcal{X}^n$ è quindi

$$\mathcal{X}^n = \underbrace{\{0,1,\ldots,k\} \times \{0,1,\ldots,k\} \ldots \times \{0,1,\ldots,k\}}_{n \text{ volte}} = \{0,1,\ldots,k\}^n.$$

Ricordando che, per la v.a. binomiale relativa a k prove di parametro θ,

$$f_X(x;\theta) = \mathbb{P}(X = x;\theta) = \binom{k}{x}\theta^x(1-\theta)^{k-x}, \qquad x = 0, 1, \ldots, k,$$

dalla (1.1) si ottiene:

$$\begin{aligned} f_n(x_1, \ldots, x_n;\theta) = f_n(\mathbf{x}_n;\theta) &= \prod_{i=1}^{n} f_X(x_i;\theta) = \prod_{i=1}^{n}\binom{k}{x_i}\theta^{x_i}(1-\theta)^{k-x_i} \\ &= \left[\prod_{i=1}^{n}\binom{k}{x_i}\right]\theta^{\sum_{i=1}^n x_i}\,(1-\theta)^{\sum_{i=1}^n (k-x_i)} = \\ &= \left[\prod_{i=1}^{n}\binom{k}{x_i}\right]\theta^{n\overline{x}_n}\,(1-\theta)^{nk-n\overline{x}_n}, \end{aligned}$$

dove l'ultima uguaglianza discende dal fatto che $\overline{x}_n = \sum_{i=1}^n x_i/n$.
Il modello statistico per un campione casuale binomiale è quindi individuato da:

$$\left\{\mathcal{X}^n = \{0, 1, \ldots, k\}^n, \quad f_n(\mathbf{x}_n;\theta) = \left[\prod_{i=1}^{n}\binom{k}{x_i}\right]\theta^{n\overline{x}_n}\,(1-\theta)^{n(k-\overline{x}_n)}, \quad \theta \in \Theta = [0,1]\right\}.$$

□

1.36 Esempio (Modello geometrico). Supponiamo che $X_1, \ldots, X_n$ sia un campione casuale geometrico di parametro incognito θ:

$$X_1, \ldots, X_n|\theta \quad \text{i.i.d,} \qquad\qquad X_i|\theta \sim \text{Geom}(\theta).$$

Dato che lo spazio $\mathcal{X}$ per la variabile di base coincide con $\overline{\mathbb{N}} = \{0, 1, \ldots\}$, ogni campione potenzialmente osservabile è costituito da n numeri ciascuno dei quali è un numero appartenente all'insieme dei numeri naturali. Lo spazio $\mathcal{X}^n$ è quindi:

$$\mathcal{X}^n = \underbrace{\overline{\mathbb{N}} \times \overline{\mathbb{N}} \times \ldots \times \overline{\mathbb{N}}}_{n\ volte} = (\overline{\mathbb{N}})^n.$$

Ricordando che per la v.a. geometrica di parametro θ si ha

$$f_X(x;\theta) = \mathbb{P}(X = x;\theta) = (1-\theta)^x\theta, \qquad x = 0, 1, 2, \ldots,$$

dalla (1.1) si ottiene

$$f_n(\mathbf{x}_n;\theta) = \prod_{i=1}^{n}(1-\theta)^{x_i}\theta = (1-\theta)^{\sum_{i=1}^n x_i}\theta^n.$$

Il modello statistico per un campione casuale geometrico è quindi individuato da:

$$\left\{\mathcal{X}^n = (\overline{\mathbb{N}})^n, \quad f_n(\mathbf{x}_n;\theta) = (1-\theta)^{n\overline{x}_n}\theta^n, \quad \theta \in \Theta = [0,1]\right\}.$$

□

1.37 Esempio (Modello di Poisson). Supponiamo che $X_1, \ldots, X_n$ sia un campione casuale poissoniano di parametro incognito θ:

$$X_1, \ldots, X_n|\theta \quad \text{i.i.d.,} \qquad\qquad X_i|\theta \sim \text{Pois}(\theta).$$

Dato che lo spazio $\mathcal{X}$ per la variabile di base coincide con $\overline{\mathbb{N}}$ si ha che $\mathcal{X}^n = (\overline{\mathbb{N}})^n$. Ricordando che, per la v.a. di Poisson di parametro θ,

$$f_X(x;\theta) = \mathbb{P}(X = x;\theta) = \frac{e^{-\theta}\,\theta^x}{x!} \qquad x = 0, 1, \ldots,$$

per la (1.1),

$$\begin{aligned} f_n(\mathbf{x}_n;\theta) &= \prod_{i=1}^n f_X(x_i;\theta) = \prod_{i=1}^n \left[\frac{e^{-\theta}\,\theta^{x_i}}{x_i!}\right] \\ &= \frac{e^{-n\,\theta}\;\theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} = \frac{1}{\prod_{i=1}^n x_i!} e^{-n\,\theta}\theta^{n\overline{x}_n} \end{aligned}$$

Il modello statistico per un campione casuale di Poisson è quindi individuato da:

$$\left\{\mathcal{X}^n = (\overline{\mathbb{N}})^n, \quad f_n(\mathbf{x}_n;\theta) = \frac{1}{\prod_{i=1}^n x_i!} e^{-n\,\theta}\theta^{n\overline{x}_n}, \quad \theta \in \Theta = \mathbb{R}^+\right\}.$$

□

1.38 Esempio (Modello esponenziale negativo). Supponiamo che $X_1, \ldots, X_n$ sia un campione casuale esponenziale negativo di parametro incognito θ:

$$X_1, \ldots, X_n|\theta \quad \text{i.i.d.}, \qquad\qquad X_i|\theta \sim \text{EN}(\theta).$$

Dato che per la v.a. di base, lo spazio $\mathcal{X}$ dei possibili risultati è l'insieme $\mathbb{R}^+$, ogni campione di n elementi potenzialmente osservabile è costituito da n numeri, ciascuno dei quali può essere un numero appartenente all'insieme dei numeri reali positivi. Lo spazio $\mathcal{X}^n$ è:

$$\mathcal{X}^n = \underbrace{\mathbb{R}^+ \times \mathbb{R}^+ \ldots \times \mathbb{R}^+}_{n\ volte} = (\mathbb{R}^+)^n.$$

Ricordando che, per una v.a. esponenziale negativa di parametro θ,

$$f_X(x;\theta) = \theta\, e^{-\theta x}, \qquad x > 0$$

si avrà[9]

$$f_n(\mathbf{x}_n;\theta) = \prod_{i=1}^n f_X(x_i;\theta) = \prod_{i=1}^n \theta\, e^{-\theta x_i} = \theta^n\, e^{-\theta\sum_{i=1}^n x_i}, \qquad x_i > 0.$$

Il modello statistico per un campione casuale esponenziale negativo è quindi[10]:

$$\left\{\mathcal{X}^n = (\mathbb{R}^+)^n, \quad f_n(\mathbf{x}_n;\theta) = \theta^n\, e^{-\theta n\overline{x}_n}, \quad \theta \in \Theta = \mathbb{R}^+\right\}.$$

□

1.39 Esempio (Modello esponenziale). Per quanto visto nell'esempio precedente, possiamo dire che il modello statistico per un campione casuale esponenziale è:

$$\left\{\mathcal{X}^n = (\mathbb{R}^+)^n, \quad f_n(\mathbf{x}_n;\theta) = \theta^{-n}\, e^{-\frac{n\overline{x}_n}{\theta}}, \quad \theta \in \Theta = \mathbb{R}^+\right\}.$$

[9] Per rendere la notazione più leggera usiamo le funzioni indicatrici del supporto solo nei casi in qui questo dipende dal parametro incognito.

[10] A rigore, qui e negli esempi successivi, dovremmo includere la produttoria della funzioni indicatrici: $\prod_{i=1}^n I_{\mathbb{R}^+}(x_i)$. Evitiamo di farlo per alleggerire la notazione.

□

1.40 Esempio (Modello normale). Supponiamo che $X_1, \ldots, X_n$ sia un campione casuale normale di parametro incognito $\theta = (\theta_1, \theta_2)$:

$$X_1, \ldots, X_n|\theta \quad \text{i.i.d.}, \qquad\qquad X_i|\theta \sim \mathrm{N}(\theta_1, \theta_2).$$

Dato che lo spazio $\mathcal{X}$ per la variabile di base coincide con $\mathbb{R}$ si ha che $\mathcal{X}^n = \mathbb{R}^n$.
Ricordando che, per la v.a. normale di parametro $\theta = (\theta_1, \theta_2)$,

$$f_X(x;\theta) = \frac{1}{\sqrt{2\pi\theta_2}} \exp\left\{-\frac{1}{2\theta_2}(x-\theta_1)^2\right\}, \qquad x \in \mathbb{R},$$

per la (1.1)

$$f_n(\mathbf{x}_n;\theta) \quad = \quad \prod_{i=1}^n f_X(x_i;\theta) = (2\pi\theta_2)^{-n/2} \exp\left\{-\frac{1}{2\theta_2}\sum_{i=1}^n (x_i-\theta_1)^2\right\}, \qquad x_i \in \mathbb{R}.$$

Il modello statistico per un campione casuale normale è:

$$\left\{\mathcal{X}^n = \mathbb{R}^n, \quad f_n(\mathbf{x}_n;\theta) = (2\pi\theta_2)^{-n/2} \exp\left\{-\frac{1}{2\theta_2}\sum_{i=1}^n (x_i-\theta_1)^2\right\}, \quad \theta \in \Theta = \mathbb{R}\times\mathbb{R}^+\right\}.$$

□

1.41 Esempio (Modello normale con varianza nota). In modo analogo, il modello statistico per un campione casuale normale con varianza nota, $\theta_2 = \sigma_0^2$ è:

$$\left\{\mathcal{X}^n = \mathbb{R}^n, \quad f_n(\mathbf{x}_n;\theta) = \frac{1}{(2\pi\sigma_0^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^n (x_i-\theta)^2\right\}, \quad \theta \in \Theta = \mathbb{R}\right\}.$$

□

1.42 Esempio (Modello normale con valore atteso noto) Il modello statistico per un campione casuale normale con media nota, $\theta_1 = \mu_0$ è:

$$\left\{\mathcal{X}^n = \mathbb{R}^n, \quad f_n(\mathbf{x}_n;\theta) = \frac{1}{(2\pi\theta)^{n/2}} \exp\left\{-\frac{1}{2\theta}\sum_{i=1}^n (x_i-\mu_0)^2\right\}, \quad \theta \in \Theta = \mathbb{R}^+\right\}.$$

□

1.43 Esempio (Modello gamma). Supponiamo che $X_1, \ldots, X_n$ sia un campione casuale gamma di parametro incognito $\theta = (\theta_1, \theta_2)$:

$$X_1, \ldots, X_n \quad \text{i.i.d.}, \qquad\qquad X_i|\theta \sim \mathrm{Gamma}(\theta_1, \theta_2).$$

Dato che lo spazio $\mathcal{X}$ per la variabile di base coincide con $\mathbb{R}^+$ si ha che $\mathcal{X}^n = (\mathbb{R}^+)^n$.
Ricordando che, per la v.a. gamma di parametro $\theta = (\theta_1, \theta_2)$,

$$f_X(x;\theta) = \frac{\theta_2^{\theta_1}}{\Gamma(\theta_1)} x^{\theta_1-1} e^{-\theta_2 x}, \qquad x > 0.$$

per la (1.1)

$$\begin{aligned} f_n(\mathbf{x}_n;\theta) &= \prod_{i=1}^{n} f_X(x_i;\theta) = \prod_{i=1}^{n} \frac{\theta_2^{\theta_1}}{\Gamma(\theta_1)} x_i^{\theta_1-1} e^{-\theta_2 x_i} \\ &= \left[\frac{\theta_2^{\theta_1}}{\Gamma(\theta_1)}\right]^n e^{-\theta_2 \sum_{i=1}^n x_i} \prod_{i=1}^{n} x_i^{\theta_1-1}, \qquad x_i > 0 \end{aligned}$$

e lo spazio $\Theta = \mathbb{R}^+ \times \mathbb{R}^+ = (\mathbb{R}^+)^2$. □

1.44 Esempio (Modello di Cauchy). Se $X_1, \ldots, X_n$ è un campione casuale di Cauchy di parametro incognito $\theta = (\theta_1, \theta_2)$, il modello statistico è

$$\left\{ \mathcal{X}^n = \mathbb{R}^n, \quad f_n(\mathbf{x}_n;\theta) = \frac{(\pi\ \theta_2)^{-n}}{\prod_{i=1}^n \left[1 + \left(\frac{x_i-\theta_1}{\theta_2}\right)^2\right]}, \quad \theta \in \Theta = \mathbb{R} \times \mathbb{R}^+ \right\}.$$

□

1.45 Esempio (Modello beta). Se $X_1, \ldots, X_n$ è un campione casuale beta di parametro incognito $\boldsymbol{\theta} = (\theta_1, \theta_2)$, il modello statistico è

$$\left\{ \mathcal{X}^n = [0,1]^n, \quad f_n(\mathbf{x}_n;\theta) = B(\theta_1,\theta_2)^{-n} \prod_{i=1}^{n} \left[x_i^{\theta_1-1}(1-x_i)^{\theta_2-1}\right], \quad \theta \in \Theta = \mathbb{R}^+ \times \mathbb{R}^+ \right\}.$$

□

Modelli con supporto dipendente dal parametro

Nel caso di campioni casuali per v.a. con supporto $\mathcal{X}_\theta$ dipendente dal parametro incognito, è consigliabile ricorrere all'uso delle funzioni indicatrici del supporto stesso. Illustriamo come procedere con il seguente esempio.

1.46 Esempio (Modello uniforme). Nel caso di un campione casuale estratto da una v.a. $X \sim \text{Unif}(\theta_1, \theta_2)$ il supporto della v.a. di base dipende da $\boldsymbol{\theta} = [\theta_1, \theta_2]$. Ogni campione potenzialmente osservabile è costituito da n numeri che appartengono ad un ipercubo nello spazio $\mathbb{R}^n$ che ha per spigoli gli intervalli $[\theta_1, \theta_2]$ e cioè:

$$\mathcal{X}_\theta^n = \underbrace{[\theta_1,\theta_2] \times [\theta_1,\theta_2] \ldots \times [\theta_1,\theta_2]}_{n\ volte} = \big([\theta_1,\theta_2]\big)^n.$$

Ricordiamo che, per una v.a. uniforme di parametro $\boldsymbol{\theta} = (\theta_1, \theta_2)$.

$$f_X(x;\boldsymbol{\theta}) = \begin{cases} (\theta_2-\theta_1)^{-1} & \text{per} \quad \theta_1 \le x \le \theta_2 \\ 0 & \text{altrove} \end{cases} = \frac{1}{\theta_2-\theta_1}\ I_{[\theta_1,\theta_2]}(x).$$

Si ha quindi che:

$$f_n(\mathbf{x}_n;\boldsymbol{\theta}) = \prod_{i=1}^{n} f_X(x_i;\boldsymbol{\theta}) = (\theta_2-\theta_1)^{-n} \prod_{i=1}^{n} I_{[\theta_1,\theta_2]}(x_i).$$

Si noti che

$$\prod_{i=1}^{n} I_{[\theta_1,\theta_2]}(x_i) = I_{(-\infty,x_{(1)}]}(\theta_1) \times I_{[x_{(n)},\infty)}(\theta_2),$$

dove $x_{(1)} = \min\{x_1, \ldots, x_n\}$ e $x_{(n)} = \max\{x_1, \ldots, x_n\}$ Il modello statistico per il campione casuale è quindi

$$\left\{ \mathcal{X}_\theta^n = \big([\theta_1, \theta_2]\big)^n,\ f_n(\mathbf{x}_n; \boldsymbol{\theta}) = \frac{1}{(\theta_2 - \theta_1)^n} I_{(-\infty,x_{(1)}]}(\theta_1) I_{[x_{(n)},+\infty)}(\theta_2),\ \Theta = \big\{(\theta_1, \theta_2) \in \mathbb{R}\times\mathbb{R},\ \theta_1 < \theta_2\big\} \right\}.$$

□

1.3.5 Famiglie esponenziali

Molti dei principali modelli parametrici che si utilizzano nelle applicazioni comuni sono casi particolari di un gruppo di modelli denominato *classe delle famiglie esponenziali.* I modelli che possono essere rappresentati come famiglie esponenziali presentano importanti proprietà inferenziali. È quindi utile verificare se un modello statistico è un membro di tale famiglia. Per semplicità, distinguiamo il caso uniparametrico da quello multiparametrico.

A - Famiglie esponenziali uniparametriche

1.47 Definizione. Una famiglia di leggi di probabilità $\mathcal{F} = (f_X(x;\theta),\ \theta \in \Theta)$ con supporto che non dipende dal parametro è detta *famiglia esponenziale uniparametrica* se $\Theta \subseteq \mathbb{R}$ e se esistono delle funzioni $h(x)$, $\eta(\theta)$, $T(x)$ e $B(\theta)$ tali che:

$$f_X(x;\theta) = h(x) \exp\{\eta(\theta)T(x) - B(\theta)\}. \tag{1.3}$$

La funzione $\eta(\theta)$ si chiama *parametro naturale* della famiglia esponenziale. □

Gli esempi che seguono mostrano che molti dei modelli parametrici considerati nei precedenti paragrafi sono famiglie esponenziali.

1.48 Esempio (Modello bernoulliano). Sia $X|\theta \sim \text{Ber}(\theta)$. Si ha allora che:

$$\begin{aligned} f_X(x;\theta) &= \theta^x (1-\theta)^{1-x} \\ &= \exp\big\{\log\big[\theta^x (1-\theta)^{1-x}\big]\big\} \\ &= \exp\big\{x \log\theta + (1-x)\log(1-\theta)\big\} \\ &= \exp\big\{x \log\frac{\theta}{1-\theta} + \log(1-\theta)\big\}. \end{aligned}$$

In questo caso la legge di probabilità può essere scritta nella forma (1.3), con

$$h(x) = 1, \qquad \eta(\theta) = \log\frac{\theta}{1-\theta}, \qquad T(x) = x, \qquad B(\theta) = -\log(1-\theta).$$

Il modello bernoulliano è quindi una famiglia esponenziale con parametro naturale uguale al logaritmo del rapporto $\theta/(1-\theta)$, denominato *odds*. Il parametro naturale $\eta(\theta) = \log[\theta/(1-\theta)]$ prende il nome di *log-odds*. □

1.49 Esempio (Modello binomiale). Sia $X|\theta \sim \text{Bin}(k, \theta)$. In modo analogo all'esempio precedente si può verificare che il modello binomiale è una famiglia esponenziale con

$$h(x) = \binom{k}{x}, \qquad \eta(\theta) = \log\frac{\theta}{1-\theta}, \qquad T(x) = x, \qquad B(\theta) = -k\log(1-\theta).$$

Anche in questo caso il parametro naturale della famiglia esponenziale è dato dal *log-odds*. □

1.50 Esempio (Modello di Poisson). Sia $X|\theta \sim \text{Pois}(\theta)$, allora

$$f_X(x;\theta) = \frac{\theta^x}{x!}\, e^{-\theta} = \frac{1}{x!}\exp\left\{x\log\theta - \theta\right\}.$$

Ponendo

$$h(x) = \frac{1}{x!}, \quad \eta(\theta) = \log\theta, \quad T(x) = x, \quad B(\theta) = \theta,$$

si verifica che il modello $\text{Pois}(\theta)$ è una famiglia esponenziale con parametro naturale $\eta(\theta) = \log\theta$. □

1.51 Esempio (Modello esponenziale negativo). Se $X|\theta \sim \text{EN}(\theta)$, si ha

$$f_X(x;\theta) = \theta\exp\{-\theta x\} = \exp\{-\theta x + \log\theta\},$$

e dunque il modello $\text{EN}(\theta)$ è una famiglia esponenziale con

$$h(x) = 1, \quad \eta(\theta) = -\theta, \quad T(x) = x, \quad B(\theta) = -\log\theta.$$

□

1.52 Esempio (Modello normale con varianza nota). Sia $X|\theta \sim \text{N}(\theta, \sigma_0^2)$, con σ_0^2 noto. In questo caso:

$$\begin{aligned} f_X(x;\theta) &= \frac{1}{\sqrt{2\pi}\sigma_0}\exp\left\{-\frac{1}{2\sigma_0^2}\,(x-\theta)^2\right\} \\ &= \frac{1}{\sqrt{2\pi}\sigma_0}\exp\left\{-\frac{x^2}{2\sigma_0^2} + \frac{x\theta}{\sigma_0^2} - \frac{\theta^2}{2\sigma_0^2}\right\} \\ &= \frac{1}{\sqrt{2\pi}\sigma_0}\exp\left\{-\frac{x^2}{2\sigma_0^2}\right\}\exp\left\{\frac{\theta\ x}{\sigma_0^2} - \frac{\theta^2}{2\sigma_0^2}\right\}, \end{aligned}$$

e si conclude che il modello $\text{N}(\theta, \sigma_0^2)$ è una famiglia esponenziale con

$$h(x) = \frac{1}{\sqrt{2\pi}\sigma_0}\exp\left\{-\frac{x^2}{2\sigma_0^2}\right\}, \quad \eta(\theta) = \frac{\theta}{\sigma_0^2}, \quad T(x) = x, \quad B(\theta) = \frac{\theta^2}{2\sigma_0^2}.$$

□

1.53 Esempio (Modello normale con valore atteso noto). Sia $X|\theta \sim \text{N}(\mu_0, \theta)$, con μ_0 nota. Poiché

$$\begin{aligned} f_X(x;\theta) &= \frac{1}{\sqrt{2\pi\theta}}\exp\left\{-\frac{1}{2\theta}(x-\mu_0)^2\right\} \\ &= \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2\theta}(x-\mu_0)^2 - \frac{1}{2}\log\theta\right\}, \end{aligned}$$

e quindi il modello $\text{N}(\mu_0, \theta)$ è una famiglia esponenziale con

$$h(x) = \frac{1}{\sqrt{2\pi}}, \quad \eta(\theta) = -\frac{1}{2\theta}, \quad T(x) = (x-\mu_0)^2, \quad B(\theta) = \frac{1}{2}\log\theta.$$

□

1.54 Esempio (Modello uniforme). Per il modello uniforme (che è un modello non regolare) si ha che:

$$f_X(x;\theta) = \frac{1}{\theta} I_{[0,\theta]}(x), \qquad \theta > 0.$$

Si osservi che la funzione indicatrice non può essere espressa in termini di una funzione esponenziale che a sua volta sia funzione sia di x che di θ. Non è possibile quindi determinare delle funzioni h, η, T e B che consentano la rappresentazione in (1.3). Il modello uniforme non è quindi una famiglia esponenziale. □

B - Famiglie esponenziali multiparametriche

La definizione di famiglia esponenziale è facilmente estesa al caso in cui il parametro del modello sia vettoriale: $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$.

1.55 Definizione. La famiglia parametrica $\mathcal{F} = (f_X(x;\boldsymbol{\theta}),\ \boldsymbol{\theta} \in \Theta)$, con supporto non dipendente da parametri, è detta *famiglia esponenziale k-parametrica* se $\Theta \subseteq \mathbb{R}^k, k > 1$, e se esistono delle funzioni di x e $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ $h(x)$, $\eta_1(\boldsymbol{\theta}), \ldots, \eta_k(\boldsymbol{\theta})$, $T_1(x), \ldots, T_k(x)$ e $B(\boldsymbol{\theta})$ tali che:

$$f_X(x;\boldsymbol{\theta}) = h(x) \exp\left\{ \sum_{j=1}^{k} \eta_j(\boldsymbol{\theta}) T_j(x) - B(\boldsymbol{\theta}) \right\}.$$

Le funzioni $\eta_1(\boldsymbol{\theta}), \ldots, \eta_k(\boldsymbol{\theta})$ si chiamano *parametri naturali* della famiglia esponenziale. □

1.56 Esempio (Modello normale). Sia $X|\boldsymbol{\theta} \sim \mathrm{N}(\theta_1, \theta_2)$, con entrambi i parametri incogniti. Si ha:

$$f_X(x;\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\theta_2}} \exp\left\{ -\frac{1}{2\theta_2} x^2 + \frac{\theta_1}{\theta_2} x - \frac{\theta_1^2}{2\theta_2} \right\} = \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2\theta_2} x^2 + \frac{\theta_1}{\theta_2} x - \frac{\theta_1^2}{2\theta_2} - \frac{1}{2}\log\theta_2 \right\}.$$

Ponendo quindi:

$$h(x) = \frac{1}{\sqrt{2\pi}} I_{\mathbb{R}}(x), \quad \eta_1(\boldsymbol{\theta}) = -\frac{1}{2\theta_2}, \quad \eta_2(\boldsymbol{\theta}) = \frac{\theta_1}{\theta_2}, \quad T_1(x) = x^2, \quad T_2(x) = x, \quad B(\boldsymbol{\theta}) = \frac{\theta_1^2}{2\theta_2} + \frac{1}{2}\log\theta_2,$$

si verifica che la famiglia $\mathrm{N}(\theta_1, \theta_2)$ è una famiglia esponenziale con $k = 2$. □

1.57 Esempio (Modello gamma). Sia $X|\boldsymbol{\theta} \sim \mathrm{Ga}(\theta_1, \theta_2)$, allora

$$f_X(x;\boldsymbol{\theta}) = \frac{\theta_2^{\theta_1}}{\Gamma(\theta_1)}\, x^{\theta_1 - 1} e^{-\theta_2 x} = \exp\big\{ \theta_1 \log\theta_2 - \log\big[\Gamma(\theta_1)\big] + (\theta_1 - 1)\log x - \theta_2 x \big\}.$$

Quindi un modello gamma con parametro vettoriale $\boldsymbol{\theta} = (\theta_1, \theta_2)$ è una famiglia esponenziale dove

$$h(x) = 1 \quad \eta_1(\boldsymbol{\theta}) = \theta_1 - 1 \quad \eta_2(\theta) = -\theta_2, \quad T_1(x) = \log x, \quad T_2(x) = x, \quad B(\boldsymbol{\theta}) = -\theta_1\log\theta_2 + \log\big[\Gamma(\theta_1)\big],$$

ed i suoi parametri naturali sono: $\eta_1(\theta) = \theta_1 - 1$ e $\eta_2(\theta) = \theta_2$. □

1.58 Esempio (Modello di Cauchy). Sia $X|\boldsymbol{\theta} \sim \mathrm{Cau}(\theta_1, \theta_2)$. Si ha che

$$f_X(x;\boldsymbol{\theta}) = \frac{1}{\pi\theta_2}\, \frac{1}{1 + \left(\dfrac{x - \theta_1}{\theta_2}\right)^2} = \frac{1}{\pi} \exp\left\{ -\log\theta_2 - \log\left(1 + \left[\frac{x - \theta_1}{\theta_2} \right]^2 \right) \right\}.$$

Dall'espressione della $f_X(x;\theta_1,\theta_2)$ si vede come non è possibile trovare funzioni η_j e T_j, per $j=1,2$, che consentano la rappresentazione caratterizzante di una famiglia esponenziale. Dunque il modello di Cauchy non è una famiglia esponenziale. □

1.3.6 Modelli statistici per campioni casuali di famiglie esponenziali

A - Famiglie esponenziali uniparametriche
Ricordando l'espressione (1.3) che definisce una famiglia esponenziale, considerato un campione casuale $\mathbf{X}_n=(X_1,\dots,X_n)$, proveniente dalla variabile di base X con legge di probabilità $\{f_X(x;\theta)=h(x)\exp\{\eta(\theta)T(x)-B(\theta)\}$, $\theta\in\Theta$, si verifica immediatamente che

$$\begin{aligned} f_n(\mathbf{x}_n;\theta) &= \prod_{i=1}^n f_X(x_i;\theta)=\prod_{i=1}^n [h(x_i)\exp\{\eta(\theta)T(x_i)-B(\theta)\}] \\ &= \prod_{i=1}^n h(x_i)\times\exp\{\eta(\theta)\sum_{i=1}^n T(x_i)-nB(\theta)\}. \end{aligned}$$

Pertanto

$$f_n(\mathbf{x}_n;\theta)=h_n(\mathbf{x}_n)\exp\{\eta(\theta)T_n(\mathbf{x}_n)-B_n(\theta)\}, \tag{1.4}$$

dove

$$h_n(\mathbf{x}_n)=\prod_{i=1}^n h(x_i), \qquad T_n(\mathbf{x}_n)=\sum_{i=1}^n T(x_i), \qquad B_n(\theta)=nB(\theta).$$

Per la (1.3), la famiglia $\mathcal{F}_n=\big(f_n(\mathbf{x}_n;\theta)=\prod_{i=1}^n f_X(x_i;\theta),\ \theta\in\Theta\big)$ è una famiglia esponenziale, caratterizzata dalle funzioni $h_n(\mathbf{x}_n),\eta(\theta),T_n(\mathbf{x}_n)$ e $B_n(\theta)$.

1.59 Esempio (Modello bernoulliano). Sia $X_1,\dots,X_n$, un campione casuale, con $X_i|\theta\sim\text{Ber}(\theta)$, $i=1,\dots,n$. È semplice verificare che, in questo caso, $f_n(\mathbf{x}_n;\theta)$ è una famiglia esponenziale uniparametrica con

$$h_n(\mathbf{x}_n)=1, \qquad \eta(\theta)=\log\left(\frac{\theta}{1-\theta}\right), \qquad T_n(\mathbf{x}_n)=\sum_{i=1}^n x_i, \qquad B_n(\theta)=-n\log(1-\theta).$$

□

1.60 Esempio (Modello di Poisson). Sia $X_1,\dots,X_n$, un campione casuale, con $X_i|\theta\sim\text{Pois}(\theta)$, $i=1,\dots,n$. È semplice verificare che, in questo caso, $f_n(\mathbf{x}_n;\theta)$ è una famiglia esponenziale uniparametrica con

$$h_n(\mathbf{x}_n)=\frac{1}{\prod_{i=1}^n x_i!}, \qquad \eta(\theta)=\log\theta, \qquad T_n(\mathbf{x}_n)=\sum_{i=1}^n x_i, \qquad B_n(\theta)=n\theta.$$

□

B - Famiglie esponenziali multiparametriche
Nel caso k-parametrico, in modo analogo, è immediato verificare che:

$$\begin{aligned} f_n(\mathbf{x}_n;\boldsymbol{\theta}) &= \prod_{i=1}^n\big[h(x_i)\big]\ \exp\Big\{\sum_{j=1}^k \eta_j(\boldsymbol{\theta})T_j(x_i)-B(\boldsymbol{\theta})\Big\} \\ &= h_n(\mathbf{x}_n)\ \exp\Big\{\sum_{j=1}^k \eta_j(\boldsymbol{\theta})\sum_{i=1}^n T_j(x_i)-nB(\boldsymbol{\theta})\Big\} \\ &= h_n(\mathbf{x}_n)\ \exp\Big\{\sum_{j=1}^k \eta_j(\boldsymbol{\theta})T_{n,j}(\mathbf{x}_n)-B_n(\boldsymbol{\theta})\Big\}, \end{aligned}$$

ovvero che $f_n(\mathbf{x}_n;\boldsymbol{\theta})$ è una famiglia esponenziale caratterizzata dalle funzioni:

$$h_n(\mathbf{x}_n),\ \ \eta_1(\boldsymbol{\theta}),\dots,\eta_k(\boldsymbol{\theta}),\ \ T_{n,1}(\mathbf{x}_n),\dots,T_{n,k}(\mathbf{x}_n),\ \ B_n(\boldsymbol{\theta}),$$

con

$$T_{n,j}(\mathbf{x}_n)=\sum_{i=1}^{n}T_j(x_i),\ j=1,\dots,k.$$

1.61 Esempio (Modello normale). Consideriamo il modello $\mathrm{N}(\theta_1,\theta_2)$. Indicando con $\boldsymbol{\theta}=(\theta_1,\theta_2)$ si ha:

$$f_X(x;\theta_1,\theta_2)=\frac{1}{\sqrt{2\pi}}\exp\left\{\frac{\theta_1}{\theta_2}x-\frac{x^2}{2\theta_2}-\frac{1}{2}\left(\frac{\theta_1^2}{\theta_2}+\ln(\theta_2)\right)\right\}.$$

Si tratta di una famiglia esponenziale con

$$\eta_1(\boldsymbol{\theta})=\frac{\theta_1}{\theta_2},\quad T_1(x)=x,\quad \eta_2(\boldsymbol{\theta})=-\frac{1}{2\theta_2},\quad T_2(x)=x^2,$$

$$B(\boldsymbol{\theta})=\frac{1}{2}\left(\frac{\theta_1^2}{\theta_2}+\ln(2\pi\theta_2)\right),\quad h(x)=\frac{1}{\sqrt{2\pi}}.$$

Nel caso di un campione casuale di dimensione n si ha che

$$\left(\sum_{i=1}^{n}T_1(x_i),\sum_{i=1}^{n}T_2(x_i)\right)=\left(\sum_{i=1}^{n}x_i,\sum_{i=1}^{n}x_i^2\right).$$

□

1.4 I principali problemi inferenziali

La classificazione dei problemi inferenziali non è univoca. Diamo qui una classificazione, non necessariamente esaustiva ma abbastanza generale, utile ai fini didattici di questo testo. Distinguiamo innanzitutto i problemi *post-sperimentali* da quelli *pre-sperimentali*. I problemi post-sperimentali sono quelli che prevedono l'uso di dati campionari. Rientrano tra questi i più usuali problemi, come ad esempio la stima del parametro incognito di un modello. I problemi pre-sperimentali sono quelli che si affrontano prima ancora di osservare i dati e consistono essenzialmente nel *disegno dell'esperimento*. Lo scopo è scegliere in modo ottimale gli elementi di un esperimento su cui lo sperimentatore e lo statistico hanno la possibilità di intervenire. L'ottimalità della scelta si riferisce qui all'informatività e all'efficienza dell'esperimento stesso. Un tipico problema di scelta dell'esperimento consiste nello stabilire il numero di osservazioni da includere nel campione. In questo caso, ad esempio, si vuole scegliere un numero di osservazioni adeguato a garantire, entro determinati limiti di costo, sufficiente accuratezza della procedura inferenziale adottata.
Consideriamo ora con maggiore dettaglio i problemi post-sperimentali. Dato un modello statistico parametrico $\{\mathcal{X}^n,\ f_n(\cdot|\theta),\ \theta\in\Theta\}$, possiamo individuare due classi di problemi inferenziali: problemi *ipotetici* e problemi *predittivi*.

Problemi ipotetici. Si tratta della classe di problemi a cui abbiamo già accennato, riguardanti l'uso dei dati campionari al fine di acquisire informazioni sul parametro incognito del modello. L'obiettivo è quindi valutare, con la maggiore accuratezza possibile, il valore vero θ^* che individua il meccanismo aleatorio che ha generato le osservazioni. Tradizionalmente i problemi ipotetici si articolano in tre sottogruppi:

(a) *Stima puntuale.* I dati campionari vengono sintetizzati attraverso una opportuna funzione T in un valore numerico $T(\mathbf{x}_n) = t$, chiamato *stima* del parametro. Evidentemente è opportuno richiedere che, per ogni campione osservato, il valore t sia un punto dello spazio parametrico Θ.

(b) *Stima mediante insiemi.* Si ricerca un insieme $S(\mathbf{x}_n) \subseteq \Theta$ (con S funzione dei dati campionari) che contenga il valore incognito θ^*. L'obiettivo della stima per insiemi è quindi meno ambizioso di quello della stima puntuale, anche se la maggiore approssimazione della conclusione è compensata da una valutazione della affidabilità che l'insieme di stima contenga θ^*. Nei più comuni problemi uniparametrici in cui $\Theta \subset \mathbb{R}$, gli insiemi di stima S sono spesso degli intervalli. In questi casi, si ha che $S(\mathbf{x}_n) = [L(\mathbf{x}_n), U(\mathbf{x}_n)]$, dove L ed U sono due opportune funzioni dei dati campionari tali che, per ogni possibile campione osservato, $L(\mathbf{x}_n) \leq U(\mathbf{x}_n)$.

(c) *Verifica di ipotesi.* In questi problemi si considera una opportuna partizione dello spazio Θ, ovvero una suddivisione in due sottospazi Θ_0 e Θ_1, la cui intersezione è vuota e la cui unione coincide con l'intero spazio Θ. Sulla base del campione osservato si deve quindi decidere se più plausibile l'ipotesi che $\theta^* \in \Theta_0$ oppure l'ipotesi complementare che $\theta^* \in \Theta_1$. Il caso più elementare consiste nel confronto tra *ipotesi semplici*, in cui Θ_0 e Θ_1 consistono rispettivamente in due punti, θ_0 e θ_1.

Problemi predittivi. Nei problemi predittivi si utilizza l'informazione fornita dal campione $\mathbf{x}_n$ osservato in un esperimento statistico per prevedere il valore del risultato di un esperimento futuro governato da una legge di probabilità che condivide lo stesso parametro θ^* con la legge che governa l'esperimento che ha generato $\mathbf{x}_n$. Anche quelli di tipo predittivo si possono suddividere in problemi di stima puntuale, stima mediante insiemi e verifica di ipotesi, riferiti in questo caso al risultato dell'esperimento futuro.
È importante osservare che la differenza principale tra problemi ipotetici e predittivi consiste nel fatto che, nel primo caso, l'oggetto di interesse (θ) non è osservabile, mentre nel secondo caso è una variabile effettivamente osservabile.

Altri rilevanti problemi inferenziali sono i seguenti.

(a) *Scelta tra modelli.* Le considerazioni precedenti presuppongono la scelta preliminare di un determinato modello statistico. Ad esempio, si dà per scontato il fatto che i dati provengano da una popolazione normale o da una popolazione di Cauchy. Un problema centrale dell'inferenza statistica consiste proprio nella scelta di uno tra più possibili modelli alternativi, sulla base dei dati osservati. Scelto il modello, si possono poi affrontare i problemi ipotetici e/o predittivi eventuali.

(b) *Controllo del modello.* In questo caso si è interessati a verificare la plausibilità di un modello adottato alla luce dei dati effettivamente osservati. Ad esempio, attraverso opportuni strumenti, può essere necessario valutare e quantificare la plausibilità che i dati provengano da una popolazione normale. Nel caso in cui i dati osservati facciano sorgere dubbi sul modello assunto, è in genere necessario mettere in discussione tale assunzione prima di affrontare ogni problema ipotetico e/o predittivo.

Questo testo si concentra sui tre problemi inferenziali di tipo ipotetico e a questi ci riferiremo da questo punto in poi.

1.5 Le principali impostazioni inferenziali

La teoria dell'inferenza statistica non è unitaria. Esistono infatti diverse impostazioni o logiche inferenziali che, pur affrontando gli stessi problemi (pre- e post-sperimentali, ipotetici o predittivi, etc.) si basano su presupposti diversi e propongono, spesso, soluzioni che divergono anche dal punto di vista operativo. Le tre principali impostazioni inferenziali, che sono quelle a cui ci riferiamo in questo testo, sono: impostazione basata sull'analisi della funzione di verosimiglianza; impostazione frequentista; impostazione bayesiana.

- *Impostazione basata sull'analisi della funzione di verosimiglianza.* Si basa su elaborazioni di una funzione, denominata *funzione di verosimiglianza* (fdv), che quantifica il grado di plausibilità che ciascun valore possibile del parametro incognito θ assume alla luce dei dati campionari effettivamente osservati. Opportuni usi della fdv consentono di affrontare i principali problemi inferenziali.

- *Impostazione frequentista.* Le procedure inferenziali frequentiste vengono selezionate in funzione del loro comportamento valutato complessivamente sull'intero spazio dei campioni, e non in riferimento a un campione effettivamente osservato. Ad esempio, nel caso della stima puntuale, si cercano funzioni dei dati campionari che con buona probabilità (nello spazio dei campioni) diano luogo a valori di stima non lontani da θ^*. Si tratta dell'impostazione inferenziale attualmente più diffusa e più comunemente utilizzata, alla quale viene dedicata ampia parte di questo testo.

- *Impostazione bayesiana.* Lo schema logico bayesiano presuppone che l'incertezza sul parametro incognito del modello sia rappresentata attraverso una distribuzione di probabilità. Tutte le informazioni pre-sperimentali su θ concorrono pertanto a precisare questa legge di probabilità, detta *distribuzione a priori del parametro.* Tale fonte di informazioni su θ viene combinata con l'informazione che il campione di dati osservato fornisce su θ attraverso la funzione di verosimiglianza. Lo strumento per effettuare questa sintesi è la formula di Bayes, con la quale si ottiene la *distribuzione a posteriori del parametro.* Da opportune sintesi della distribuzione a posteriori si ottengono gli strumenti per affrontare i principali problemi inferenziali.

1.6 Appendice

1.6.1 Funzione di ripartizione, funzione di massa di probabilità e funzione di densità

Richiamiamo brevemente alcuni concetti probabilistici di base.

1.62 Definizione La *funzione di ripartizione* (f.r.) di una v.a. semplice (unidimensionale) X è una funzione

$$F_X : \mathbb{R} \to [0,1]$$

definita ponendo

$$F_X(x) = \mathbb{P}(X \leq x), \qquad \forall x \in \mathbb{R}.$$

□

Una funzione $F : \mathbb{R} \to \mathbb{R}$ è una funzione di ripartizione se ha le seguenti caratteristiche:

1. $0 \leq F(x) \leq 1$;
2. è una funzione non decrescente ovvero $F(x) \leq F(y)$ per ogni $x, y \in \mathbb{R}$ tali che $x < y$
3. è una funzione continua a destra ovvero $F(x_0) = \lim_{x \to x_0^+} F(x), \qquad \forall x_0 \in \mathbb{R}$.
4. $\lim_{x \to -\infty} F(x) = 0, \qquad \lim_{x \to +\infty} F(x) = 1.$

Variabili aleatorie discrete Una v.a. X si dice discreta se la sua funzione di ripartizione è discreta. Una funzione di ripartizione $F_X(\cdot)$ si dice discreta se è una funzione di ripartizione costante a tratti ovvero una funzione a gradini (*step-function*) con discontinuità (di prima specie). In tali circostanze l'insieme dei punti di salto $\mathcal{X} \subset \mathbb{R}$ è costituito da un numero finito, o al più numerabile, di punti distinti. L'altezza p_x di ciascun gradino o salto della funzione $F_X(\cdot)$ vale

$$p_x = F_X(x) - \lim_{y \to x^-} F_X(y) > 0$$

e sarà $\sum_{x \in \mathcal{X}} p_x = 1$. L'insieme $\mathcal{X}$ è detto *supporto* della distribuzione discreta.

1.63 Definizione La *funzione di massa di probabilità* di una v.a. discreta X è una funzione

$$f_X : \mathbb{R} \to [0,1]$$

definita ponendo

$$f_X(x) = \mathbb{P}(X = x) = \begin{cases} p_x & \forall x \in \mathcal{X}, \\ 0 & \forall x \notin \mathcal{X} \end{cases}$$

□

Una funzione $f : \mathbb{R} \to [0,1]$ è *funzione di massa di probabilità* se esiste un sottoinsieme $\mathcal{X}$ di cardinalità finita o al più numerabile, detto supporto, tale che

1. $f(x) > 0 \quad \forall x \in \mathcal{X}$;
2. $f(x) = 0 \quad \forall x \notin \mathcal{X}$;
3. $\sum_{x \in \mathcal{X}} f(x) = 1$.

Variabili aleatorie assolutamente continue Una v.a. X si dice assolutamente continua se è assolutamente continua la sua funzione di ripartizione $F_X(\cdot)$.

1.64 Definizione Una funzione di ripartizione di una v.a. X si dice *assolutamente continua* (a.c.) se esiste una funzione reale di variabile reale $f_X(\cdot)$, integrabile su tutti gli intervalli, tale che:

1. $f_X(x) \geq 0$;
2. $F_X(x) = \int_{-\infty}^{x} f_X(t)dt, \qquad \forall x \in \mathbb{R}.$

□

Se X è una v.a. assolutamente continua allora, per ogni coppia $a, b \in \mathbb{R}$ tali che $a \leq b$, si ha che:

$$\mathbb{P}(a < X \leq b) = \int_a^b f_X(x)dx = F_X(b) - F_X(a).$$

La funzione $f_X(\cdot)$ è detta *funzione di densità di probabilità* di X. Infine, se X è una v.a. assolutamente continua si ha (quasi ovunque) che

$$f_X(x) = \frac{d}{dx}F_X(x).$$

Il supporto della distribuzione di una v.a. è definito come il più piccolo tra gli insiemi chiusi C tali che $\mathbb{P}(X \in C) = 1$. Nel caso di v.a. assolutamente continue può essere identificato come l'insieme $\mathcal{X}$ tale che $f_X(x) > 0$ per ogni $x \in \mathcal{X}$.
Una funzione $f : \mathbb{R} \to \mathbb{R}^+$ è una *funzione di densità di probabilità* se soddisfa le seguenti condizioni:

1. $f(x) \geq 0$;
2. $f(\cdot)$ è integrabile su tutti gli intervalli;
3. $\int_{-\infty}^{+\infty} f(x)dx = 1$.

1.65 Definizione Un *quantile di livello* $\alpha \in [0, 1]$ di una v.a. X (assolutamente continua) è un valore $q_\alpha \in \mathbb{R}$ tale che

$$\mathbb{P}(X \leq q_\alpha) = F_X(q_\alpha) = \int_{-\infty}^{q_\alpha} f_X(x)dx = \alpha.$$

□

Nel caso in cui la funzione F_X sia invertibile (ovvero biiettiva), a ciascun valore di α corrisponde un unico valore di q_α e viceversa. In tal caso la funzione $Q : [0, 1] \to \mathcal{X}$ definita da

$$Q(\alpha) = F_X^{-1}(\alpha) = q_\alpha$$

si chiama *funzione quantile*.
La Figura (1.7) riporta il grafico della funzione di densità della v.a. normale standardizzata e del quantile di livello α. Il quantile q_α è un valore che può assumere la v.a. X e va quindi rappresentato sull'asse delle ascisse, mentre α è la probabilità che X assuma valori nella semiretta $(-\infty, q_\alpha)$ e deve quindi essere rappresentatato come area della porzione di piano con base la semiretta e delimitata superiormente dal grafico della funzione di densità di X.

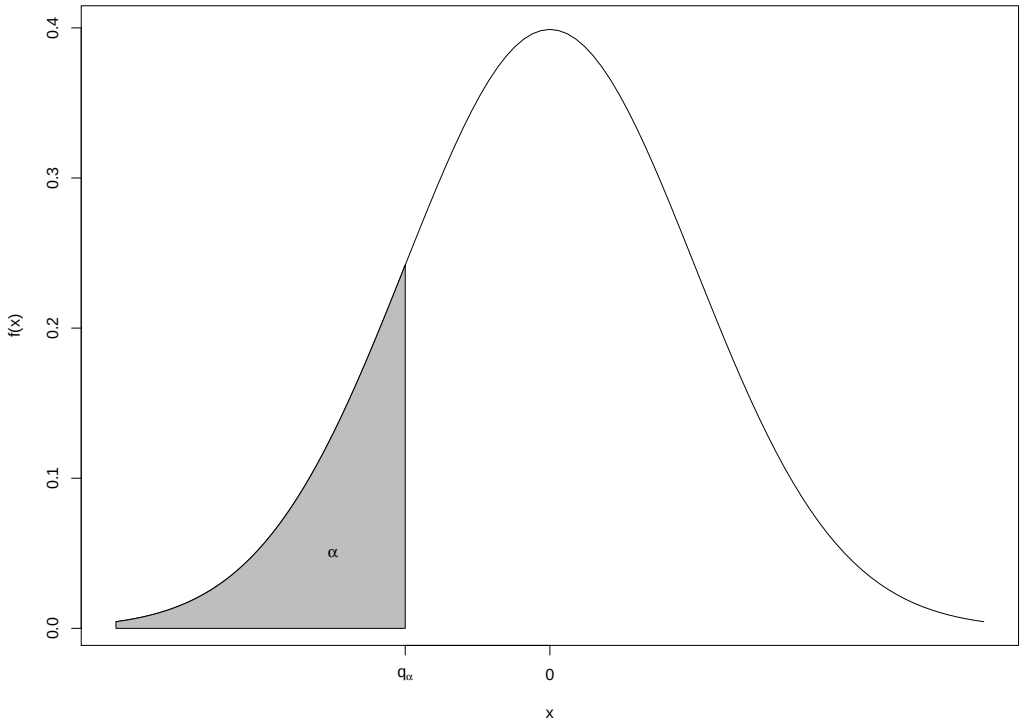

Figura 1.7: *Quantile di livello α della v.a. X.*

Capitolo 2

La funzione di verosimiglianza e i suoi usi inferenziali

Il capitolo è dedicato all'illustrazione dei principi e dei metodi dell'impostazione inferenziale basata sulla funzione di verosimiglianza (fdv). Dopo avere introdotto il concetto di fdv, viene mostrato come utilizzarla per risolvere i tre canonici problemi ipotetici (stima puntuale, stima intervallare e verifica di ipotesi). Nella seconda parte del capitolo si introduce un altro concetto fondamentale nella teoria dell'inferenza statistica: la sufficienza. A conclusione del capitolo viene discusso il cosiddetto Principio di Verosimiglianza, il fondamento logico dell'impiego dei metodi basati sulla funzione omonima.

2.1 La funzione di verosimiglianza

Lo scopo principale di un esperimento statistico è quello di fornire informazioni sul meccanismo generatore dei dati, in parte non noto. Nei problemi parametrici, l'incertezza riguarda il parametro incognito θ del modello statistico $(\mathcal{X}^n, f_n(\mathbf{x}_n;\theta), \theta \in \Theta)$. La fdv svolge esattamente questa funzione: formalizza matematicamente l'informazione che i dati osservati danno sul parametro incognito di un modello. Infatti, in corrispondenza di un campione osservato, $\mathbf{x}_n^o$, la funzione di verosimiglianza assegna ad ogni possibile valore di $\theta \in \Theta$ una misura di plausibilità (verosimiglianza) basata sui dati. Si tratta, come vedremo, di una quantificazione della compatibilità tra il campione che si è osservato e ciascuno dei possibili valori che θ può assumere. Prima di considerare la definizione formale, illustriamo un esempio.

2.1 Esempio (Modello bernoulliano). Supponiamo di non conoscere la probabilità θ che al lancio di una determinata moneta esca T. Supponiamo però di sapere che tale valore possa essere esclusivamente $\theta = 0.1$ oppure $\theta = 0.9$. Si lanci 10 volte la moneta e si osservi 8 volte T e 2 volte C. Sulla base di questo risultato è intuitivo ritenere che il valore $\theta = 0.9$ sia, tra i due, il più plausibile, ovvero il più compatibile (o *verosimile*) con il risultato dell'esperimento. La giustificazione di questa intuizione è che, se il vero valore di θ fosse proprio 0.9, sarebbe più probabile osservare 8 volte T su 10 lanci della moneta, di quanto non sarebbe se θ fosse pari a 0.1. Pensando al risultato del lancio della moneta come alla realizzazione di una v.a. bernoulliana di parametro θ in cui l'esito T corrisponde al valore 1 e l'esito C al valore 0, e supponendo che il campione osservato dia luogo a 8 successi e 2 insuccessi, la probabilità del campione osservato è $= f_{10}(\mathbf{x}_{10}^o;\theta) = \theta^{\sum x_i^o}(1-\theta)^{10-\sum x_i^o} = \theta^8(1-\theta)^2$. Se $\theta = 0.9$, allora $f_{10}(\mathbf{x}_{10}^o;\theta = 0.9) = (0.9)^8\ (0.1)^2 = 0.004304672$. Questo numero è la verosimiglianza di $\theta = 0.9$ associata al campione osservato. Se invece $\theta = 0.1$, la probabilità

dello stesso campione è $f_{10}(\mathbf{x}_{10}^o; \theta = 0.1) = 0.0000000081$. Questo valore è la verosimiglianza di $\theta = 0.1$ associata al campione osservato. Facendo il rapporto delle due probabilità ottenute si trova che lo stesso campione è 531441 volte più probabile quando $\theta = 0.9$ di quando $\theta = 0.1$. Quanto detto per i due valori $\theta = 0.1$ e $\theta = 0.9$ può essere esteso a tutti i valori che θ può assumere, ovvero ai valori nell'intervallo reale $[0, 1]$. Si ottiene così la funzione di verosimiglianza del parametro θ associata al campione osservato, la cui espressione è $\theta^8(1-\theta)^2$, $\theta \in [0, 1]$. Il grafico della fdv descrive, al variare di θ, come varia la probabilità di osservare il campione che si è realizzato. I valori di θ in cui la fdv assume valori più elevati sono quelli più verosimili. Si verifica facilmente che valore $\theta = 8/10$ è il valore maggiormente verosimile alla luce del campione osservato (punto di massimo della funzione nell'intervallo $[0, 1]$). □

2.2 Definizione (Funzione di verosimiglianza). Dato un modello statistico $(\mathcal{X}^n, f_n(\mathbf{x}_n; \theta), \theta \in \Theta)$ si chiama *funzione di verosimiglianza* di θ associata al campione osservato $\mathbf{x}_n^o = (x_1^o, \ldots, x_n^o)$, la funzione $L : \Theta \mapsto \mathbb{R}^+$ definita ponendo

$$L(\theta; \mathbf{x}_n^o) = f_n(\mathbf{x}_n^o; \theta), \qquad \theta \in \Theta.$$

□

La funzione di verosimiglianza di θ si ottiene quindi calcolando il valore che $f_n(\cdot; \theta)$ assume nel particolare campione osservato $\mathbf{x}_n^o$. Poichè i valori $x_1^o, \ldots, x_n^o$ sono dei numeri reali, $f_n(\mathbf{x}_n^o; \theta)$ è una funzione di θ. Per semplicità di notazione, nel seguito indicheremo il campione osservato con $\mathbf{x}_n$, ovvero con la stessa simbologia utilizzata per un generico campione di $\mathcal{X}^n$.
La definizione introdotta è valida sia nel caso in cui si considerano v.a. discrete che in quelli per v.a. assolutamente continue. Se le v.a. X_i sono discrete, la fdv fornisce, per ogni valore di θ, la probabilità del campione realizzato e quindi assume valori nell'intervallo $[0, 1]$. Se le v.a. X_i sono assolutamente continue, la fdv fornisce, per ogni valore di θ, la densità di probabilità del campione realizzato, $\mathbf{x}_n$. L'interpretazione resta comunque inalterata rispetto al caso discreto: i valori di θ per i quali L è più elevata sono i valori del parametro più compatibili con il campione dato. Per questo motivo i valori che $L(\theta; \mathbf{x}_n)$ assegna ai diversi valori di θ sono detti *verosimiglianze* e, complessivamente, la fdv quantifica l'*evidenza sperimentale* che il campione osservato fornisce per ciascun valore di θ in Θ.

2.3 Osservazione.

1. I valori della funzione di verosimiglianza non possono essere interpretati come probabilità assegnate a θ, alla luce del risultato $\mathbf{x}_n$. Il parametro θ non è infatti una v.a. e il valore $L(\theta; \mathbf{x}_n)$ rappresenta, come si è detto, la probabilità o la densità di probabilità dell'evento $\mathbf{X}_n = \mathbf{x}_n$.

2. La fdv è definita a meno di costanti che non dipendono da θ. Per una costante arbitraria $c > 0$, eventualmente dipendente dai dati campionari, la funzione $cL(\theta; \mathbf{x}_n)$ fornisce sul parametro incognito le stesse informazioni che otteniamo considerando la funzione $L(\theta; \mathbf{x}_n)$. In generale, possiamo sempre esprimere $f_n(\mathbf{x}_n; \theta)$ e quindi anche la verosimiglianza $L(\theta; \mathbf{x}_n)$ attraverso il prodotto di due funzioni h (di $\mathbf{x}_n$) e g (di $\mathbf{x}_n$ e θ) tali che $\forall \mathbf{x}_n \in \mathcal{X}^n$ e $\forall \theta \in \Theta$ si ha

   $$L(\theta; \mathbf{x}_n) = f_n(\mathbf{x}_n; \theta) = h(\mathbf{x}_n) g(\theta; \mathbf{x}_n).$$

   La funzione $g(\theta; \mathbf{x}_n)$ rappresenta il *nucleo* della funzione di verosimiglianza di θ. Si osservi inoltre che, mentre il fattore $h(\mathbf{x}_n)$ è indispensabile per la definizione corretta di $f_n(\mathbf{x}_n; \theta)$, ovvero per la distribuzione campionaria di $\mathbf{X}_n$, tale quantità risulta irrilevante (e può essere trascurata) per la funzione $L(\theta; \mathbf{x}_n)$, dal momento che non dipende da θ.

3. (Funzione di verosimiglianza associata a campioni casuali). La definizione di fdv vale per campioni generici (non necessariamente i.i.d.). Si ottiene una notevole semplificazione analitica quando $\mathbf{X}_n$ è un campione casuale. In questo caso, infatti, si ha che
$$L(\theta; \mathbf{x}_n) = \prod_{i=1}^{n} f_X(x_i; \theta).$$

□

La fdv assume, per definizione, valori nello spazio $\mathbb{R}^+$. La definizione che segue introduce una una misura relativa (compresa tra 0 e 1) dell'evidenza che il campione osservati di dati $\mathbf{x}_n$ assegna ai valori di θ in Θ.

2.4 Definizione (Funzione di verosimiglianza relativa). Dato un modello statistico $(\mathcal{X}^n, f_n(\mathbf{x}_n; \theta), \theta \in \Theta)$, si definisce funzione di verosimiglianza relativa associata al campione osservato $\mathbf{x}_n$ la funzione $\overline{L} : \Theta \mapsto [0, 1]$ definita ponendo
$$\overline{L}(\theta; \mathbf{x}_n) = \frac{L(\theta; \mathbf{x}_n)}{\sup_{\theta \in \Theta} L(\theta; \mathbf{x}_n)}, \qquad \theta \in \Theta. \tag{2.1}$$

□

La funzione di verosimiglianza relativa $\overline{L}(\theta; \mathbf{x}_n)$ è proporzionale a $L(\theta; \mathbf{x}_n)$, dal momento che $\sup_{\theta \in \Theta} L(\theta; \mathbf{x}_n)$ è una costante rispetto a θ. Nei problemi più semplici $L(\theta; \mathbf{x}_n)$ ha un unico punto di massimo (che, vedremo, si chiama *stima di massima verosimiglianza*) $\theta = \widehat{\theta}_{mv}$ appartenente a Θ. Si ha quindi che
$$\overline{L}(\theta; \mathbf{x}_n) = \frac{L(\theta; \mathbf{x}_n)}{\max_{\theta \in \Theta} L(\theta; \mathbf{x}_n)} = \frac{L(\theta; \mathbf{x}_n)}{L(\widehat{\theta}_{mv}; \mathbf{x}_n)}, \qquad \theta \in \Theta.$$

La fdv relativa consente di confrontare direttamente i diversi possibili valori di θ tra loro e con l'ipotesi $\theta = \widehat{\theta}_{mv}$ privilegiata dal campione osservato.

2.5 Esempio (Modello bernoulliano). Riprendiamo in esame il modello bernoulliano e consideriamo un generico campione osservato $\mathbf{x}_n$ di dimensione n. In questo caso
$$L(\theta; \mathbf{x}_n) = f_n(\mathbf{x}_n; \theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}.$$

Se consideriamo il campione di dimensione $n = 10$ in cui $\sum_{i=1}^n x_i = 8$ e $(n - \sum_{i=1}^n x_i) = 2$, la fdv di θ risulta essere
$$L(\theta; \mathbf{x}_n) = \theta^8(1-\theta)^2, \qquad \theta \in [0, 1].$$

Con semplici calcoli (vedi Es. 2.9 per la verifica) si mostra che l'unico punto di massimo assoluto di questa funzionesi ottiene per $\theta = 0.8$. La fdv relativa è quindi
$$\overline{L}(\theta; \mathbf{x}_n) = \left(\frac{\theta}{0.8}\right)^8 \left(\frac{1-\theta}{0.2}\right)^2, \qquad \theta \in [0, 1].$$

La Figura 2.1 mostra il grafico di $\overline{L}(\theta; \mathbf{x}_n)$ per $\theta \in [0, 1]$ (linea continua). Consideriamo ora un nuovo esperimento in cui $n = 50$ e il numero di successi osservati nel campione è pari a 40. La fdv relativa è quindi
$$\overline{L}(\theta; \mathbf{x}_n) = \left(\frac{\theta}{0.8}\right)^{40} \left(\frac{1-\theta}{0.2}\right)^{10}, \qquad \theta \in [0, 1].$$

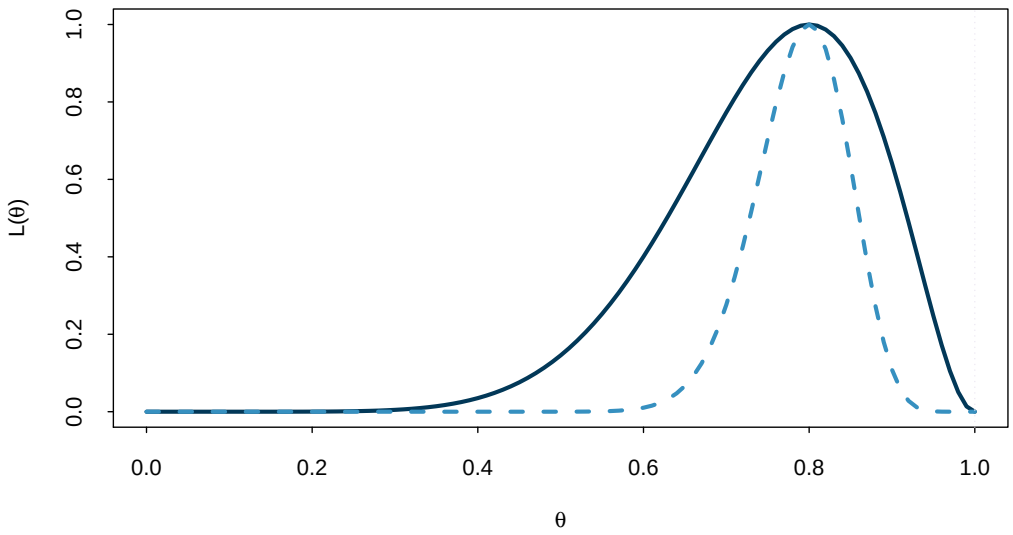

Figura 2.1: *Grafico della funzione di verosimiglianza relativa di* θ, $\overline{L}$, *associata a due campioni provenienti da una popolazione bernoulliana di numerosità* $n = 10$ *(linea continua) e* $n = 50$ *(linea tratteggiata), entrambi con* $\overline{x}_n = 0.8$.

Si verifica facilmente che anche in questo caso $\widehat{\theta}_{mv} = 0.8$. La linea tratteggiata nella Figura 2.1 corrisponde al grafico di $\overline{L}(\theta; \mathbf{x}_n)$ in questo secondo caso. Va osservato che, nonostante il punto di massimo resti inalterato, quando $n = 50$ la fdv relativa risulta ben più concentrata intorno al suo punto di massimo che nel caso $n = 10$. All'aumentare del numero di osservazioni, si riduce la lunghezza dell'intervallo di valori del parametro che ricevono dal campione una misura di evidenza superiore a un livello prescelto. L'aumento della dimensione campionaria consente quindi alla fdv di discernere meglio tra i possibili valori di θ. □

2.2 Usi inferenziali della funzione di verosimiglianza

La funzione di verosimiglianza costituisce il punto di partenza per affrontare i tre problemi inferenziali riguardanti il parametro incognito del modello (stima puntuale, stima mediante regioni e verifica di ipotesi). A tale fine si considerano opportune sintesi della fdv che, nel suo insieme, rappresenta la totalità dell'informazione sperimentale sul parametro incognito del modello.

2.2.1 Stima puntuale: stima di massima verosimiglianza

Premettiamo una definizione.

2.6 Definizione (Stima di un parametro). Dato il modello statistico $\{\mathcal{X}^n,\ f_n(\mathbf{x}_n;\theta),\ \theta \in \Theta\}$ e il campione osservato $\mathbf{x}_n$, si chiama *stima* di θ una funzione $\widehat{\theta} : \mathcal{X}^n \to \Theta$. Il valore (scalare o vettoriale) $\widehat{\theta}(\mathbf{x}_n)$ viene utilizzato come valutazione numerica di θ. □

Si osservi che la definizione data è piuttosto generica e vale sia nel caso scalare che vettoriale: la dimensione della stima coincide con quella del vettore dei parametri. In molti casi è abbastanza semplice proporre delle ragionevoli funzioni dei dati campionari per la stima puntuale. Ad esempio, nel modello bernoulliano una stima intuitiva della probabilità θ è data dalla media aritmetica dei dati campionari, $\overline{x}_n$. Non sempre è però semplice individuare una funzione dei dati che sia una stima ragionevole del parametro. Abbiamo quindi bisogno di un metodo generale di stima, che non si basi solo sull'interpretazione che siamo in grado di dare al parametro di uno specifico modello. Il metodo che qui illustriamo sfrutta l'informazione che ci fornisce la fdv. Il modo più naturale per ottenere da questa funzione una stima del parametro consiste nell'individuare i valori più verosimili di θ alla luce del campione osservato, ovvero il punto o i punti di massimo assoluto di $L(\theta;\mathbf{x}_n)$ (se esistenti).

2.7 Definizione (Stima di massima verosimiglianza). Dato il modello statistico $\{\mathcal{X}^n,\ f_n(\mathbf{x}_n;\theta),\ \theta \in \Theta\}$ e il campione osservato $\mathbf{x}_n$, si chiama *stima di massima verosimiglianza* (smv) di θ un valore $\widehat{\theta}_{mv}(\mathbf{x}_n) \in \Theta$ tale che $L(\widehat{\theta}_{mv};\mathbf{x}_n) \geq L(\theta;\mathbf{x}_n),\ \forall\theta \in \Theta$:

$$\widehat{\theta}_{mv}(\mathbf{x}_n) = \arg\max_{\theta\in\Theta} L(\theta;\mathbf{x}_n).$$

□

L'esistenza della smv non è garantita per tutti i modelli statistici. Inoltre esistono modelli in cui la smv esiste ma non è unica. Tuttavia, nei più comuni casi uniparametrici, la smv esiste ed è unica.

Consideriamo ora il problema della ricerca della smv per due classi di problemi rilevanti (problemi regolari di stima e problemi con supporto della v.a. dipendente dal parametro incognito).

Stima di massima verosimiglianza in problemi regolari di stima

Per semplicità consideriamo in questa sezione il caso uniparametrico e assumiamo che Θ sia un sottoinsieme di $\mathbb{R}$. Quando il modello statistico soddisfa le proprietà della definizione che segue, la ricerca della smv può essere effettuata con le usuali tecniche con cui si determinano i punti di massimo di una funzione reale di variabile reale derivabile nel dominio.

2.8 Definizione (Problema regolare di stima). Dati una v.a. X di base, il modello statistico $\{\mathcal{X}^n,\ f_n(\mathbf{x}_n;\theta),\ \theta \in \Theta\}$, un campione osservato $\mathbf{x}_n$ e la funzione di verosimiglianza ad esso associata, $L(\theta;\mathbf{x}_n)$, si ha un *problema regolare di stima* se:

- il supporto della v.a. di base X non dipende da θ.
- esistono finite le derivate prima e seconda di $L(\theta;\mathbf{x}_n)$ rispetto a θ.

□

In questi casi un punto $\widehat{\theta}_{mv}$ interno di Θ è un massimo per la funzione $L(\theta;\mathbf{x}_n)$ se sono verificate le seguenti due condizioni:

$$\left.\frac{d}{d\theta}\ L(\theta;\mathbf{x}_n)\right|_{\theta=\widehat{\theta}_{mv}} = 0 \qquad \text{e} \qquad \left.\frac{d^2}{d\theta^2}\ L(\theta;\mathbf{x}_n)\right|_{\theta=\widehat{\theta}_{mv}} < 0. \tag{2.2}$$

In molti degli esempi che tratteremo, è più semplice calcolare le derivate della *funzione di log-verosimiglianza*:

$$\ell(\theta; \mathbf{x}_n) = \log L(\theta; \mathbf{x}_n).$$

Dal momento che la funzione logaritmo (con base maggiore di uno) è una funzione crescente, i punti di massimo di $L(\theta; \mathbf{x}_n)$ e $\ell(\theta; \mathbf{x}_n)$ coincidono. Pertanto, il punto $\widehat{\theta}_{mv}$ è un massimo per la funzione $L(\theta; \mathbf{x}_n)$ se sono verificate le seguenti due condizioni:

$$\frac{d}{d\theta}\ell(\theta; \mathbf{x}_n)\bigg|_{\theta=\widehat{\theta}_{mv}} = 0 \qquad \text{e} \qquad \frac{d^2}{d\theta^2}\ell(\theta; \mathbf{x}_n)\bigg|_{\theta=\widehat{\theta}_{mv}} < 0.$$

Le equazioni $d/d\theta L(\theta; \mathbf{x}_n) = 0$ e $d/d\theta \ell(\theta; \mathbf{x}_n) = 0$ prendono rispettivamente il nome di *equazione di verosimiglianza* e *equazione di log-verosimiglianza*.
Nell'esempio che segue illustriamo la determinazione della smv nel caso del modello bernoulliano. Il Paragrafo 2.4 riporta i calcoli espliciti per la determinazione dells smv per i principali modelli uniparametrici.

2.9 Esempio (Smv per modello bernoulliano). Consideriamo ancora una volta il modello bernoulliano e assumiamo che lo spazio Θ sia l'intervallo chiuso $[0, 1]$. La funzione di log-verosimiglianaza è

$$\ell(\theta; \mathbf{x}_n) = \log L(\theta; \mathbf{x}_n) = \left(\sum_{i=1}^{n} x_i\right) \log\theta + \left(n - \sum_{i=1}^{n} x_i\right) \log(1-\theta).$$

L'equazione di log-verosimiglianza

$$\frac{d}{d\theta}\ell(\theta; \mathbf{x}_n) = \frac{d}{d\theta}\log L(\theta; \mathbf{x}_n) = \frac{\sum x_i}{\theta} - \frac{n - \sum x_i}{1-\theta} = 0$$

ha un'unica soluzione nel punto $\widehat{\theta}_{mv} = \overline{x}_n = \sum x_i/n$. Si tratta della smv in quanto la derivata seconda di $\ell(\theta; \mathbf{x}_n)$ calcolata in $\theta = \overline{x}_n$ è

$$\frac{d^2}{d\theta^2}\ell(\theta; \mathbf{x}_n)\bigg|_{\theta=\overline{x}_n} = \quad - \frac{n}{\overline{x}_n(1-\overline{x}_n)} < 0.$$

□

2.10 Esempio (Smv per altri modelli notevoli). Si procede in modo analogo a quanto visto nel precedente esempio 2.9 per i principali modelli uniparametrici regolari considerati. In particolare: (i) per il modello di Poisson (vedi Es. 2.20), $\widehat{\theta}_{mv} = \overline{x}_n$; per il modello Bin$(k, n)$ (vedi Es. 2.22), $\widehat{\theta}_{mv} = \overline{x}_n/k$;

□

Stima di massima verosimiglianza nei modelli con supporto dipendente da θ

Quando il supporto della v.a. dipende dal parametro del modello, il problema di stima è non regolare e non si può ricorrere al metodo appena illustrato. In questi casi, infatti, la fdv presenta punti di discontinuità (in cui la fdv non è derivabile) che sono proprio i punti di massima verosimiglianza.

2.11 Esempio (**Modello uniforme**). Consideriamo separatamente due casi: campione casuale da v.a. uniforme in $[0, \theta]$ e da v.a. uniforme in $[\theta - 1, \theta + 1]$.

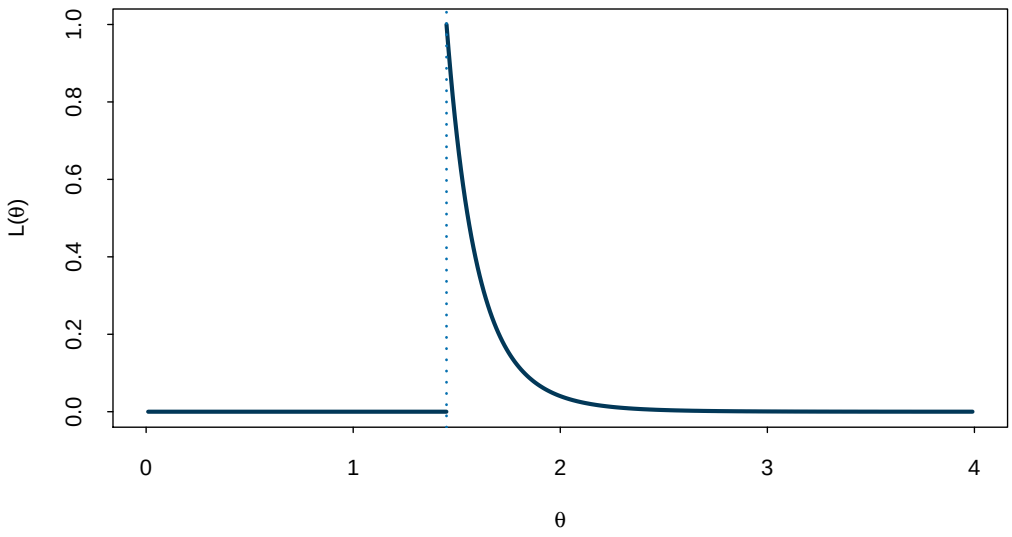

Figura 2.2: *Grafico delle funzione di verosimiglianza relativa di* θ, $\overline{L}$, *per un campione di dimensione estratto da una popolazione Unif*$[0, \theta]$, *con* $n = 10$, $x_{(10)=1.45}$.

A. La funzione di densità della variabile aleatoria di base $X \sim \text{Unif}[0, \theta]$ è $f_X(x; \theta) = \theta^{-1} I_{[0,\theta]}(x)$. La funzione di verosimiglianza per il campione osservato $\mathbf{x}_n$ è

$$L(\theta; \mathbf{x}_n) = f_n(\mathbf{x}_n; \theta) = \frac{1}{\theta^n} \prod_{i=1}^{n} I_{[0,\theta]}\ (x_i) = \frac{1}{\theta^n}\ I_{[x_{(n)},\infty]}(\theta),$$

dove $x_{(n)} = \max\{x_1, \ldots, x_n\}$ indica il massimo valore tra le osservazioni campionarie osservate. L'ultima uguaglianza si giustifica osservando che

$$\prod_{i=1}^{n} I_{[0,\theta]}\ (x_i) = 1 \Leftrightarrow x_i \in [0, \theta],\ i = 1, \ldots, n\ \Leftrightarrow 0 \leq x_{(n)} \leq \theta \Leftrightarrow x_{(n)} \leq \theta \leq +\infty.$$

La funzione di verosimiglianza vale quindi zero nell'intervallo $[0, x_{(n)}]$, assume valori positivi decrescenti in $[x_{(n)}, +\infty)$ e ha quindi un unico punto di massimo in $\theta = x_{(n)}$, punto di discontinuità di prima specie per $L(\theta; \mathbf{x}_n)$ (ove la funzione non è derivabile e vale $[x_{(n)}]^{-n}$). Abbiamo quindi che $\widehat{\theta}_{mv} = x_{(n)}$. La Figura 2.2 mostra il grafico di $\overline{L}(\theta; \mathbf{x}_n)$ nel caso di un campione osservato in cui $n = 10$ e $x_{(10)} = 1.45$.

B. Consideriamo ora il modello uniforme nell'intervallo $[\theta - 1, \theta + 1]$. La funzione di densità della variabile aleatoria di base $X \sim \text{Unif}[\theta - 1, \theta + 1]$ è:

$$f_X(x; \theta) = \frac{1}{2}\ I_{[\theta-1,\theta+1]}(x).$$

La fdv associata al campione osservato $\mathbf{x}_n$ è:

$$L(\theta; \mathbf{x}_n) = f_n(\mathbf{x}_n; \theta) = \frac{1}{2^n} \prod_{i=1}^{n} I_{[\theta-1, \theta+1]}(x_i) = \frac{1}{2^n}\ I_{[x_{(n)}-1, x_{(1)}+1]}(\theta)$$

dove $x_{(1)} = \min\{x_1, \ldots, x_n\}$, $x_{(n)} = \max\{x_1, \ldots, x_n\}$ sono rispettivamente il minimo e il massimo campionario. La funzione di verosimiglianza è costante sull'intervallo $[x_{(n)} - 1, x_{(1)} + 1]$ e vale 0 al di fuori di tale intervallo. Pertanto, tutti i punti di questo intervallo sono punti di massimo per $L(\theta; \mathbf{x}_n)$ e la stima di massima verosimiglianza non è unica. □

2.2.2 Stima mediante regioni: insiemi di verosimiglianza

La procedura di stima di massima verosimiglianza seleziona il valore o i valori di θ, se esistenti, massimamente compatibili con il campione osservato. Sembra tuttavia ragionevole non scartare i valori di θ che ricevono dal campione osservato un supporto superiore ad una soglia adeguatamente elevata.

2.12 Definizione (Insieme di verosimiglianza). Dato il modello statistico $\{\mathcal{X}^n,\ f_n(\mathbf{x}_n; \theta), \theta \in \Theta\}$, il campione osservato $\mathbf{x}_n$ e un valore $q \in [0, 1]$, si definisce *insieme di verosimiglianza di livello q* l'insieme di valori di θ

$$L_q(\mathbf{x}_n) = \{\theta \in \Theta : \overline{L}(\theta; \mathbf{x}_n) \geq q\}.$$

□

La definizione data, valida per uno spazio parametrico di dimensione arbitraria, non garantisce che, nel caso in cui $\Theta \subseteq \mathbb{R}$, l'insieme L_q sia un intervallo, anche se questo è il caso di molti dei più comuni modelli che consideriamo in questo testo. Si noti inoltre che, quanto più elevato è il valore di q, tanto minore risulta la dimensione di L_q. Al contrario, quanto minore q, tanto maggiore la dimensione di L_q. Come casi limite abbiamo che

$$q \to 0 \Rightarrow L_q(\mathbf{x}_n) \to \Theta \qquad \text{e che} \qquad q \to 1 \Rightarrow L_q(\mathbf{x}_n) \to \widehat{\theta}_{mv}.$$

Consideriamo un esempio in cui è possibile determinare esplicitamente gli estremi dell'intervallo L_q per un parametro scalare.

2.13 Esempio (Modello normale, varianza nota). Consideriamo per la v.a. di base un modello normale con valore atteso incognito θ e varianza nota, pari a σ^2. Per questo modello gli estremi dell'intervallo di log-verosimiglianza di livello q possono essere determinati analiticamente. La fdv associata ad un campione casuale osservato è

$$L(\theta; \mathbf{x}_n) = f_n(\mathbf{x}_n; \theta) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n\ e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2} \propto e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2}.$$

La stima di massima verosimiglianza per questo modello (vedi Esempio 2.27) è $\widehat{\theta}_{mv} = \overline{x}_n$, la media campionaria di $\mathbf{x}_n$. La funzione di verosimiglianza relativa è:

$$\overline{L}(\theta; \mathbf{x}_n) = \frac{L(\theta, \mathbf{x}_n)}{L(\widehat{\theta}_{mv}, \mathbf{x}_n)} = \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(x_i - \theta)^2 - \sum_{i=1}^{n}(x_i - \overline{x}_n)^2\right]\right\}.$$

La quantità $\sum(x_i-\theta)^2$ può essere scomposta nel modo seguente:

$$\begin{aligned}\sum_{i=1}^{n}(x_i-\theta)^2 &= \sum_{i=1}^{n}(x_i-\overline{x}_n+\overline{x}_n-\theta)^2=\sum_{i=1}^{n}\left[(x_i-\overline{x}_n)^2+(\overline{x}_n-\theta)^2+2(x_i-\overline{x}_n)(\overline{x}_n-\theta)\right]\\ &= \sum_{i=1}^{n}(x_i-\overline{x}_n)^2+n(\overline{x}_n-\theta)^2+2(\overline{x}_n-\theta)\sum_{i=1}^{n}(x_i-\overline{x}_n)\\ &= \sum_{i=1}^{n}(x_i-\overline{x}_n)^2+n(\overline{x}_n-\theta)^2,\end{aligned}$$

dove l'ultima uguaglianza sussiste ricordando che $\sum(x_i-\overline{x}_n)=0$. La funzione di verosimiglianza relativa diventa

$$\begin{aligned}\overline{L}(\theta;\mathbf{x}_n) &= \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(x_i-\overline{x}_n)^2+n(\overline{x}_n-\theta)^2-\sum_{i=1}^{n}(x_i-\overline{x}_n)^2\right]\right\}\\ &= \exp\left\{-\frac{n}{2\sigma^2}(\overline{x}_n-\theta)^2\right\}.\end{aligned}$$

L'intervallo di verosimiglianza per θ di livello q è dato dall'insieme di valori di θ tali che:

$$\begin{aligned}L_q &= \{\theta\in\Theta:\overline{L}(\theta;\mathbf{x}_n)\geq q\}=\left\{\theta\in\Theta:\exp\{-\frac{n}{2\sigma^2}(\overline{x}_n-\theta)^2\}\geq q\right\}\\ &= \left\{\theta\in\Theta:-\frac{n}{2\sigma^2}(\overline{x}_n-\theta)^2\geq\ln q\right\}\\ &= \left\{\theta\in\Theta:\overline{x}_n-k_q\frac{\sigma}{\sqrt{n}}\ \leq\ \theta\ \leq\ \overline{x}_n+k_q\frac{\sigma}{\sqrt{n}}\right\}\\ &= \left[\overline{x}_n-k_q\frac{\sigma}{\sqrt{n}},\overline{x}_n+k_q\frac{\sigma}{\sqrt{n}}\right],\qquad k_q=\sqrt{-2\ln q}.\end{aligned}$$

Si noti che l'ipotesi $q\in(0,1)$ garantisce l'esistenza di k_q. La Figura 2.3 riporta il grafico della funzione di verosimiglianza relativa associata a due campioni di diversa numerosità ma con stessa media aritmetica ($\overline{x}_n=2$) (la linea tratteggiata corrisponde alla fdv associata al campione di dimensione maggiore). Si osservi anche la simmetria delle fdv e la diversa lunghezza degli insiemi di verosimiglianza. □

Per molti modelli non è possibile risolvere analiticamente la disequazione che definisce l'insieme $L_q(\mathbf{x}_n)$ e quindi determinare le formule che definiscono gli estremi dell'intervallo di verosimiglianza. In questi casi tali valori si possono tuttavia ottenere numericamente. In alternativa, come vedremo, è anche possibile ricorrere ad approssimazioni della fdv per ottenere formule esplicite degli estremi (approssimati) degli intervalli di livello q.

2.14 Esempio (**Modello bernoulliano**). In questo caso si ha che, per un generico campione osservato $\mathbf{x}_n$,

$$\overline{L}(\theta;\mathbf{x}_n)=\left(\frac{\theta}{\overline{x}_n}\right)^{\sum x_i}\left(\frac{1-\theta}{1-\overline{x}_n}\right)^{n-\sum x_i}$$

e non è quindi possibile determinare le formule esplicite degli estremi dell'insieme L_q. Risulta tuttavia agevole determinare numericamente le soluzioni della disequazione $\overline{L}(\theta;\mathbf{x}_n)\geq q$ in

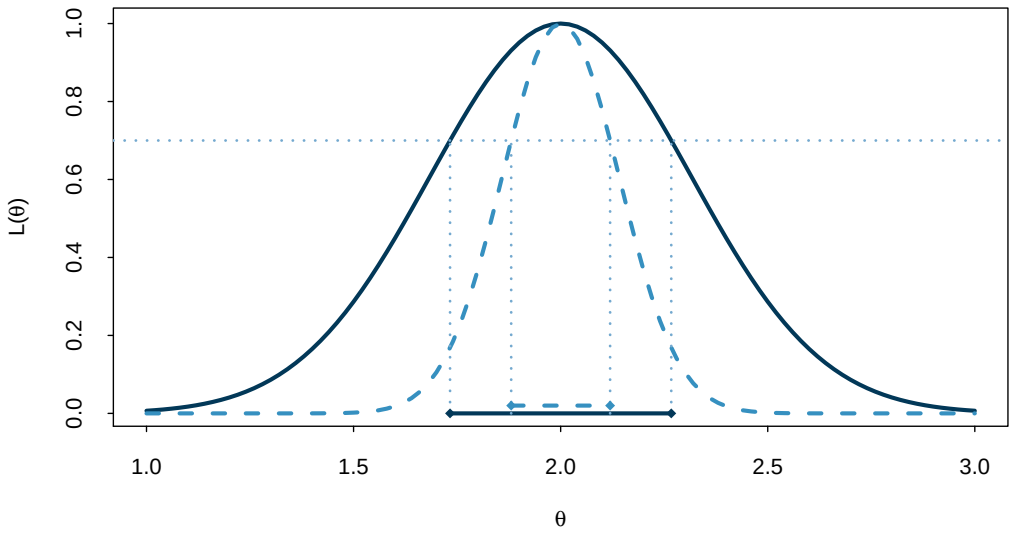

Figura 2.3: *Insiemi L_q per il parametro θ del modello $N(\theta, \sigma^2)$.*

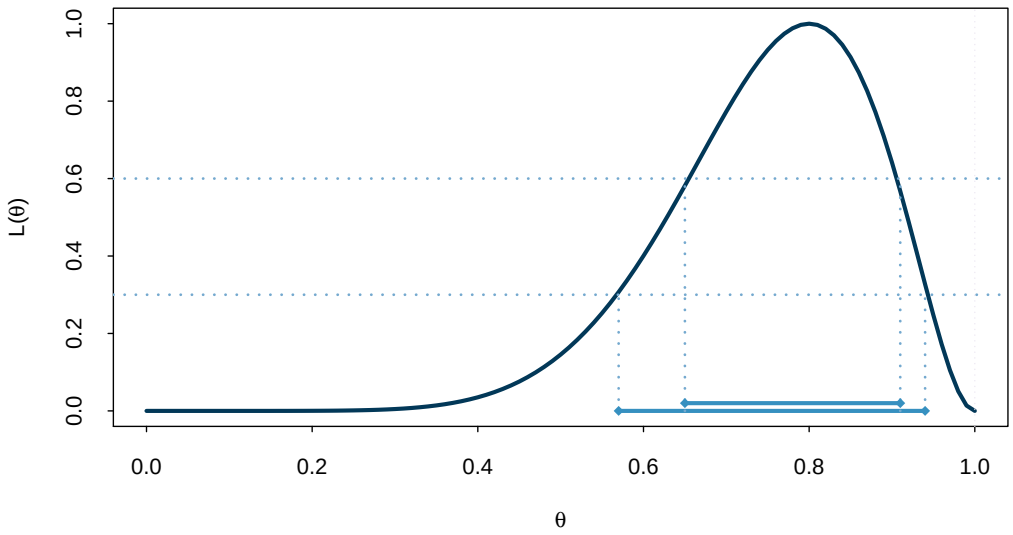

Figura 2.4: *Insiemi L_q per il modello bernoulliano, per $q = 0.3$ e $q = 0.6$.*

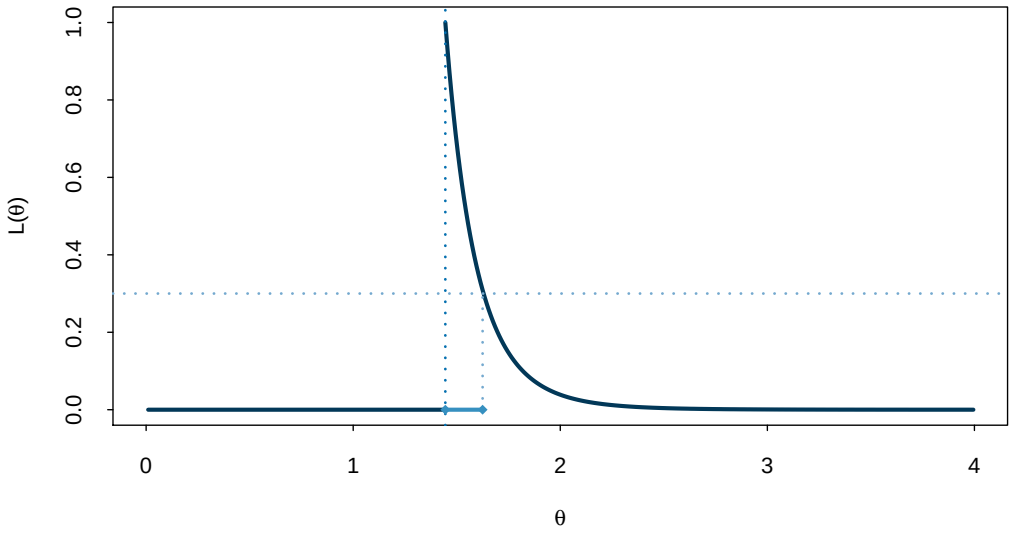

Figura 2.5: *Insieme L_q per modello Unif*$[0, \theta]$.

corrispondenza di campioni effettivi. La Figura 2.4 mostra gli insiemi L_q che si ottengono ponendo $q = 0.3$ e $q = 0.6$. Si osservi che, come avviene in generale, a un valore più grande di q corrisponde una minore lunghezza dell'intervallo di stima. □

Insiemi di verosimiglianza nei modelli con supporto dipendente da θ

La determinazione dell'intervallo L_q non presenta particolari problemi nel caso di modelli con supporto dipendente da θ.

2.15 Esempio (Modello uniforme). Nel caso di un campione casuale dalla popolazione uniforme in $[0, \theta]$ abbiamo visto che

$$L(\theta; \mathbf{x}_n) = \frac{1}{\theta^n} I_{[x_{(n)}, +\infty)}(\theta), \quad \theta > 0,$$

dove $x_{(n)} = \max\{x_1, \ldots, x_n\}$, In questo modello $\widehat{\theta}_{mv} = x_{(n)}$ e quindi

$$\overline{L}(\theta; \mathbf{x}_n) = \left[\frac{x_{(n)}}{\theta}\right]^n I_{[x_{(n)}, +\infty)}(\theta).$$

Dall'esame grafico si deduce che, per un qualunque $q \in [0, 1]$, l'insieme di verosimiglianza di livello q è $L_q(\mathbf{x}_n) = [x_{(n)}, x_{(n)} q^{-1/n}]$. La figura 2.5 mostra (in grigio) l'insieme $L_q \subseteq \Theta = [0, 1]$ nel caso in cui il massimo campionario osservato è uguale a 0.45, $n = 10$ e $q = 0.3$ e quindi assumendo ad esempio $q = 0.147$, si ottiene l'intervallo di verosimiglianza $[0.45, 0.54]$.

□

2.2.3 Verifica di ipotesi: rapporti di verosimiglianze

Nei problemi di verifica di ipotesi siamo interessati a confrontare la plausibilità di congetture alternative sul parametro θ, utilizzando i dati a disposizione. Consideriamo due sottoinsiemi dello spazio dei parametri Θ, che indichiamo con Θ_1 e Θ_2 e supponiamo che $\Theta_1 \cap \Theta_2 = \emptyset$. Indichiamo con H_i l'ipotesi che $\theta \in \Theta_i$, i=1,2, ovvero che il valore del parametro che individua il modello che ha generato i dati sia un valore in Θ_i. Un caso particolare si ha quando si assume che gli insiemi Θ_i siano costituiti da punti dello spazio parametrico, ovvero che $\Theta_1 = \{\theta_1\}$ e $\Theta_2 = \{\theta_2\}$. In questo caso il confronto tra le due ipotesi (dette *ipotesi semplici*) può essere effettuato attraverso il *rapporto delle verosimiglianze*:

$$\lambda_{12}(\mathbf{x}_n) = \frac{L(\theta_1; \mathbf{x}_n)}{L(\theta_2; \mathbf{x}_n)} = \frac{\overline{L}(\theta_1; \mathbf{x}_n)}{\overline{L}(\theta_2; \mathbf{x}_n)}$$

Evidentemente, l'ipotesi H_i è più verosimile dell'ipotesi H_j se $\lambda_{ij}(\mathbf{x}_n) > 1$ ($i, j = 1, 2$ e $i \neq j$); le ipotesi sono equivalenti se $\lambda_{12}(\mathbf{x}_n) = 1$.

Il rapporto delle verosimiglianze, oltre a indicare se una ipotesi sia più verosimile di un'altra, misura il *grado di evidenza* che i dati assegnano a una ipotesi rispetto ad un'altra. Se, ad esempio, in corrispondenza di $\mathbf{x}_n$ si ha che $\lambda_{ij}(\mathbf{x}_n) = 8$, alla luce dei dati a disposizione l'ipotesi H_j risulta 8 volte più plausible dell'ipotesi H_j.

2.16 Esempio (**Rapporti di verosimiglianza nel modello bernoulliano**). Si consideri il modello bernoulliano e un campione casuale di $n = 10$ osservazioni per le quali $\sum x_i = 8$. In questo caso

$$\overline{L}(\theta; \mathbf{x}_n) = \frac{\theta^{\sum x_i}\ (1-\theta)^{10-\sum x_i}}{(0.8)^{\sum x_i}\ (0.2)^{10-\sum x_i}} = \frac{\theta^8\ (1-\theta)^2}{(0.8)^8\ (0.2)^2}.$$

Se si deve decidere quale tra le due ipotesi semplici $H_1 : \ \theta = \theta_1 = 0.3$ contro l'ipotesi $H_2 : \ \theta = \theta_2 = 0.7$ è più verosimile, si calcola $\overline{L}(0.3; \mathbf{x}_n) = 0.005 < \overline{L}(0.7; \mathbf{x}_n) = 0.773$ e si sceglie, di conseguenza $\theta_2 = 0.7$ perchè più verosimile tra i due valori. Il rapporto delle verosimiglianze $\lambda_{12}(\mathbf{x}_n) = 1/154.6$ ci dice inoltre che, alla luce dei dati osservati, l'ipotesi H_2 è circa 155 più plausibile dell'ipotesi H_1. □

Confronto tra ipotesi composte

Quando Θ_i non è un singolo punto di Θ, l'ipotesi H_i si dice *composta*. La fdv è una funzione di punto definita in Θ e non consente di assegnare valutazioni di verosimiglianza ad insiemi non puntuali dello spazio Θ. Questo fatto determina una difficoltà nel caso di confronto tra ipotesi composte. La valutazione della plausibilità di un sottoinsieme non puntuale Θ_i di Θ può essere sffettuata con la quantità

$$\sup_{\theta \in \Theta_i} L(\theta; \mathbf{x}_n).$$

In questo caso il confronto tra due ipotesi composte H_1 e H_2 si effettua con il *rapporto di verosimiglianze massimizzate*

$$\lambda_{12}^m(\mathbf{x}_n) = \frac{\sup_{\theta \in \Theta_1} L(\theta; \mathbf{x}_n)}{\sup_{\theta \in \Theta_2} L(\theta; \mathbf{x}_n)}.$$

Nell'impostazione frequentista (si veda il Capitolo 6), si assume di solito che $\Theta = \Theta_1 \cup \Theta_2$ e, per il confronto tra ipotesi composte, si usa comunemente la seguente definizione di rapporto di verosimiglianze massimizzate:

$$\lambda_{12}^m(\mathbf{x}_n) = \frac{\sup_{\theta \in \Theta_1} L(\theta; \mathbf{x}_n)}{\sup_{\theta \in \Theta} L(\theta; \mathbf{x}_n)}.$$

Si noti tuttavia che, in questo secondo caso, $0 \leq \lambda_{12}^m \leq 1$ e quindi il valore di indifferenza tra le due ipotesi non coincide con l'unità.

2.2.4 La funzione di log-verosimiglianza nei tre problemi ipotetici

Nei precedenti paragrafi abbiamo fatto spesso ricorso alla funzione di log-verosimiglianza $\ell(\theta; \mathbf{x}_n) = \log L(\theta; \mathbf{x}_n)$ (dove la funzione log è presa a base maggiore di uno e quindi crescente) in luogo della fdv per risolvere problemi inferenziali riguardanti θ. Come si è detto in precedenza, $\ell(\theta; \mathbf{x}_n)$ è spesso analiticamente più semplice da trattare di $L(\theta; \mathbf{x}_n)$ pur contenendo la stessa informazione che il campione osservato $\mathbf{x}_n$ esprime riguardo a θ. Questo è vero anche per la funzione di log-verosimiglianza relativa. Assumendo che la fdv abbia un massimo interno a Θ abbiamo infatti che

$$\overline{\ell}(\theta; \mathbf{x}_n) = \log \overline{L}(\theta; \mathbf{x}_n) = \log \frac{L(\theta; \mathbf{x}_n)}{L(\widehat{\theta}_{mv}; \mathbf{x}_n)} = \ell(\theta; \mathbf{x}_n) - \ell(\widehat{\theta}_{mv}; \mathbf{x}_n).$$

Le funzioni $L(\theta; \mathbf{x}_n), \overline{L}(\theta; \mathbf{x}_n), \ell(\theta; \mathbf{x}_n)$ e $\overline{\ell}(\theta; \mathbf{x}_n)$ possono essere utilizzate indifferentemente per determinare stime di massima verosimiglianza e per confrontare ipotesi. Infatti, se ad esempio consideriamo le funzioni L ed ℓ, si ha che

$$\widehat{\theta}_{mv} = \arg\max_{\theta \in \Theta} L(\theta; \mathbf{x}_n) = \arg\max_{\theta \in \Theta} \ell(\theta; \mathbf{x}_n).$$

Inoltre, per ogni coppia (θ_1, θ_2), grazie al fatto che la funzione logaritmica è crescente,

$$\ell(\theta_1; \mathbf{x}_n) > \ell(\theta_2; \mathbf{x}_n) \Leftrightarrow L(\theta_1; \mathbf{x}_n) > L(\theta_2; \mathbf{x}_n).$$

Per quanto riguarda gli insiemi di verosimiglianza, questi coincidono con gli *insiemi di log-verosimiglianza di livello* $\log q$:

$$L_q = \big\{\theta \in \Theta : \overline{L}(\theta; \mathbf{x}_n) \geq q\big\} = \big\{\theta \in \Theta : \overline{\ell}(\theta; \mathbf{x}_n) \geq \log q\big\}.$$

2.3 Informazione di Fisher osservata

Nell'Esempio 2.5 (modello bernoulliano) abbiamo osservato che, nel caso di due campioni di dimensioni $n = 10$ e $n = 50$ che danno luogo alla stessa smv (0.8), la maggiore informazione campionaria che si ha per $n = 50$ si concretizza nella maggiore concentrazione della fdv intorno al punto di massimo. Nel caso in cui vi sia un unico punto di massimo, il grado di concentrazione della fdv intorno a tale valore è un modo generale abbastanza intuitivo per valutare l'informatività della stessa fdv. Nel caso di problemi regolari di stima (Definizione 2.8), il grado di concentrazione della fdv è quantificato dalla nozione che segue.

2.17 Definizione (Informazione osservata). Dato un modello statistico $(\mathcal{X}^n, f_n(\mathbf{x}_n; \theta), \theta \in \Theta)$ la cui stima di massima verosimiglianza $\widehat{\theta}_{mv}$ è un punto interno all'insieme $\Theta \subseteq \mathbb{R}^1$ e la cui funzione di log-verosimiglianza è derivabile due volte in $\theta = \widehat{\theta}_{mv}$ (problema regolare di stima), si definisce *informazione di Fisher osservata* la quantità:

$$\mathcal{I}_n(\widehat{\theta}_{mv}; \mathbf{x}_n) = -\left.\frac{d^2}{d\theta^2} \log L(\theta; \mathbf{x}_n)\right|_{\theta = \widehat{\theta}_{mv}} = -\left.\frac{d^2}{d\theta^2}\, \ell(\theta; \mathbf{x}_n)\right|_{\theta = \widehat{\theta}_{mv}}.$$

La funzione

$$\mathcal{I}_n(\theta; \mathbf{x}_n) = -\frac{d^2}{d\theta^2}\, \ell(\theta; \mathbf{x}_n)$$

si chiama *funzione di informazione* (o *informazione di Fisher*). □

2.18 Osservazione. La precedente definizione richiede la derivabilità della funzione di log-verosimoglianza. Pertanto la funzione di informazione (e quindi l'informazione osservata) non possono essere determinate nei modelli con supporto dipendente dal parametro incognito. □

La funzione di informazione dipende dalla coppia $(\theta, \mathbf{x}_n)$ mentre l'informazione di Fisher osservata è una funzione dei dati campionari e non dipende dal parametro incognito. Per semplicità, quando useremo la notazione $\mathcal{I}_n$ oppure $\mathcal{I}_n(\widehat{\theta})$ intenderemo che si tratta dell'informazione osservata. La funzione di informazione può anche essere espressa come

$$\mathcal{I}_n(\theta; \mathbf{x}_n) = -\frac{d}{d\theta} S(\theta; \mathbf{x}_n)$$

dove $S(\theta; \mathbf{x}_n) = \frac{d}{d\theta} \log L(\theta; \mathbf{x}_n)$ si chiama *funzione di punteggio* (in inglese *score function*).

Ricordando che, in un punto di massimo di $\ell(\theta; \mathbf{x}_n)$, la derivata seconda $\ell''(\widehat{\theta}_{mv}; \mathbf{x}_n)$ è negativa, necessariamente $\mathcal{I}_n$ è un numero reale positivo $\forall \mathbf{x}_n \in \mathcal{X}^n$. In condizioni di regolarità, la derivata seconda di $\ell(\theta)$ in $\widehat{\theta}_{mv}$ è negativa ed è una misura della concavità della log-verosimiglianza nel punto di massimo. Quanto maggiore è il valore assoluto della derivata seconda (e quindi di $-\ell''(\widehat{\theta}_{mv})$), tanto più accentuata è la concavità della log-verosimiglianza in $\widehat{\theta}_{mv}$. Quanto più accentuata è la concavità in $\ell(\widehat{\theta}_{mv})$, tanto più rapidamente decresce la funzione $\ell(\theta)$ in un intorno di $\widehat{\theta}_{mv}$. Così, anche per valori di θ in un intorno piccolo del punto di massimo, la differenza $\ell(\widehat{\theta}_{mv}) - \ell(\theta)$ può essere rilevante. Del resto, osservando che

$$\ell(\widehat{\theta}; \mathbf{x}_n) - \ell(\theta; \mathbf{x}_n) = \log \frac{L(\widehat{\theta}_{mv}; \mathbf{x}_n)}{L(\theta; \mathbf{x}_n)} = \log \frac{\overline{L}(\widehat{\theta}_{mv}; \mathbf{x}_n)}{\overline{L}(\theta; \mathbf{x}_n)}, \tag{2.3}$$

si deduce che valori elevati di $\mathcal{I}_n$ corrispondono a valori elevati del rapporto tra le verosimiglianze in $\theta = \widehat{\theta}_{mv}$ e punti di θ in un intorno di $\widehat{\theta}_{mv}$.

2.19 Esempio (Modello bernoulliano). In questo caso abbiamo visto che $\ell(\theta; \mathbf{x}_n) = n\overline{x}_n \log\theta + n(1-\overline{x}_n)\log(1-\theta)$. Quindi

$$\ell'(\theta; \mathbf{x}_n) = \frac{n\overline{x}_n}{\theta} - \frac{n(1-\overline{x}_n)}{1-\theta} \quad \text{e} \quad \ell''(\theta; \mathbf{x}_n) = -\frac{n\overline{x}_n}{\theta^2} - \frac{n(1-\overline{x}_n)}{(1-\theta)^2}$$

da cui, ricordando che $\widehat{\theta}_{mv} = \overline{x}_n$, si ottiene

$$\mathcal{I}_n = -\ell''(\theta; \mathbf{x}_n)\Big|_{\theta=\overline{x}_n} = \frac{n}{\overline{x}_n(1-\overline{x}_n)}.$$

Questo risultato dimostra formalmente quanto emerso in precedenti esempi numerici particolari: a parità di valore osservato per $\widehat{\theta}_{mv}$, maggiore la dimensione campionaria, maggiore l'informazione che otteniamo dalla fdv. Nel caso specifico, per $n = 10$ e $\widehat{\theta}_{mv} = 0.8$, si ha $\mathcal{I}_n = 62.5$, mentre per $n = 50$ si ha che $\mathcal{I}_n = 312.5$. Il rapporto tra il valore di $\mathcal{I}_n$ nei due casi è pari a 5. □

2.20 Esempio(**Modello di Poisson**). Consideriamo un campione estratto da una variabile aleatoria di base di Poisson, con legge di probabilità $f_X(x;\theta) = e^{-\theta}\theta^x/x!, \quad \theta > 0$. Per questo modello la distribuzione di probabilità del campione casuale $\mathbf{X}_n$, la funzione di verosimiglianza e la funzione di log-verosimiglianza di θ associata al campione $\mathbf{x}_n$ sono, rispettivamente

$$f_n(\mathbf{x}_n;\theta) = \frac{e^{-n\theta}\ \theta^{n\overline{x}_n}}{\prod_{i=1}^n x_i!}, \qquad L(\theta; \mathbf{x}_n) = e^{-n\theta}\theta^{n\overline{x}_n}, \qquad \ell(\theta; \mathbf{x}_n) = -n\theta + n\ \overline{x}_n \log\theta,$$

dove l'espressione di $L(\theta; \mathbf{x}_n)$ coincide con il nucleo della fdv (sono stati eliminati tutti i valori costanti rispetto a θ). La soluzione dell'equazione di log-verosimiglianza

$$\ell'(\theta; \mathbf{x}_n) = -n + \frac{n\ \overline{x}_n}{\theta} = 0$$

permette di ottenere come punto candidato ad essere la smv il valore $\widehat{\theta}_{mv} = \overline{x}_n$ La derivata seconda di $\ell(\theta; \mathbf{x}_n)$ nel punto $\widehat{\theta}_{mv}$ è negativa:

$$\ell''(\theta; \mathbf{x}_n)\Big|_{\theta=\widehat{\theta}_{mv}} = -\frac{n\ \overline{x}_n}{\theta^2}\Big|_{\theta=\overline{x}_n} = -\frac{n}{\overline{x}_n} < 0.$$

Pertanto $\widehat{\theta}_{mv}$ è la (unica) smv di θ e l'informazione osservata di Fisher $\mathcal{I}_n = n/\overline{x}_n$ cresce al crescere di n. □

2.4 Analisi della verosimiglianza in alcuni modelli statistici

Per i più comuni modelli uniparametrici riportiamo di seguito l'espressione di $L(\theta; \mathbf{x}_n)$, $\ell(\theta; \mathbf{x}_n)$ e le seguenti quantità: equazione di log-verosimiglianza, $\widehat{\theta}_{mv}$, $\ell''(\theta; \mathbf{x}_n)$ e $\mathcal{I}_n$.

2.21 Esempio (**Modello bernoulliano**). Riassumendo quanto visto nei precedenti esempi, la stima di massima verosimiglianza è $\widehat{\theta} = \overline{x}_n$ e che

$$\ell''(\theta)\Big|_{\theta=\widehat{\theta}} = -\frac{n}{\overline{x}_n(1-\overline{x}_n)} < 0.$$

Quindi l'informazione osservata di Fisher per questo modello è

$$\mathcal{I}_n = \frac{n}{\overline{x}_n(1-\overline{x}_n)}.$$

□

2.22 Esempio (**Modello binomiale**). Consideriamo una variabile aleatoria di base binomiale, con legge di probabilità

$$f_X(x;\theta) = \binom{k}{x}\theta^x(1-\theta)^{(k-x)}, \qquad x = 0, 1, \ldots, k.$$

In questo caso, la distribuzione di probabilità congiunta di un campione casuale $\mathbf{X}_n$, la funzione di verosimiglianza e la funzione di log-verosimiglianza sono, rispettivamente:

$$f_n(\mathbf{x}_n;\theta) = \left[\prod_{i=1}^{n}\binom{k}{x_i}\right]\theta^{n\overline{x}_n}\ (1-\theta)^{nk-n\overline{x}_n}, \qquad L(\theta;\mathbf{x}_n) = \theta^{n\overline{x}_n}\ (1-\theta)^{nk-n\overline{x}_n},$$

$$\ell(\theta;\mathbf{x}_n) = n\ \overline{x}_n \log\theta + n(k-\overline{x}_n)\log(1-\theta),$$

dove, dall'espressione di $L(\theta; \mathbf{x}_n)$, sono stati eliminati tutti i valori costanti rispetto a θ. La soluzione dell'equazione di log-verosimiglianza

$$\ell'(\theta; \mathbf{x}_n) = \frac{n\ \overline{x}_n}{\theta} - \frac{n(k-\overline{x}_n)}{(1-\theta)} = 0,$$

è la stima di massima verosimiglianza di θ:

$$\widehat{\theta}_{mv} = \frac{\overline{x}_n}{k}.$$

Infatti, per la derivata seconda di $\ell(\theta)$ nel punto $\theta = \widehat{\theta}_{mv}$ si ha

$$\ell''(\theta; \mathbf{x}_n)\Big|_{\theta=\widehat{\theta}_{mv}} = -\frac{n\ \overline{x}_n}{\theta^2} - \frac{n(k-\overline{x}_n)}{(1-\theta)^2}\Big|_{\theta=\overline{x}_n/k} = -\frac{nk^3}{\overline{x}_n(k-\overline{x}_n)} < 0.$$

L'informazione osservata di Fisher è

$$\mathcal{I}_n = \frac{nk^3}{\overline{x}_n(k-\overline{x}_n)}\ .$$

□

2.23 Esempio (**Modello geometrico**). Consideriamo un campione estratto da una variabile aleatoria di base geometrica, che rappresenta il numero di prove da effettuare prima di avere un successo in una successione di prove bernoulliane, con probabilità di successo θ. La variabile aleatoria $X \sim \text{Geom}(\theta)$ assume valori nell'insieme $\mathcal{X} = \{0, 1, 2, \ldots\}$, con legge di probabilità

$$f_X(x; \theta) = (1-\theta)^x\theta, \qquad x = 0, 1, 2, \ldots.$$

La distribuzione di massa di probabilità del campione casuale $\mathbf{X}_n$, la funzione di verosimiglianza e la funzione di log-verosimiglianza sono, rispettivamente:

$$\begin{aligned} f_n(\mathbf{x}_n; \theta) &= \prod_{i=1}^n (1-\theta)^{x_i}\theta = (1-\theta)^{\sum_{i=1}^n x_i}\ \theta^n, \qquad L(\theta; \mathbf{x}_n) = (1-\theta)^{\sum_{i=1}^n x_i}\ \theta^n, \\ \ell(\theta; \mathbf{x}_n) &= \sum_{i=1}^n x_i \log(1-\theta) + n\log\theta. \end{aligned}$$

La soluzione dell'equazione di log-verosimiglianza

$$\ell'(\theta) = -\frac{\sum_{i=1}^n x_i}{1-\theta} + \frac{n}{\theta} = 0$$

fornisce la smv

$$\widehat{\theta}_{mv} = \frac{n}{n + \sum_{i=1}^n x_i},$$

poichè $\ell''(\theta) = -\dfrac{\sum_{i=1}^n x_i}{(1-\theta)^2} - \dfrac{n}{\theta^2}$ è negativa per ogni $0 \le \theta \le 1$. L'informazione osservata di Fisher è

$$\mathcal{I}_n = \frac{\sum_{i=1}^n x_i}{(1-\theta)^2} + \frac{n}{\theta^2}\Big|_{\theta=\widehat{\theta}_{mv}} = \frac{(n + \sum_{i=1}^n x_i)^3}{n\ \sum_{i=1}^n x_i}.$$

□

2.24 Esempio (**Modello di Poisson**). Vedi Esempio 2.20. □

2.25 Esempio (**Modello esponenziale negativo**). Per questo modello si ha che

$$f_n(\mathbf{x}_n; \theta) = \prod_{i=1}^n \theta\ e^{-\theta x_i} = \theta^n\ e^{-\theta\sum_{i=1}^n x_i}, \qquad L(\theta; \mathbf{x}_n) = \theta^n\ e^{-\theta n\overline{x}_n}, \qquad \ell(\theta; \mathbf{x}_n) = n\ \log\theta - n\overline{x}_n\theta.$$

La soluzione dell'equazione di log-verosimiglianza

$$\ell'(\theta; \mathbf{x}_n) = \frac{n}{\theta} - n\,\overline{x}_n = 0,$$

fornisce la smv di θ

$$\widehat{\theta}_{mv} = \frac{1}{\overline{x}_n},$$

in quanto la derivata seconda di $\ell(\theta; \mathbf{x}_n)$ è: $\ell''(\theta; \mathbf{x}_n) = -\frac{n}{\theta^2} < 0$. L'informazione osservata di Fisher è

$$\mathcal{I}_n = n\overline{x}_n{}^2.$$

□

2.26 Esempio (**Modello esponenziale**). La legge di probabilità congiunta del campione casuale $\mathbf{X}_n$, la funzione di verosimiglianza e la funzione di log-verosimiglianza sono:

$$f_n(\mathbf{x}_n; \theta) = \theta^{-n}\, e^{-\sum x_i/\theta} \qquad L(\theta; \mathbf{x}_n) = \theta^{-n}\, e^{-n\overline{x}_n/\theta} \qquad \ell(\theta; \mathbf{x}_n) = -n\log\theta - \frac{n\overline{x}_n}{\theta}.$$

L'equazione di log-verosimiglianza $\ell'(\theta) = -n/\theta + n\overline{x}_n/\theta^2 = 0$ ha come soluzione

$$\widehat{\theta}_{mv} = \overline{x}_n,$$

che è la smv di θ poichè

$$\ell''(\theta; \mathbf{x}_n)\Big|_{\theta=\widehat{\theta}} = \frac{n}{\theta^2} - \frac{2\,n\,\overline{x}_n}{\theta^3}\Big|_{\theta=\overline{x}_n} = -\frac{n}{\overline{x}_n{}^2} < 0.$$

L'informazione osservata è

$$\mathcal{I}_n = \frac{n}{\overline{x}_n{}^2}.$$

□

2.27 Esempio (**Modello normale**). Per il modello normale con varianza nota si ha che

$$L(\theta; \mathbf{x}_n) = \exp\left\{-\frac{n}{2\sigma^2}\,(\overline{x}_n - \theta)^2\right\}, \qquad \ell(\theta; \mathbf{x}_n) = -\frac{n}{2\sigma^2}\,(\overline{x}_n - \theta)^2,$$

e l'equazione di log-verosimiglianza è

$$\ell'(\theta; \mathbf{x}_n) = \frac{n(\overline{x}_n - \theta)}{\sigma^2} = 0.$$

La stima di massima verosimiglianza è

$$\widehat{\theta}_{mv} = \overline{x}_n,$$

in quanto $\ell''(\theta; \mathbf{x}_n) = -n/\sigma^2 < 0$ e

$$\mathcal{I}_n = \frac{n}{\sigma^2}.$$

Per il modello normale con valore atteso μ noto e varianza incognita θ, le funzioni di verosimiglianza e di log-verosimiglianza sono:

$$L(\theta; \mathbf{x}_n) = \frac{1}{\theta^{n/2}} \exp\left\{-\frac{\sum(x_i - \mu)^2}{2\theta}\right\}, \qquad \ell(\theta; \mathbf{x}_n) = -\frac{n}{2}\log\theta - \frac{\sum(x_i - \mu)^2}{2\theta}.$$

L'equazione di log-verosimiglianza

$$\ell'(\theta;\mathbf{x}_n) = -\frac{n}{2\theta} + \frac{\sum(x_i-\mu)^2}{2\theta^2} = 0$$

ha soluzione

$$\widehat{\theta}_{mv} = \frac{\sum(x_i-\mu)^2}{n}$$

che è la smv di θ poichè

$$\ell''(\theta;\mathbf{x}_n)\Big|_{\theta=\widehat{\theta}_{mv}} = \quad \frac{n}{2\theta^2} - \frac{\sum(x_i-\mu)^2}{\theta^3}\Big|_{\theta=\sum(x_i-\mu)^2/n} = \quad -\frac{n^3}{2\big[\sum(x_i-\mu)^2\big]^2} < 0.$$

L'informazione osservata è quindi

$$\mathcal{I}_n = \frac{n^3}{2\big[\sum(x_i-\mu)^2\big]^2} = \frac{n}{2\widehat{\theta}^2_{mv}}.$$

□

2.5 Approssimazione normale della fdv

Sotto opportune condizioni di regolarità, la funzione di verosimiglianza di parametro unidimensionale può essere approssimata, in un intorno di $\widehat{\theta}_{mv}$, da una funzione proporzionale a una densità $\mathrm{N}(\widehat{\theta}_{mv}, [\mathcal{I}_n(\widehat{\theta}_{mv})]^{-1})$. Siano $L(\theta)$ ed $\ell(\theta)$ le funzioni di verosimiglianza e di log-verosimiglianza associate ad un campione osservato e $\mathcal{I}_n(\widehat{\theta}_{mv})$ l'informazione osservata di Fisher (omettiamo per semplicità la dipendenza dal campione osservato) e sia $\Theta \subseteq \mathbb{R}$. Se $\ell(\theta)$ ammette un punto di massimo $(\widehat{\theta}_{mv})$ interno a Θ (che supponiamo unico) e se può essere sviluppata attorno a questo punto in serie di Taylor fino al secondo grado, possiamo scrivere:

$$\ell(\theta) = \ell(\widehat{\theta}_{mv}) + (\theta-\widehat{\theta}_{mv})\,\ell'(\theta)\big|_{\theta=\widehat{\theta}_{mv}} + \frac{1}{2}\,(\theta-\widehat{\theta}_{mv})^2\ell''(\theta)\big|_{\theta=\widehat{\theta}_{mv}} + R(\theta)$$

dove il resto $R(\theta)$ è una funzione dei dati osservati e del parametro che assumiamo essere tanto più trascurabile quanto più elevata la numerosità campionaria. Dato che $\ell'(\widehat{\theta}_{mv}) = 0$, trascurando il resto, si ottiene:

$$\ell(\theta) - \ell(\widehat{\theta}_{mv}) = \log \overline{L}(\theta) \approx \frac{1}{2}(\theta-\widehat{\theta}_{mv})^2\ell''(\theta)\big|_{\theta=\widehat{\theta}_{mv}}.$$

Pertanto

$$\overline{L}(\theta;\mathbf{x}_n) \approx \exp\left\{\frac{1}{2}(\theta-\widehat{\theta})^2\ell''(\theta)\big|_{\theta=\widehat{\theta}}\right\} = \exp\left\{-\frac{1}{2}(\theta-\widehat{\theta}_{mv})^2\mathcal{I}_n(\widehat{\theta}_{mv})\right\}.$$

L'ultima eguaglianza è dovuta al fatto che $\mathcal{I}_n(\widehat{\theta}_{mv}) = -\ell''(\theta)\big|_{\theta=\widehat{\theta}_{mv}}$. Possiamo quindi utilizzare, come approssimazione di $\overline{L}(\theta;\mathbf{x}_n)$, la funzione $\widetilde{L}:\ \Theta \to [0,1]$, definita da

$$\widetilde{L}(\theta;\mathbf{x}_n) = \exp\left\{-\frac{1}{2}(\theta-\widehat{\theta}_{mv})^2\mathcal{I}_n\right\} \qquad \theta\in\Theta. \tag{2.4}$$

L'approssimazione $\widetilde{L}$ per $\overline{L}$ è proporzionale a una densità gaussiana di parametri $(\widehat{\theta}_{mv}, [\mathcal{I}_n(\widehat{\theta}_{mv})]^{-1})$.

Il grafico di $\widetilde{L}$ in Θ ha quindi la forma tipica di una densità normale. Questo non vuol dire però che $\widetilde{L}$ sia interpretabile come se fosse funzione di densità, dal momento che, come si è più volte sottolineato, il parametro θ non è una variabile aleatoria.

Da quanto visto discende che, nei modelli regolari, l'informazione osservata $\mathcal{I}_n(\widehat{\theta}_{mv})$ controlla sia la curvatura di ℓ in $\widehat{\theta}_{mv}$ che la dispersione dell'approssimazione normale intorno allo stesso punto. L'approssimazione normale di $\overline{L}$ è tanto migliore quanto più piccolo è l'intorno di $\widehat{\theta}_{mv}$ che si considera e, a parità di altre condizioni, quanto più elevata è la dimensione campionaria.

2.5.1 Insiemi di verosimiglianza approssimati

La (2.4) consente di determinare intervalli di verosimiglianza di livello q approssimati per il parametro di un modello statistico regolare. L'utilità di questo risultato risiede nel fatto che molto spesso, anche per i più comuni modelli, le soluzioni della disequazione che definisce l'insieme L_q non sono determinabili in forma esplicita. Ciò vuol dire che non siamo in grado (o comunque non è agevole) determinare gli estremi dell'insieme di verosimiglianza. In questi casi è quindi necessario ricorrere a soluzioni numeriche. Un'alternativa è appunto rappresentata dagli insiemi di verosimiglianza che si possono ottenere dall'approssimazione normale, definiti come segue:

$$\widetilde{L}_q = \left\{\theta \in \Theta :\ \widetilde{L}(\theta; \mathbf{x}_n) \geq q\right\}, \qquad q \in (0,1).$$

Dalla (2.4) discende infatti che $L_q \approx \widetilde{L}_q$ dove

$$\widetilde{L}_q = \left\{\theta \in \Theta :\ \exp\left\{-\frac{1}{2}(\theta - \widehat{\theta}_{mv})^2 \mathcal{I}_n(\widehat{\theta}_{mv})\right\} \geq q\right\}, \quad q \in (0,1).$$

Con calcoli del tutto analoghi a quelli svolti per determinare l'insieme L_q per il modello $\mathrm{N}(\theta, \sigma^2)$, si verifica facilmente che

$$\widetilde{L}q = \left[\widehat{\theta}_{mv} - k_q\sqrt{[\mathcal{I}_n(\widehat{\theta}_{mv})]^{-1}}, \widehat{\theta}_{mv} + k_q\sqrt{[\mathcal{I}_n(\widehat{\theta}_{mv})]^{-1}}\right], \quad k_q = \sqrt{-2\ln q}. \tag{2.5}$$

Si noti per prima cosa che per ogni $q \in (0,1)$, $-2\ln q > 0$ e quindi che $k_q \in \mathbb{R}^+$. Inoltre, si osservi che, maggiore è $\mathcal{I}_n(\widehat{\theta}_{mv})$ (ovvero l'informazione campionaria), minore è la lunghezza dell'intervallo di verosimiglianza per θ.

La perdita di precisione nel calcolo dell'insieme di verosimiglianza che si ha ricorrendo all'approssimazione $\widetilde{L}_q$ risulta compensata, in particolar modo per n adeguatamente elevato, dalla disponibilità di formule esplicite per il calcolo degli estremi degli intervalli di stima.

2.28 Esempio (**Modello bernoulliano**). Ricordando che, per il modello in esame,

$$\widehat{\theta}_{mv} = \overline{x}_n \qquad \text{e che} \qquad \mathcal{I}_n(\widehat{\theta}_{mv}) = \frac{n}{\overline{x}_n(1-\overline{x}_n)}$$

si ha che

$$\widetilde{L}_q = \left[\overline{x}_n - k_q\sqrt{\frac{\overline{x}_n(1-\overline{x}_n)}{n}},\ \overline{x}_n + k_q\sqrt{\frac{\overline{x}_n(1-\overline{x}_n)}{n}}\right].$$

Se ad esempio si considera un campione di dimensione $n = 10$ in cui $\overline{x}_n = 1/2$ e se si assume che $\sigma^2 = 1$, scegliendo il valore $q = 0.147$ si ottiene

$$\left[\frac{1}{2} - \sqrt{-2\ln(0.147)}\frac{1/2}{\sqrt{10}}, \frac{1}{2} + \sqrt{-2\ln(0.147)}\frac{1/2}{\sqrt{10}}\right] = [0.19, 0.81]\,.$$

Se, a parità del valore di $\overline{x}_n$, avessimo $n = 100$, l'intervallo dello stesso livello q sarebbe $[0.40, 0.60]$ e quindi di lunghezza minore di quello ottenuto con $n = 10$. □

2.29 Esempio (**Modello di Poisson**). Ricordando che, per il modello in esame,

$$\widehat{\theta}_{mv} = \overline{x}_n \qquad \text{e che} \qquad \mathcal{I}_n(\widehat{\theta}_{mv}) = \frac{n}{\overline{x}_n}$$

si ha che

$$\widetilde{L}_q = \left[\overline{x}_n - k_q \sqrt{\frac{\overline{x}_n}{n}},\ \overline{x}_n + k_q \sqrt{\frac{\overline{x}_n}{n}} \right].$$

□

2.30 Esempio (Modello normale, valore atteso noto). Abbiamo visto in precedenza che, per un modello normale con media nota μ e varianza incognita $\theta = \sigma^2$, si ha:

$$\widehat{\theta}_{mv} = \frac{1}{n} \sum (x_i - \mu)^2 \qquad \text{e} \qquad \mathcal{I}_n(\widehat{\theta}_{mv}) = \frac{n}{2\widehat{\theta}^2_{mv}}.$$

Gli estremi dell'intervallo $\widetilde{L}_q$ per $\theta = \sigma^2$ sono dati da:

$$\widehat{\theta}_{mv} \pm k_q \sqrt{[\mathcal{I}_n(\widehat{\theta}_{mv})]^{-1}} = \widehat{\theta}_{mv} \pm \widehat{\theta}_{mv} k_q \sqrt{\frac{2}{n}}, \qquad q \in (0, 1).$$

La figura 2.6 mostra la fdv esatta (linea continua) e l'approssimazione normale (linea tratteggiata) nel caso di un campione di piccola dimensione ($n = 10$). L'approssimazione risulta accurata solo per un intorno molto piccolo di $\widehat{\theta}_{mv}$. L'approssimazione dell'insieme di verosimiglianza esatta con quello approssimato risulta abbastanza grossolana. Al crescere di n l'approssimazione tende a migliorare sensibilmente. □

2.6 Il caso multiparametrico

Quando $\Theta \subseteq \mathbb{R}^k$, $k > 1$, il parametro incognito è un vettore $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$. La definizione di funzione di verosimiglianza e di quasi tutte le quantità ad essa collegata restano valide, con ovvie generalizzazioni. Ad esempio, una stima di massima verosimiglianza di $\boldsymbol{\theta}$ è un vettore $\widehat{\boldsymbol{\theta}}_{mv} = (\widehat{\theta}_1, \ldots, \widehat{\theta}_k)$ che massimizza la funzione $L(\boldsymbol{\theta}; \mathbf{x}_n)$, mentre la regione di stima L_q è ora un sottoinsieme di $\mathbb{R}^k$. Richiedono definizioni più generali di quelle date in precedenza l'informazione di Fisher e le quantità a questa collegate. Nella definizione che segue assumiamo che il vettore stima di massima verosimiglianza $\widehat{\theta}_{mv}$ sia un punto interno all'insieme Θ e che la funzione di log-verosimiglianza ammetta derivate parziali continue fino al secondo ordine.

2.31 Definizione (Matrice di informazione). Dato un modello statistico $(\mathcal{X}^n, f_n(\mathbf{x}_n; \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta)$, con $\Theta \subseteq \mathbb{R}^k$, $k > 1$, si chiama *matrice di informazione di Fisher* la matrice $\mathcal{I}_n(\theta)$ di dimensioni (k, k) con temine generico

$$[\mathcal{I}_n(\boldsymbol{\theta}; \mathbf{x}_n)]_{ij} = -\frac{\partial}{\partial \theta_j} \frac{\partial}{\partial \theta_i} \log L(\boldsymbol{\theta}; \mathbf{x}_n), \qquad i, j = 1, \ldots, k.$$

La matrice

$$[\mathcal{I}_n(\widehat{\boldsymbol{\theta}}_{mv}; \mathbf{x}_n)]_{ij} = -\frac{\partial}{\partial \theta_j} \frac{\partial}{\partial \theta_i} \log L(\boldsymbol{\theta}; \mathbf{x}_n)|_{\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_{mv}}, \qquad i, j = 1, \ldots, k,$$

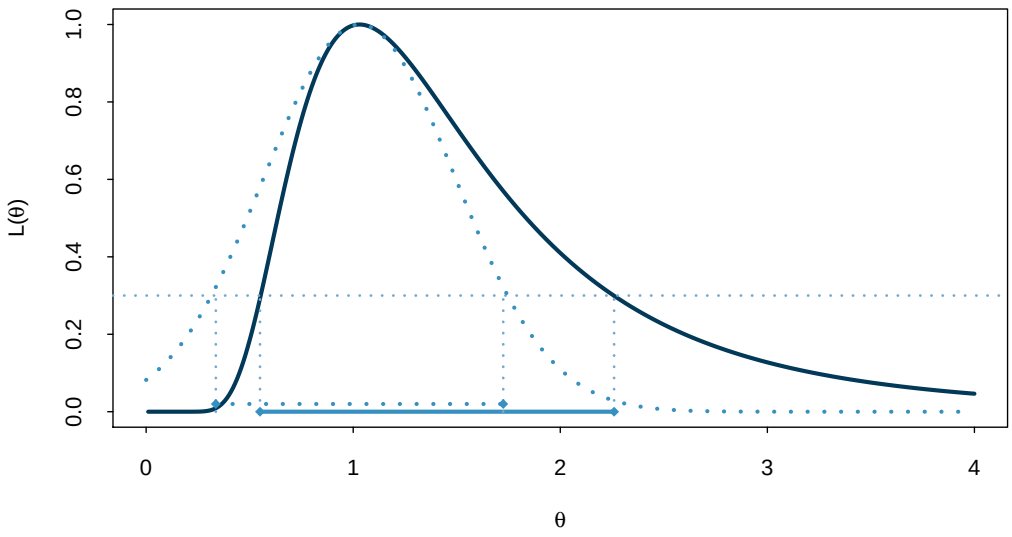

Figura 2.6: *Insieme* $\overline{L}_q$ *e approssimazione* $\widetilde{L}_q$ *per il parametro del modello* $N(\mu, \theta)$.

si chiama *matrice di informazione osservata*. □

La matrice di informazione di Fisher è quindi la versione multidimensionale dell'informazione di Fisher. La funzione di punteggio (score function) diventa, nel caso $k > 1$, un vettore:

$$S(\boldsymbol{\theta}; \mathbf{x}_n) = \left[\frac{\partial}{\partial \theta_1} \log L(\boldsymbol{\theta}; \mathbf{x}_n), \ldots, \frac{\partial}{\partial \theta_k} \log L(\boldsymbol{\theta}; \mathbf{x}_n) \right].$$

Si può verificare che, nel caso in cui $\widehat{\boldsymbol{\theta}}_{mv}$ sia unico, vale la seguente approssimazione normale:

$$\overline{L}(\boldsymbol{\theta}; \mathbf{x}_n) \simeq \widetilde{L}(\boldsymbol{\theta}; \mathbf{x}_n) = \exp\left\{ -\frac{1}{2} (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{mv})^T \mathcal{I}_n(\widehat{\boldsymbol{\theta}}_{mv}) (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{mv}) \right\}.$$

L'approssimazione $\widetilde{L}$ della fdv relativa è quindi proporzionale a una densità $\mathrm{N}_k(\widehat{\boldsymbol{\theta}}_{mv}, [\mathcal{I}_n(\widehat{\boldsymbol{\theta}}_{mv})]^{-1})$, la cui matrice delle varianze e covarianze è la matrice inversa della matrice di informazione osservata.

2.6.1 Stima di massima verosimiglianza

Quando il parametro $\boldsymbol{\theta}$ è un vettore, la determinazione della smv richiede la massimizzazione di una funzione di più variabili. Se la funzione di log-verosimiglianza è differenziabile, il sistema che si ottiene ponendo uguali a zero le derivate parziali di $\ell(\boldsymbol{\theta}; \mathbf{x}_n)$ rispetto alle componenti θ_j di $\boldsymbol{\theta}$, $j = 1, \ldots, k$ (detto *sistema delle log-verosimiglianze*) individua punti in Θ che possono essere smv. La verifica che un punto candidato ad essere un massimo lo sia effettivamente coinvolge le derivate parziali di secondo ordine. Per il caso $k = 2$, affinchè un punto $\widehat{\boldsymbol{\theta}}_{mv} = (\widehat{\theta}_1, \widehat{\theta}_2) \in \Theta$ sia una smv di $\boldsymbol{\theta} = (\theta_1, \theta_2)$, devono valere le seguenti condizioni.

1. Le componenti del vettore $\widehat{\boldsymbol{\theta}}_{mv} = (\widehat{\theta}_1, \widehat{\theta}_2)$ devono essere soluzioni del sistema di equazioni di log-verosimiglianza, ovvero:

$$\begin{cases} \dfrac{\partial}{\partial\theta_1}\ell(\theta_1,\theta_2)\,\Big|_{\theta_1=\widehat{\theta}_1,\theta_2=\widehat{\theta}_2} = 0 \\ \\ \dfrac{\partial}{\partial\theta_2}\ell(\theta_1,\theta_2)\,\Big|_{\theta_1=\widehat{\theta}_1,\theta_2=\widehat{\theta}_2} = 0 \end{cases}.$$

2. La derivata parziale seconda $\partial^2\ell(\theta_1,\theta_2)/\partial\theta_1^2$, calcolata in $\widehat{\boldsymbol{\theta}}_{mv}$, deve essere negativa:

$$\frac{\partial^2}{\partial\theta_1^2}\ell(\theta_1,\theta_2)|_{\theta_1=\widehat{\theta}_1,\theta_2=\widehat{\theta}_2} < 0,$$

3. Il determinante J della matrice hessiana di $\ell(\theta_1,\theta_2)$ (matrice delle derivate parziali seconde) calcolato in $\widehat{\boldsymbol{\theta}}_{mv}$ deve essere positivo:

$$|J| = \left\{ \frac{\partial^2}{\partial\theta_1^2}\ell(\theta_1,\theta_2)\frac{\partial^2}{\partial\theta_2^2}\ell(\theta_1,\theta_2) - \left[\frac{\partial^2}{\partial\theta_1\partial\theta_2}\ell(\theta_1,\theta_2)\right]^2 \right\}\Big|_{\theta_1=\widehat{\theta}_1,\theta_2=\widehat{\theta}_2} > 0.$$

2.32 Esempio (**Modello normale**).
In questo caso la fdv e la log-verosimiglianza sono

$$L(\theta_1,\theta_2;\mathbf{x}_n) = \frac{1}{\theta_2^{n/2}}\exp\left\{-\frac{\sum(x_i-\theta_1)^2}{2\theta_2}\right\}, \qquad \ell(\theta_1,\theta_2) = -\frac{n}{2}\log\theta_2 - \frac{\sum(x_i-\theta_1)^2}{2\theta_2}.$$

Il sistema di equazioni di log-verosimiglianza è quindi

$$\begin{cases} \dfrac{\partial}{\partial\theta_1}\ell(\theta_1,\theta_2) = \dfrac{1}{\theta_2}\sum(x_i-\theta_1) = 0 \\ \\ \dfrac{\partial}{\partial\theta_2}\ell(\theta_1,\theta_2) = -\dfrac{n}{2\theta_2} + \dfrac{\sum(x_i-\theta_1)^2}{2\theta_2^2} = 0 \end{cases}$$

che ha soluzione $\widehat{\boldsymbol{\theta}}_{mv} = (\widehat{\theta}_1, \widehat{\theta}_2) = \big(\overline{x}_n, \sum(x_i-\overline{x}_n)^2/n\big) = (\overline{x}_n, \widehat{\sigma}_n^2)$. Le derivate parziali seconde di ℓ sono:

$$\frac{\partial^2}{\partial\theta_1^2}\ell(\theta_1,\theta_2) = -\frac{n}{\theta_2}, \quad \frac{\partial^2}{\partial\theta_2^2}\ell(\theta_1,\theta_2) = \frac{n}{2\,\theta_2^2} - \frac{\sum(x_i-\theta_1)^2}{\theta_2^3}, \quad \frac{\partial^2}{\partial\theta_1\partial\theta_2}\ell(\theta_1,\theta_2) = -\frac{\sum(x_i-\theta_1)}{\theta_2^2}.$$

La matrice hessiana è pertanto

$$\begin{pmatrix} -\frac{n}{\theta_2}, & -\frac{\sum(x_i-\theta_1)}{\theta_2^2} \\ -\frac{\sum(x_i-\theta_1)}{\theta_2^2} & \frac{n}{2\,\theta_2^2} - \frac{\sum(x_i-\theta_1)^2}{\theta_2^3} \end{pmatrix}.$$

Osservando che i termini che corrispondono alle derivate seconde parziali miste si annullano in $\theta_1 = \overline{x}_n$, il determinante della matrice calcolato in $\overline{x}_n, \widehat{\sigma}_n^2$ diventa

$$-\frac{n}{\theta_2}\left(\frac{n}{2\theta_2^2} - \frac{\sum(x_i-\theta_1)^2}{\theta_2^3}\right)\Big|_{\theta_1=\overline{x}_n,\theta_2=\widehat{\sigma}_n^2} = \frac{n^2}{2(\widehat{\sigma}_n^2)^3} > 0.$$

Tutte le condizioni richieste sono soddisfatte e $\widehat{\boldsymbol{\theta}}_{mv} = (\overline{x}_n, \widehat{\sigma}_n^2)$ è la smv vettoriale di (θ_1,θ_2).
□

2.7 Inferenza in presenza di parametri di disturbo

In molti problemi inferenziali, il modello statistico in uso e la corrispondente fdv dipendono da un parametro vettoriale $\boldsymbol{\theta} = (\theta_1, \theta_2)$ Per semplicità assumiamo che θ_1 e θ_2 siano due scalari. Si possono verificare le seguenti situazioni.

a. Siamo interessati al parametro θ_1, ma le procedure inferenziali per questo parametro dipendono anche da θ_2. Questo caso si verifica, ad esempio, quando vogliamo stimare il parametro θ_1 nel modello $\mathrm{N}(\theta_1, \theta_2)$ attraverso gli insiemi L_q o quando vogliamo utilizzare i rapporti delle verosimiglianze, che dipendono anche dalla varianza incognita (θ_2) del modello.

b. Siamo interessati ad una opportuna funzione $g(\theta_1, \theta_2) = \lambda$ delle componenti di $\boldsymbol{\theta}$, ma la funzione di verosimiglianza, riscritta in modo da essere funzione di $\boldsymbol{\theta}$ attraverso la funzione $g(\theta_1, \theta_2)$, dipende anche da un'altra funzione $h(\theta_1, \theta_2) = \gamma$. Le procedure di stima di $g(\theta_1, \theta_2)$ dipendono quindi dalla funzione $h(\theta_1, \theta_2)$. Questo caso si verifica, ad esempio, quando consideriamo due v.a. indipendenti di Poisson di parametri θ_1 e θ_2 e siamo interessati ad effettuare inferenza sul parametro $\lambda = \theta_1 + \theta_2$. In questo caso si può verificare che la funzione di verosimiglianza si può riscrivere in funzione dei parametri (λ, γ), in cui $\gamma = \theta_1/(\theta_1 + \theta_2)$.

In entrambi i casi si tratta di problemi inferenziali in presenza di parametri di disturbo, in cui, in generale, assumiamo che la fdv possa essere espressa in termini di un parametro vettoriale, che indichiamo con la coppia (λ, γ), e in cui l'interesse effettivo verte sulla sola componente λ dell'intero vettore, il *parametro di interesse*. La restante componente del parametro, γ, il *parametro di disturbo*, pur non essendo di diretto interesse inferenziale, non può comunque essere ignorata poichè la funzione di verosimiglianza dipende dalla coppia (λ, γ) nel suo insieme e le procedure inferenziali per λ ottenute da questa sono funzioni di γ. Nei casi più semplici, quelli di tipo (a), λ e γ corrispondono a due sottovettori del parametro $\boldsymbol{\theta}$. Ad esempio, nel caso del modello $\mathrm{N}(\theta_1, \theta_2)$, se si è interessati a effettuare inferenza sul valore atteso, il parametro di interesse è $\lambda = \theta_1$ e il parametro di disturboè invece $\gamma = \theta_2$. In altri casi più complessi, quelli di tipo (b), il parametro di interesse $\lambda = g(\theta_1, \theta_2)$ è una funzione delle componenti del vettore dei parametri e $\gamma = h(\theta_1, \theta_2)$ è un'altra funzione dei parametri da cui dipendono sia la funzione di verosimiglianza scritta nella parametrizzazione (λ, γ) che le procedure inferenziali su λ.

Il problema dell'inferenza in presenza di un parametro di disturbo è, in genere, molto complesso. La situazione ideale, molto rara in pratica, è quella in cui la fdv si fattorizza nel modo seguente:

$$L(\lambda, \gamma; \mathbf{x}_n) = L_1(\lambda; \mathbf{x}_n) \times L_2(\gamma; \mathbf{x}_n),$$

dove L_1 e L_2 sono due funzioni rispettivamente solo di λ e di γ. In questo caso, in cui si realizza una sostanziale separazione dell'informazione sperimentale per il parametro di interesse e quello di disturbo, le procedure inferenziali per λ (smv, insiemi L_q, rapporti di verosimiglianze) si determinano sfruttando solo il fattore $L_1(\lambda; \mathbf{x}_n)$ della fdv e ignorando $L_2(\gamma; \mathbf{x}_n)$. Si parla infatti di *L-indipendenza* (o *ortogonalità*) tra i due parametri. In genere la fdv non presenta questa particolare struttura ed è quindi necessario individuare dei metodi che, partendo dalla fdv completa $Ł(\lambda, \gamma; \mathbf{x}_n)$, ci consentano di eliminare i parametri di disturbo. Non esiste tuttavia un unico metodo di eliminazione dei parametri di disturbo. Al contrario esistono molte proposte alternative, che danno luogo a surrogati della fdv originale, che però hanno il vantaggio di dipendere dal solo parametro di interesse. Ci si riferisce a queste funzioni di λ associate al campione osservato $\mathbf{x}_n$ come a *psuedo-verosimiglianze*. Tra i

metodi più importanti ricordiamo le verosimiglianze profilo, plug-in, marginale, condizionata e integrata [1].

2.8 Sufficienza

I dati campionari $(x_1, \ldots, x_n)$ forniscono informazioni sul parametro incognito del modello attraverso la funzione di verosimiglianza. Ci chiediamo ora se è possibile riassumere le n osservazioni in elaborazioni sintetiche (possibilmente di dimensioni inferiori a n) senza perdere informazioni su θ rispetto a quelle fornite dall'intero campione osservato. Più precisamente, siamo interessati a verificare se, per un determinato modello statistico, esistano delle *statistiche* (vedi Definizione 3.1) che, pur sintetizzando i dati campionari osservati, diano la stessa informazione che l'intero campione $\mathbf{x}_n$ fornisce su θ. L'obiettivo è identificare delle funzioni T dei dati $\mathbf{x}_n$ tali che la conoscenza del valore $T(\mathbf{x}_n) = t$ sia equivalente, ai fini dell'inferenza su θ, alla conoscenza degli n valori campionari $(x_1, \ldots, x_n)$, ovvero tali che la parte di informazione che necessariamente si perde passando da $\mathbf{x}_n$ a $T(\mathbf{x}_n)$ sia inessenziale ai fini inferenziali. Poichè $\mathbf{x}_n$ è un oggetto n-dimensionale (un punto di $\mathcal{X}^n$), il vantaggio si ha quando queste statistiche assumono valori in uno spazio $\mathcal{T}$ di dimensioni inferiori ad n. Dal momento che la fdv formalizza tutta l'informazione che i dati campionari forniscono sul parametro incognito di un modello, il valore osservato t di una statistica T ha lo stesso valore informativo dell'intero campione $\mathbf{x}_n$ se per ottenere la fdv è sufficiente conoscere il valore $T(\mathbf{x}_n) = t$

2.33 Esempio Nel caso di un modello statistico bernoulliano e di un campione osservato $\mathbf{x}_n$ la fdv è (nel caso i.i.d.)

$$\begin{aligned} L(\theta; \mathbf{x}_n) &= \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i} \\ &= \theta^t(1-\theta)^{n-t}, \end{aligned}$$

dove $t = \sum_{i=1}^n x_i$. In questo caso è sufficiente conoscere il valore della somma delle osservazioni campionarie e non i singoli valori delle x_i per ricostruire completamente la fdv. □

2.34 Definizione (Sufficienza). Dato il modello statistico $(\mathcal{X}^n, f_n(\mathbf{x}_n; \theta), \theta \in \Theta)$, una statistica $T : \mathcal{X}^n \to \mathcal{T}$ è *sufficiente* se esistono due funzioni $h : \mathcal{X}^n \to \mathbb{R}^+$ (funzione dei dati campionari $\mathbf{x}_n$) e $g : \Theta \times \mathcal{T} \to \mathbb{R}^+$ (di θ e di $T(\mathbf{x}_n)$) tali che

$$L(\theta; \mathbf{x}_n) = f_n(\mathbf{x}_n; \theta) = h(\mathbf{x}_n) \cdot g(\theta, T(\mathbf{x}_n)), \quad \forall \theta \in \Theta, \quad \forall \mathbf{x}_n \in \mathcal{X}^n.$$

□

In base alla definizione, nota anche come *Criterio di fattorizzazione di Fisher*, una statistica T è quindi sufficiente per il dato modello se il *nucleo* della fdv, g, dipende dai dati esclusivamente attraverso il valore t assunto dalla statistica T nel campione osservato $\mathbf{x}_n$.

2.35 Esempio (**Modello di Poisson**). Consideriamo $\mathbf{x}_n$, un campione casuale estratto da una popolazione di Poisson. In questo caso

$$L(\theta; \mathbf{x}_n) = f_n(\mathbf{x}_n; \theta) = \frac{1}{\prod_{i=1}^n x_i!}\theta^{\sum_{i=1}^n x_i} e^{-n\theta} = h(\mathbf{x}_n) \cdot g(\theta, T(\mathbf{x}_n)),$$

dove

$$h(\mathbf{x}_n) = \frac{1}{\prod_{i=1}^n x_i!} \qquad \text{e} \qquad g(\theta; T(\mathbf{x}_n)) = \theta^{\sum_{i=1}^n x_i} e^{-n\theta}.$$

[1] Per una trattazione più ampia dell'inferenza in presenza di parametri di disturbo, si veda, ad esempio, Piccinato (2009), pp. 140-148.

La statistica $T(\mathbf{x}_n) = \sum_{i=1}^n x_i$, è quindi una statistica sufficiente per il modello considerato. □

2.36 Esempio (**Modello esponenziale negativo**). Per il modello esponenziale negativo, la fdv è

$$L(\theta; \mathbf{x}_n) = f_n(\mathbf{x}_n; \theta) = \theta^n e^{-\theta \sum_{i=1}^n x_i} \prod_{i=1}^n I_{(0,+\infty)}(x_i) \propto \theta^n e^{-\theta n\ \overline{x}_n}. \tag{2.6}$$

da cui, per il criterio di fattorizzazione, si vede facilmente che $\sum_{i=1}^n x_i$ è una statistica sufficiente. □

2.37 Esempio (**Modello normale, varianza nota**). Per il modello normale $\mathrm{N}(\theta, \sigma^2)$, in cui θ è il parametro incognito e la varianza σ^2 è nota, la funzione di verosimiglianza ottenuta in (2.3) è:

$$\begin{aligned} L(\theta; \mathbf{x}_n) &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{n}{2\sigma^2} \sum (x_i - \theta)^2\right\} \prod_{i=1}^n I_{\mathbb{R}}(x_i) \\ &\propto \exp\left\{-\frac{n}{2\sigma^2} \sum (x_i - \theta)^2\right\} \\ &\propto \underbrace{\exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \overline{x}_n)^2\right\}}_{\text{in } h(\mathbf{x}_n)} \cdot \underbrace{\exp\left\{-\frac{n}{2\sigma^2}(\overline{x}_n - \theta)^2\right\}}_{g(\overline{x}_n, \theta)} \end{aligned}$$

Pertanto

$$L(\theta; \mathbf{x}_n) \propto \exp\left\{-\frac{n}{2\sigma^2}(\overline{x}_n - \theta)^2\right\}$$

da cui si evince che $\overline{x}_n$ è una statistica sufficiente per il modello. □

2.38 Esempio (**Modello normale, valore atteso noto**). In questo caso

$$L(\theta; \mathbf{x}_n) = \frac{1}{(2\pi\theta)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\theta} \sum_{i=1}^n (x_i - \mu)^2\right\} \prod_{i=1}^n I_{\mathbb{R}}(x_i).$$

Se si pone $h(\mathbf{x}_n) = \frac{1}{(2\pi)^{\frac{n}{2}}} \prod_{i=1}^n I_{\mathbb{R}}(x_i)$ e $g(t(\mathbf{x}_n); \theta) = \theta^{-n/2} \exp\left\{-\frac{1}{2\theta} \sum_{i=1}^n (x_i - \mu)^2\right\}$, si verifica che la statistica $\sum_{i=1}^n (x_i - \mu)^2$ è sufficiente per il modello considerato.

□

2.39 Esempio (Modello normale). Consideriamo ora il modello $\mathrm{N}(\theta_1, \theta_2)$ (entrambi i parametri incogniti) Abbiamo mostrato in precedenza che

$$\sum (x_i - \theta_1)^2 = n\widehat{\sigma}^2 + n(\overline{x}_n - \theta_1)^2,$$

dove $\widehat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \overline{x}_n)^2$. In questo caso la funzione di verosimiglianza è

$$\begin{aligned} L(\theta_1, \theta_2; \mathbf{x}_n) &= \left(\frac{1}{\sqrt{2\pi\theta_2}}\right)^n \exp\{-\frac{n\ \widehat{\sigma}_n}{2\ \theta_2}\} \cdot \exp\{-\frac{n}{2\ \theta_2}(\overline{x}_n - \theta_1)^2\} \cdot \prod_{i=1}^n I_{\mathbb{R}}(x_i) \\ &\propto \frac{1}{(\theta_2)^{\frac{n}{2}}} \cdot \exp\{-\frac{n\ \widehat{\sigma}_n}{2\ \theta_2}\} \cdot \exp\{-\frac{n}{2\ \theta_2}(\overline{x}_n - \theta_1)^2\}. \end{aligned}$$

Nell'espressione precedente possiamo ignorare il fattore $h(\mathbf{x}_n) = (\sqrt{2\pi})^{-n}\prod_{i=1}^n I_{\mathbb{R}}(x_i)$ che non dipende dai parametri incogniti. L'espressione restante dipende dai dati attraverso la coppia $(\overline{x}_n, \widehat{\sigma}_n^2)$, che quindi è una statistica sufficiente per il modello. La dimensione della statistica sufficiente vettoriale coincide con quella del parametro incognito, (θ_1, θ_2). □

Statistiche sufficienti in modelli con supporto dipendente dal parametro

Nei modelli con supporto dipendente dal parametro incognito, il fattore contenente il prodotto delle funzioni indicatrici $\prod_{i=1}^n I_{\mathcal{X}_\theta}(x_i)$ che appare in $f_n(\mathbf{x}_n;\theta)$ non può essere trascurato quando si considera la funzione di verosimiglianza, ed entra quindi come fattore nella funzione $g(t;\theta)$.

2.40 Esempio (**Modello uniforme**). Per il modello uniforme nell'intervallo $[0,\theta]$ si ha che

$$L(\theta;\mathbf{x}_n) = f_n(\mathbf{x}_n;\theta) = \frac{1}{\theta^n}\prod_{i=1}^n I_{[0,\theta]}(x_i) = \frac{1}{\theta^n} I_{[x_{(n)},+\infty)}(\theta).$$

Il massimo valore del campione $x_{(n)}$ è quindi una statistica sufficiente. □

Nell'Esempio 2.39 abbiamo osservato che, per il modello normale con entrambi i parametri incogniti, la dimensione della statistica sufficiente vettoriale coincide con il numero di parametri incogniti. Tuttavia, in generale, la dimensione della statistica sufficiente non coincide necessariamente con quella del vettore $\boldsymbol{\theta}$.

2.41 Esempio (**Modello uniforme**). Supponiamo di estrarre un campione di ampiezza n da una popolazione uniforme sull'intervallo $[\theta-1, \theta+1]$, $\theta \in \mathbb{R}$. La funzione di verosimiglianza associata al campione è

$$L(\theta;\mathbf{x}_n) = \frac{1}{2^n}\prod_{i=1}^n I_{[\theta-1,\theta+1]}(x_i) = \frac{1}{2^n} I_{[x_{(n)}-1,x_{(1)}+1]}(\theta).$$

In questo caso $h(\mathbf{x}_n) = 1/2^n$ e $g(T(\mathbf{x}_n);\theta)) = I_{[x_{(n)}-1,x_{(1)}+1]}(\theta)$ dipende dalle osservazioni campionarie attraverso la coppia di valori $(x_{(1)}, x_{(n)})$, che quindi costituisce una statistica sufficiente bidimensionale anche se il parametro incognito è uno solo. □

Già dai primi esempi considerati possiamo trarre tre considerazioni rilevanti:

(i) la statistica sufficiente per un modello *non è unica* (possono esistere infatti statistiche diverse ma del tutto equivalenti in termini informativi su θ);

(ii) esistono statistiche che sono sufficienti ma non essenziali (ovvero contenenti informazioni superflue);

(iii) due distinte statistiche sufficienti possono sintetizzare in misura diversa le informazioni campionarie.

Illustriamo questi tre aspetti con un esempio.

2.42 Esempio (Modello di Poisson). Per il modello di Poisson è semplice verificare che, oltre alla statistica $T(\mathbf{x}_n) = \sum_{i=1}^n x_i$, sono sufficienti anche (e non solo) le statistiche $T_1(\mathbf{x}_n) = \overline{x}_n$, $T_2(\mathbf{x}_n) = (\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2)$ e la statistica $T_3(\mathbf{x}_n) = (x_{(1)}, x_{(2)}, \ldots, x_{(n)})$ (vettore contenente le osservazioni ordinate in senso crescente, detto *campione ordinato*).

(i) La statistica T_1 è del tutto equivalente alla statistica T ai fini della ricostruzione di $L(\theta;\mathbf{x}_n)$. Più in generale, ogni funzione biunivoca di una statistica sufficiente è ancora una statistica sufficiente.

(ii) La statistica T_2 è certamente sufficiente, ma contiene una componente inessenziale ($\sum x_i^2$) per la ricostruzione della fdv. Si noti che, per ogni campione osservato $\mathbf{x}_n$, il valore della statistica $T_1(\mathbf{x}_n)$ si può ottenere come funzione di $T_2(\mathbf{x}_n)$, ma che non è vero il viceversa.

(iii) La statistica T_3 è sufficiente per il modello, ma contiene informazioni inutili rispetto alla determinazione della funzione di verosimiglianza. Infatti $\sum_{i=1}^n x_{(i)} = \sum_{i=1}^n x_i$: la conoscenza degli n valori del campione ordinato consentono quindi di ricostruire la fdv, ma la conoscenza dei singoli valori $x_{(i)}, i = 1, \ldots, n$ è del tutto superflua al tale scopo.

La statistica T, che sintetizza in un numero naturale l'informazione dell'intero campione casuale, riassume quindi in modo più efficiente l'informazione campionaria su θ di quanto faccia T_3, che è invece un vettore di n numeri. L'aspetto cruciale consiste anche in questo caso nel fatto che la statistica T può essere ottenuta come funzione di T_3, ma non viceversa. □

2.8.1 Statistiche sufficienti minimali

Tra tutte le statistiche sufficienti di un modello, siamo interessati ad individuare quelle che forniscono la massima riduzione possibile rispetto al campione osservato.

2.43 Definizione (Sufficienza minimale I). Dato un modello statistico $(\mathcal{X}^n, f_n(\mathbf{x}_n; \theta), \theta \in \Theta)$, una statistica sufficiente T si dice *minimale* se, per ogni altra statistica sufficiente T' per il modello, T è una funzione di T'. □

Questa definizione formalizza quanto osservato nel caso specifico del modello di Poisson. La comprensione di questa definizione necessita tuttavia un approfondimento che riguarda il concetto di partizione dello spazio dei campioni indotto da una statistica. Ricordiamo che una statistica T è per definizione un'applicazione da $\mathcal{X}^n$ (spazio campionario) in $\mathcal{T}$ (insieme dei valori che può assumere T). Indichiamo con A_t l'insieme degli elementi di $\mathcal{X}^n$ in corrispondenza dei quali la statistica T assume il valore $t \in \mathcal{T}$:

$$A_t = \{\mathbf{x}_n \in \mathcal{X}^n : \ T(\mathbf{x}_n) = t\}.$$

Poichè per ogni coppia $(t, t') \in \mathcal{T} : \ t \neq t'$

$$A_t \cap A_{t'} = \emptyset, \quad \text{e} \quad \bigcup_{t \in \mathcal{T}} A_t = \mathcal{X}^n,$$

possiamo affermare che la statistica T induce una partizione dello spazio campionario, costituito dagli insiemi $A_t, \ t \in \mathcal{T}$.

Se T è una statistica sufficiente, la partizione indotta da T si chiama *partizione sufficiente*. Dato un valore t che può assumere T, tutti i campioni che appartengono al medesimo insieme A_t della partizione sufficiente, danno luogo alla stessa funzione di verosimiglianza. Infatti, fissato un valore t in $\mathcal{T}$,, per tutti i campioni $\mathbf{x}_n \in A_t$ si ha che

$$L(\theta; \mathbf{x}_n) \propto g(\theta, T(\mathbf{x}_n)) = g(\theta; t).$$

Se consideriamo due statistiche sufficienti T_1 e T_2 non in corrispondenza biunivoca, indicati con $\mathcal{T}_1$ e $\mathcal{T}_2$ gli insiemi dei valori che possono assumere, ad esse corrispondono due distinte partizioni sufficienti:

$$\{A_{t_i}, t_i \in \mathcal{T}_i\}, \qquad \text{dove} \qquad A_{t_i} = \{\mathbf{x}_n \in \mathcal{X}^n : T(\mathbf{x}_n) = t_i\}, \qquad i = 1, 2.$$

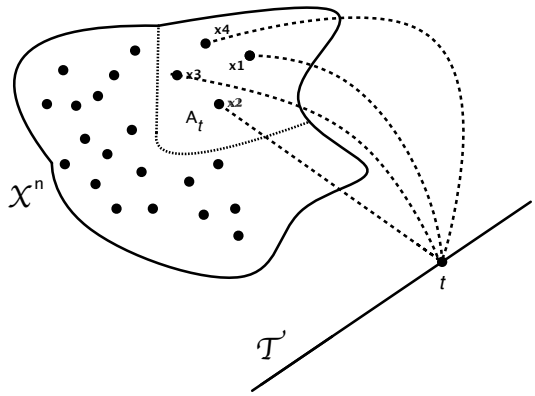

Figura 2.7: *Insieme A_t della partizione dello spazio dei campioni $\mathcal{X}^n$ indotta da una statistica $T(\cdot)$.*

Dire che T_1 è funzione di T_2 vuol dire che, per ogni generica coppia di campioni $\mathbf{x}_n$ e $\mathbf{y}_n$, il fatto che $T_2(\mathbf{x}_n) = T_2(\mathbf{y}_n)$ implica che $T_1(\mathbf{x}_n) = T_1(\mathbf{y}_n)$ (e non viceversa). Possiamo, ad esempio, ottenere il valore di $t_1 = \sum_{i=1}^n x_i$ conoscendo i valori $t_2 = (x_{(1)}), \ldots, x_{(n)})$ ma non possiamo ricostruire i valori del campione ordinato, t_2, partendo dalla conoscenza di t_1 (e infatti possiamo ottenere $\sum_{i=1}^n x_i$ facendo la somma di $[x_{(1)}, \ldots, x_{(n)}]$ ma non possiamo fare il viceversa). In termini di partizioni, ciò vuol dire che ogni insieme A_{t_2} della partizione indotta da T_2 è contenuto in un insieme A_{t_1} della partizione indotta da T_1: la partizione indotta da T_1 risulta quindi meno fine di quella indotta da T_2. Da quanto detto, discende quindi che la partizione indotta da una statistica sufficiente minimale è la partizione *meno fine* tra tutte le partizioni sufficienti per un dato modello.

Le precedenti considerazioni implicano che la *sufficienza* di una statistica T richiede che la funzione di verosimiglianza sia la stessa per *tutti* i campioni appartenenti ad un elemento A_t della partizione che induce; la *minimalità* richiede che ciò si verifichi per *tutti e i soli* campioni di A_t. Questo vuol dire che, se T è sufficiente e minimale, a campioni appartenenti ad insiemi distinti della partizione indotta da T corrispondono funzioni di verosimiglianza non equivalenti. Le precedenti considerazioni si riassumono nella seguente definizione, che esplicita la precedente:

2.44 Definizione (Sufficienza minimale II). Dato un modello statistico $(\mathcal{X}^n, f_n(\mathbf{x}_n; \theta), \theta \in \Theta)$, una statistica sufficiente T si dice *minimale* se la funzione di verosimiglianza è la stessa, a meno di fattori costanti, per tutti e i soli campioni $\mathbf{x}_n$ tali che $T(\mathbf{x}_n) = t$, per ogni $t \in \mathcal{T}$. □

2.45 Osservazione (Partizione di verosimiglianza). Anche la funzione di verosimiglianza, benchè non sia una statistica (in quanto dipende da θ), induce una partizione dello spazio dei campioni, detta *partizione di verosimiglianza*. In ciascun elemento della partizione di verosimiglianza si trovano tutti e i soli campioni che danno luogo a verosimiglianze equivalenti (ovvero, proporzionali tra loro). Una statistica sufficiente è quindi minimale se la partizione che induce coincide con la partizione di verosimiglianza. □

2.46 Osservazione (Unicità sostanziale della statistica sufficiente minimale). Dalla precedente osservazione discende che tutte le statistiche sufficienti minimali inducono la stessa partizione, che coincide con quella di verosimiglianza. Le statistiche sufficienti minimali sono quindi legate tra loro da relazioni biunivoche. Possiamo allora affermare che

le statistiche sufficienti minimali costituiscono una classe di equivalenza, ovvero una classe di funzioni diverse dei dati (come, ad esempio, somma e media campionaria nel modello di Poisson e ogni loro funzione biunivova) alle quali però corrisponde la stessa partizione dello spazio dei campioni. In questo senso possiamo parlare di unicità *sostanziale* della statistica sufficiente minimale. □

2.47 Esempio (Modello di Poisson). Consideriamo il modello di Poisson e le statistiche sufficienti $T_1(\mathbf{x}_n) = \sum_{i=1}^n x_i$ e $T_2(\mathbf{x}_n) = (x_{(1)}, \ldots, x_{(n)})$. Risulta evidente che T_1 è funzione di T_2: presi due campioni distinti di dimensione n che danno luogo alla stesso campione ordinato, questi danno luogo anche alla stessa somma ($T_2(\mathbf{x}_n) = T_2(\mathbf{y}_n) \Rightarrow T_1(\mathbf{x}_n) = T_2(\mathbf{y}_n)$ e non viceversa). La partizione indotta da T_1 è quindi meno fine di quella indotta da T_2. Infatti, se ad esempio per $n = 3$ consideriamo i due campioni $(5, 3, 2)$ e $(1, 7, 2)$, questi appartengono a due insiemi A_{t_2} distinti della partizione indotta da T_2 (perchè danno luogo a campioni ordinati distinti), ma sono nello stesso insieme A_{t_1} della partizione indotta da T_1 (poichè producono lo stesso valore della somma campionaria). Poichè però i due campioni danno luogo alla stessa verosimiglianza, la statistica T_2 non è minimale, dal momento che nella partizione sufficiente che induce, campioni distinti ma equivalenti (a cui è associata una stessa funzione di verosimiglianza) appartengono a elementi distinti della partizione. La statistica sufficiente T_1 è invece minimale poichè a campioni che appartengono a insiemi distinti della partizione indotta da T_1 corrispondono funzioni di verosimiglianza di θ non equivalenti. □

Le considerazioni precedenti si concretizzano in un criterio operativo di verifica di sufficienza minimale di una statistica (per la dimostrazione del quale si rimanda a Casella-Berger (2002, p. 281)).

2.48 Teorema (Criterio di Lehmann-Scheffé). Dato un modello statistico $(\mathcal{X}^n, f_n(\mathbf{x}_n; \theta), \theta \in \Theta)$, supponiamo che esista una statistica T tale che, per qualsiasi coppia di campioni $\mathbf{x}_n$ e $\mathbf{y}_n$, il rapporto delle verosimiglianze $L(\theta; \mathbf{x}_n)/L(\theta; \mathbf{y}_n)$ è costante rispetto a θ se e solo se $T(\mathbf{x}_n) = T(\mathbf{y}_n)$. Allora T è una statistica sufficiente minimale per il modello. □

2.49 Esempio (Modello normale). Consideriamo il modello $\mathrm{N}(\theta, 1)$ (varianza nota). Il rapporto delle verosimiglianze per due campioni $\mathbf{x}_n$ e $\mathbf{y}_n$ è

$$\frac{L(\theta; \mathbf{x}_n)}{L(\theta; \mathbf{y}_n)} = \frac{\exp\{-n(\theta - \overline{x}_n)^2/2\}}{\exp\{-n(\theta - \overline{y}_n)^2/2\}}$$

Se assumiamo che il rapporto è costante rispetto a θ (ovvero se i due campioni danno luogo a stessa verosimiglianza), necessariamente $\overline{x}_n = \overline{y}_n$ (e i due campioni $\mathbf{x}_n$ e $\mathbf{y}_n$ appartengono allo stesso insieme della partizione indotta dalla statistica media campionaria). Se invece $\overline{x}_n = \overline{y}_n$ (i due campioni $\mathbf{x}_n$ e $\mathbf{y}_n$ appartengono allo stesso insieme della partizione indotta dalla statistica media campionaria), il rapporto $L(\theta; \mathbf{x}_n)/L(\theta; \mathbf{y}_n)$ non dipende da θ (i due campioni producono la stessa verosimiglianza). La media campionaria soddisfa la condizione necessaria e sufficiente del criterio di Lehmann-Scheffé ed è quindi una statistica sufficiente minimale per il modello considerato. □

2.50 Esempio (Modello uniforme). In questo caso, dati due campioni $\mathbf{x}_n$ e $\mathbf{y}_n$, si ha che

$$\frac{L(\theta; \mathbf{x}_n)}{L(\theta; \mathbf{y}_n)} = \frac{I_{[x_{(n)}, +\infty)}(\theta)}{I_{[y_{(n)}, +\infty)}(\theta)}.$$

Questo rapporto non dipende da θ se e solo se $x_{(n)} = y_{(n)}$. La statistica *massimo campionario* è quindi sufficiente minimale per il modello uniforme in $[0, \theta]$. □

2.51 Osservazione. La statistica sufficiente T minimale di un modello non è unica: ogni statistica T', funzione biunivoca di T è infatti sufficiente minimale, dal momento che alle due statistiche corrisponde la stessa partizione (minimale) di $\mathcal{X}^n$. □

Definizione frequentista di sufficienza

La definizione del concetto di sufficienza adottata in questo testo è conseguenza del ruolo centrale assegnato alla funzione di verosimiglianza. Nella maggior parte dei testi, la definizione che noi abbiamo utilizzato costituisce un criterio pratico per individuare statistiche sufficienti (da cui il nome di Criterio di fattorizzazione). La seguente caratterizzazione della sufficienza (la cui dimostrazione è tratta da veda Piccinato (2009), pp. 136-7) è in genere utilizzata come definizione.

2.52 Teorema. Dato un modello statistico $(\mathcal{X}^n, f_n(\mathbf{x}_n;\theta), \theta \in \Theta)$, la statistica T è sufficiente se e solo se la distribuzione di probabilità di $\mathbf{X}_n$ condizionata a $T = t$ (per qualunque $t \in \mathcal{T}$) non dipende da θ. □

Dimostrazione. Per semplicità consideriamo il caso in cui $f_T(\cdot;\theta) = \mathbb{P}_\theta(\cdot)$ sia una funzione di massa di probabilità (ovvero T una v.a. discreta). Per dimostrare il teorema dobbiamo mostrare che valgono le due implicazioni.

a) Supponiamo che T sia sufficiente. Per il criterio di fattorizzazione esistono allora due funzioni h e g tali che

$$\mathbb{P}_\theta(\mathbf{X}_n = \mathbf{x}_n) = h(\mathbf{x}_n) \cdot g(\theta, t),$$

dove $t = T(\mathbf{x}_n)$. Si ha quindi che (per la definizione di probabilità condizionata)

$$\mathbb{P}_\theta(\mathbf{X}_n = \mathbf{x}_n | T = t) = \frac{\mathbb{P}_\theta(\mathbf{X}_n = \mathbf{x}_n, T = t)}{\mathbb{P}_\theta(T = t)} = \frac{\mathbb{P}_\theta(\mathbf{X}_n = \mathbf{x}_n)}{\mathbb{P}_\theta(T = t)},$$

dove l'ultima uguaglianza si ottiene osservando che l'evento $(\mathbf{X}_n = \mathbf{x}_n)$ è contenuto nell'evento $(T = t)$ e quindi che l'intersezione dei due eventi coincide con l'evento $(\mathbf{X}_n = \mathbf{x}_n)$. Osserviamo che il denominatore dell'ultima espressione può essere riscritto come

$$\mathbb{P}_\theta(T = t) = \sum_{\mathbf{x}_n : T(\mathbf{x}_n) = t} h(\mathbf{x}_n) g(\theta, t) = g(\theta, t) \sum_{\mathbf{x}_n : T(\mathbf{x}_n) = t} h(\mathbf{x}_n) = g(\theta, t) \overline{h}(\mathbf{x}_n).$$

Sostituendo in (2.7) si ha quindi

$$\mathbb{P}_\theta(\mathbf{X}_n = \mathbf{x}_n | T = t) = \frac{h(\mathbf{x}_n) \cdot g(\theta, t)}{\overline{h}(\mathbf{x}_n) \cdot g(\theta, t)} = \frac{h(\mathbf{x}_n)}{\overline{h}(\mathbf{x}_n)}$$

che non dipende da θ.

b) Supponiamo ora invece che $\mathbb{P}_\theta(\mathbf{X}_n = \mathbf{x}_n | T = t)$ non dipenda da θ, ovvero che sia funzione dei soli dati campionari. Indicando con $h(\mathbf{x}_n)$ tale funzione e ricorrendo alla definizione di probabilità condizionata abbiamo che

$$\mathbb{P}_\theta(\mathbf{X}_n = \mathbf{x}_n) = \mathbb{P}_\theta(\mathbf{X}_n = \mathbf{x}_n | T = t) \cdot \mathbb{P}_\theta(T = t) = h(\mathbf{x}_n) \cdot \mathbb{P}_\theta(T = t),$$

ovvero che $\mathbb{P}_\theta(\mathbf{X}_n = \mathbf{x}_n)$ può essere fattorizzata nel prodotto di una funzione dei soli dati, $h(\mathbf{x}_n)$, e una funzione del parametro, $\mathbb{P}_\theta(T = t)$, che dipende dai dati solo attraverso la statistica $T(\mathbf{x}_n)$.

L'idea intuitiva di questo risultato (o di questa definizione) è che il campione fornisce informazioni su θ grazie al fatto che la sua distribuzione di probabilità dipende da θ. Se il

condizionamento a $T = t$ cancella il legame tra campione e parametro, ciò vuol dire che tutta l'informazione su θ offerta dal campione è riassunta dalla conoscenza di $T = t$.

2.53 Osservazione. Ricordando che i campioni diversi che danno luogo allo stesso valore t di una statistica T si trovano tutti nello stesso insieme A_t della partizione, il risultato precedente implica che la probabilità dei campioni che appartengono allo stesso A_t, subordinatamente al fatto di sapere di limitarsi ai campioni di tale insieme, non dipende da θ. □

2.54 Esempio (Modello di Poisson). Riprendendo ancora il modello di Poisson, mostriamo che $\mathbb{P}_\theta\left(\mathbf{X}_n = \mathbf{x}_n \middle| \sum X_i = t\right)$ non dipende da θ. Si osservi innanzitutto che

$$\mathbb{P}_\theta\left(\mathbf{X}_n = \mathbf{x}_n \middle| \sum X_i = t\right) = \frac{\mathbb{P}_\theta\left(\mathbf{X}_n = \mathbf{x}_n, \sum X_i = t\right)}{\mathbb{P}_\theta\left(\sum X_i = t\right)} = \frac{\mathbb{P}_\theta\left(\mathbf{X}_n = \mathbf{x}_n\right)}{\mathbb{P}_\theta\left(\sum X_i = t\right)},$$

dove l'ultima uguaglianza deriva dal fatto che l'evento $(\mathbf{X}_n = \mathbf{x}_n)$ è contenuto nell'evento $(\sum_{i=1}^n X_i = t)$. Osservando che

$$\mathbb{P}_\theta(\mathbf{X}_n = \mathbf{x}_n) = f_n(\mathbf{x}_n; \theta) = \frac{1}{x_! \dots x_n!} e^{-n\theta} \theta^{\sum_{i=1}^n x_i}$$

e che, ricordando che la somma di n v.a. di Poisson indipendenti ha distruzione di Poisson di parametro $n\theta$,

$$\mathbb{P}_\theta\left(\sum X_i = t\right) = \frac{1}{t!} e^{-n\theta} (n\theta)^t,$$

si ottiene che

$$\mathbb{P}_\theta\left(\mathbf{X}_n = \mathbf{x}_n \middle| \sum_{i=1}^n X_i = t\right) = \frac{t!}{n^t x_1! \dots x_n!},$$

che è indipendente da θ. □

2.8.2 Statistiche sufficienti e famiglie esponenziali

Per un generico modello che forma una famiglia esponenziale è particolarmente semplice individuare una statistica sufficiente minimale. Nel caso di famiglia esponenziale uniparametrica, si è visto nella (1.4) che

$$L(\theta; \mathbf{x}_n) = f_n(\mathbf{x}_n; \theta) = h_n(\mathbf{x}_n) \exp\{\eta(\theta) T_n(\mathbf{x}_n) - B_n(\theta)\},$$

e quindi, per il criterio di fattorizzazione, la statistica $T_n(\mathbf{x}_n) = \sum_{i=1}^n T(x_i)$, è sufficiente per il modello considerato. È semplice verificare che, in base al criterio di Lehmann-Scheffé, la statistica $\sum_{i=1}^n T(x_i)$ è anche minimale. Infatti, dati due campioni $\mathbf{x}_n$ e $\mathbf{y}_n$, si ha che:

(i) se $T_n(\mathbf{x}_n) = T_n(\mathbf{y}_n)$, allora $L(\theta; \mathbf{x}_n)/L(\theta; \mathbf{y}_n) = h_n(\mathbf{x}_n)/h_n(\mathbf{y}_n)$, che è costante rispetto a θ;

(ii) se $L(\theta; \mathbf{x}_n)/L(\theta; \mathbf{y}_n)$ è costante rispetto a θ, necessariamente si deve avere che $T_n(\mathbf{x}_n) = T_n(\mathbf{y}_n)$.

Il risultato si estende al caso di modelli k-parametrici ($\Theta \subset \mathbb{R}^k, k > 1$) che sono famiglie esponenziali. Ricordiamo che, in questo caso, la fdv è

$$L(\boldsymbol{\theta}; \mathbf{x}_n) = f_n(\mathbf{x}_n; \boldsymbol{\theta}) = h_n(\mathbf{x}_n) \exp\left\{\sum_{j=1}^k \eta_j(\boldsymbol{\theta}) T_j(\mathbf{x}_n) - B_n(\boldsymbol{\theta})\right\}.$$

Per il criterio di fattorizzazione, il vettore

$$\mathbf{T}(\mathbf{x}_n) = (T_1(\mathbf{x}_n), \ldots, T_j(\mathbf{x}_n), \ldots, T_k(\mathbf{x}_n)),$$

in cui $T_j(\mathbf{x}_n) = \sum_{i=1}^n T_j(x_i)$, $j = 1, \ldots, k$, è una statistica sufficiente per il modello corrispondente. La minimalità si verifica in modo analogo al caso $k = 1$ con il criterio di Lehmann-Scheffé.

2.55 Esempio Ricordando che, per i modelli che seguono, le leggi di probabilità sono famiglie esponenziali, possiamo affermare che:

- per il modello Bernuolliano, di Poisson, esponenziale negativo e normale con varianza nota, la statistica $T(\mathbf{x}_n) = \sum_{i=1}^n x_i$ è una statistica sufficiente minimale;
- per il modello normale con valore atteso noto, μ, la statistica $T(\mathbf{x}_n) = \sum_{i=1}^n (x_i - \mu)^2$ è sufficiente minimale;
- per il modello $\mathrm{N}(\theta_1, \theta_2)$ il vettore $\mathbf{T}(\mathbf{x}_n) = \left(\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2\right)$ è sufficiente minimale.

□

2.9 Il Principio di verosimiglianza

Il concetto di verosimiglianza, dovuto allo statistico inglese R.A. Fisher (anni 20) formalizza matematicamente, attraverso un sistema di pesi su tutti i valori del parametro, l'informazione che i dati osservati forniscono su un parametro. Questo ruolo è stato formalizzato da A. Birnbaum (1962) nel cosidetto Principio di verosimiglianza (PdV).

> **Principio di verosimiglianza (PdV)**. Siano $\{\mathcal{X}^n, f_n(\cdot;\theta), \theta \in \Theta\}$ e $\{\mathcal{Y}^m, p_m(\cdot;\theta), \theta \in \Theta\}$ due modelli statistici in cui lo spazio dei parametri è lo stesso. Se le funzioni di verosimiglianza $L(\theta;\mathbf{x}_n)$ e $L(\theta;\mathbf{y}_m)$ associate a due campioni osservati $\mathbf{x}_n \in \mathcal{X}^n$ e $\mathbf{y}_m \in \mathcal{Y}^m$ sono proporzionali tra loro, ovvero se
>
> $$L(\theta, \mathbf{x}_n) = c \cdot L(\theta;\mathbf{y}_m), \qquad \forall \theta \in \Theta,$$
>
> dove $c > 0$ è una costante rispetto a θ (che può dipendere da $\mathbf{x}_n$ e $\mathbf{y}_m$), i risultati campionari $\mathbf{x}_n$ e $\mathbf{y}_m$ forniscono la stessa informazione sul parametro incognito.

Il PdV formalizza il fatto che la fdv incorpora tutta l'informazione campionaria riguardo al parametro. Afferma infatti che se due campioni distinti, provenienti anche da due modelli diversi (ma che condividono lo spazio dei parametri), danno luogo a verosimiglianze equivalenti, le conclusioni inferenziali su θ che si hanno utilizzando i due campioni devono essere le stesse. Ad esempio, si devono ottenere con i due campioni stesse stime puntuali, intervallari e stesse scelte nei problemi di verifica di ipotesi.

La formulazione data è a volte indicata come versione *forte* del PdV, mentre per la versione *debole* si assume che lo spazio campionario e la legge di probabilità corrispondenti ai due campioni osservati siano gli stessi.

2.56 Esempio Sia θ la probabilità incognita che si realizzi un evento. Consideriamo due v.a X e Y, e supponiamo che $X \sim \mathrm{Binom}(y, \theta)$ (per y intero positivo fissato), ovvero che X sia la v.a. che conta il numero di successi (x) in y prove bernoulliane indipendenti con probabilità di successo pari a θ. Supponiamo inoltre che Y sia la v.a. che conta il numero

(y) di prove bernoulliane (di parametro θ) che si devono effettuare per il numero fissato x di successi. Le funzioni di massa di probabilità [2] sono

$$f_X(x;\theta) = \binom{y}{x}\theta^x(1-\theta)^{y-x} \qquad p_Y(y;\theta) = \binom{y-1}{x-1}\theta^x(1-\theta)^{y-x}.$$

I modelli statistici associati sono

$$\{\mathcal{X}, f_X(x;\theta), \theta \in [0,1]\} \quad \{\mathcal{Y}, p_Y(y;\theta), \theta \in [0,1]\}.$$

in cui

$$\mathcal{X} = \{0,1,2,\ldots,y\} \qquad \text{e} \qquad \mathcal{Y} = (x, x+1, x+2, \ldots).$$

Consideriamo un esperimento associato al primo modello e supponiamo che su $y = 10$ prove si osservino $x = 4$ successi. La fdv associata a $x = 4$ è

$$L_x(\theta;x) = \binom{10}{4}\theta^4(1-\theta)^6.$$

Consideriamo poi un esperimento associato al secondo modello in cui, per arrivare a $x = 4$ successi sono stati necessari $y = 10$ prove bernoulliane. La fdv associata a $y = 10$ in questo caso è

$$L_y(\theta;y) = \binom{9}{3}\theta^4(1-\theta)^6.$$

Si ha quindi che, con questi risultati, $L_x(x;\theta) \propto L_y(y;\theta)$, ovvero che i due esperimenti, condotti in modi diversi, conducono alla stessa fdv. Per il PdV, indipendentemente dal modo con il quale i dati sono stati generati, le conclusioni inferenziali su θ devono essere le stesse. □

Le procedure inferenziali basate sulle sintesi della fdv obbediscono automaticamente al PdV: nell'esempio precedente, stima di massima verosimiglianza, insiemi L_q e rapporti di verosimiglianza sono i medesimi se si utilizzano i risultati campionari osservati ($y = 10, x = 4$), sebbene provenienti da due distinte procedure sperimentali. Non tutte le impostazioni inferenziali rispettano però il PdV. Infatti, come vedremo, nell'impostazione frequentista, la procedura con cui si ottengono i dati non è irrilevante. Si verifica infatti che, nel problema considerato nell'esempio (in cui, lo ricordiamo, le fdv sono equivalenti), è possibile giustificare l'uso di stime di θ distinte a seconda della procedura che ha generato i dati: nel caso del primo esperimento la stima è x/y, qui $4/10$, mentre con il secondo esperimento, la stima è $(x-1)/(y-1)$, qui $3/9$, con evidente violazione del PdV.

[2] La v.a. Y prende il nome di binomiale negativa ($Y \sim \text{BinNeg}(x,\theta)$, per x intero positivo fissato). L'espressione di p_Y si giustifica osservando che, considerando prove bernoulliane indipendenti, per conseguire complessivamente x successi alla prova y si deve avere un successo alla prova y (evento che ha probabilità θ) e averne conseguiti $x-1$ nelle $y-1$ prove precedenti (evento che ha probabilità $\binom{y-1}{x-1}\theta^{x-1}(1-\theta)^{(y-1)-(x-1)}$). Il prodotto delle due probabilità dà $p_Y(y;\theta)$.

Capitolo 3

Inferenza frequentista: statistiche e distribuzioni campionarie

Il capitolo introduce la trattazione dell'impostazione frequentista dell'inferenza statistica, alla quale sono dedicati anche i capitoli 4, 5 e 6. In particolare, questo capitolo si enuncia dapprima il principio fondativo del metodo frequentista, ovvero il *Principio del campionamento ripetuto*; successivamente si introducono i concetti fondamentali di statistica campionaria e distribuzione campionaria, che sono alla base dei metodi inferenziali per affrontare i problemi di stima puntuale, intervallare e di verifica delle ipotesi. A ciascuno di questi viene dedicato uno dei capitoli che seguono. Attenzione particolare è riservata ai risultati relativi alle distribuzioni di alcune statistiche (tra cui media e varianze campionarie) nel caso del campionamento da popolazioni normali. Vengono poi presentati alcuni risultati asintotici sulla distribuzione delle successioni di medie e somme campionarie.

3.1 Principio del campionamento ripetuto

L'approccio frequentista rappresenta il modello teorico inferenziale attualmente dominante. Questo paradigma si impose tra il 1920 e il 1940, in contrapposizione alla logica bayesiana fino ad allora prevalente. Gli elementi di base di tale impostazione di pensiero, alternativa sia all'impostazione basata sulla verosimiglianza che a quella bayesiana, sono già presenti in alcuni lavori di R.A Fisher pubblicati negli anni '20 del secolo scorso. Negli anni '30 i principali contributi a questa impostazione furono dovuti agli statistici J. Neyman e E.S. Pearson. Solo negli anni '40, A. Wald formulò una sistemazione generale del pensiero frequentista, basata sulla teoria delle decisioni statistiche. La logica di base dell'impostazione che qui esaminiamo è sintetizzata dal cosidetto *Principio del campionamento ripetuto*.

> **Principio del campionamento ripetuto (PCR)**. Le procedure statistiche devono essere valutate in base al loro comportamento in ripetizioni ipotetiche dell'esperimento che genera i dati, supponendo che siano effettuate nelle stesse condizioni.

Si tratta di un principio concettuale (non tecnico-operativo) che necessita di essere chiarito. A tal fine, ripercorriamo gli elementi di un problema inferenziale parametrico. Dato un modello statistico $\{\mathcal{X}^n, f_n(\mathbf{x}_n; \theta), \theta \in \Theta\}$, supponiamo di voler effettuare inferenza sul

parametro incognito θ attraverso i dati campionari $\mathbf{x}_n = (x_1, \ldots, x_n)$, realizzazione della v.a. n-dimensionale $\mathbf{X}_n$. Supponiamo che esista un vero valore $\theta = \theta^*$, incognito, che identifica la legge $f_n(\cdot; \theta^*)$ che ha generato le osservazioni. Per effettuare inferenza sul parametro incognito , si considerano opportune funzioni dei dati osservati (ad esempio per la stima puntuale, mediante intervalli o per la verifica di ipotesi), che indichiamo genericamente con $T(x_1, \ldots, x_n)$. Prima di effettuare l'esperimento che genera i dati, il campione $\mathbf{X}_n = (X_1, \ldots, X_n)$ e tutte le funzioni $T(\mathbf{X}_n) = T(X_1, \ldots, X_n)$ sono oggetti aleatori. Poichè la legge di probabilità $f_n(\cdot; \theta^*)$ di $\mathbf{X}_n$ dipende dal parametro incognito, anche la legge di probabilità associata al generico oggetto aleatorio T dipenderà da θ^*: è proprio la dipendenza di questa legge da θ^* che lega i valori osservati di T al parametro incognito.

Il principio del campionamento ripetuto afferma che le procedure inferenziali devono essere valutate in ipotetiche ripetizioni dell'esperimento che genera i dati. Questo vuol dire, in buona sostanza, che le procedure inferenziali basate sull'uso delle funzioni T devono fornire buoni risultati con elevate probabilità, ovvero che tali procedure devono essere valutate attraverso le caratteristiche delle distribuzioni di probabilità ad esse associate. Ad esempio, se consideriamo il modello $\mathrm{N}(\theta, 1)$, la media $\overline{x}_n$ delle osservazioni è una stima intuitiva di θ^* (e sappiamo anche che è la stima di massima verosimiglianza). Dal punto di vista del PCR, la validità del suo uso dipende dalle caratteristiche della distribuzione di probabilità della v.a. $\overline{X}_n$. In particolare, si guarda alla tendenza che $f_{\overline{X}_n}(\cdot; \theta^*)$ ha di generare valori osservati vicini a θ^*. La v.a. $\overline{X}_n$ è pertanto una buona procedura per stimare θ^* se $f_{\overline{X}_n}(\cdot; \theta^*)$ si concentra intorno al valore θ^*. In questo caso, poichè la v.a. $\overline{X}_n$ ha distribuzione $\mathrm{N}(\theta^*, 1/n)$, abbiamo una garanzia probabilistica che il valore osservato $\overline{x}_n$ sia adeguatamente vicino a θ^* e tale garanzia è tanto più forte quanto più elevata è la numerosità campionaria (che fa decrescere la varianza di $\overline{X}_n$). Quanto detto è vero qualunque sia il vero valore θ^* del parametro.

Secondo il PCR, quindi, una procedura inferenziale viene valutata, tenendo conto del suo comportamento complessivo nello spazio $\mathcal{X}^n$. Deve cioè essere elevata la probabilità (ovvero la *frequenza*) con cui la procedura funziona in modo soddisfacente. Per questo motivo l'impostazione inferenziale ispirata al PCR viene denominata *frequentista*.

Alcune considerazioni sono d'obbligo. La prima è che l'ottica frequentista è di tipo *presperimentale*. Le procedure inferenziali vengono valutate attraverso le loro caratteristiche probabilistiche, e non in corrispondenza di un campione osservato. I dati osservati $\mathbf{x}_n$ vengono quindi utilizzati, una volta scelta una procedura T, per determinare il valore numerico $T(\mathbf{x}_n)$. La seconda osservazione, strettamente legata alla precedente, è che le procedure inferenziali frequentiste e quelle basate sulla funzione di verosimiglianza sono, se non operativamente, almeno concettualmente in contrasto. Il Principio di verosimiglianza, che ispira e motiva le procedure inferenziali basate sulla sintesi della funzione di verosimiglianza (vedi Capitolo 3) assegna un ruolo centrale al campione effettivamente osservato. La giustificazione di queste procedure, in chiaro contrasto con quella dei metodi frequentisti, è quindi di tipo *post-sperimentale*. Va però osservato che, al di là delle contrapposizioni logiche tra le due logiche inferenziali, le procedure che si ottengono attraverso la sintesi della fdv presentano in genere valide proprietà anche dal punto di vista frequentista.

3.2 Definizioni preliminari

Una *statistica campionaria* è una funzione dei dati campionari, ovvero una funzione $T : \mathcal{X}^n \to \mathcal{T}$, con $\mathcal{T} \subseteq \mathbb{R}^k$. Questo vuol dire che la statistica T assume valori reali scalari ($k = 1$) o vettoriali ($k > 1$). Prima di osservare i dati, la statistica T è una funzione della v.a. $X_1, \ldots, X_n$. Pertanto $T(\mathbf{X}_n)$ ha una distribuzione di massa di probabilità (se le X_i sono v.a. discrete) o una funzione di densità (se le X_i sono v.a. assolutamente continue). Precisiamo la definizione di statistica nel modo seguente.

3.1 Definizione (Statistica e distribuzione campionaria). Si chiama *statistica campionaria* una funzione

$$T : \mathcal{X}^n \to \mathcal{T}, \qquad \mathcal{T} \subseteq \mathbb{R}^k, \quad k \geq 1.$$

La funzione di massa di probabilità o di densità di $T(\mathbf{X}_n)$, $f_T(\cdot;\theta)$, si chiama *distribuzione campionaria* della statistica. □

Si noti che la distribuzione campionaria di una statistica dipende (in genere, ma non necessariamente) dal parametro incognito del modello. Nel Capitolo 4 l'attenzione si concentrerà sullo studio delle statistiche utilizzate specificamente allo scopo di ottenere stime puntuali dei parametri, ovvero sulle funzioni T per le quali i valori t in $\mathcal{T}$ sono punti di Θ. In questo capitolo ci soffermiamo invece sullo studio delle proprietà di alcune statistiche campionarie rilevanti, indipendentemente dalla loro funzione nei problemi inferenziali. Come vedremo, molte delle statistiche che considereremo saranno anche, a seconda dei modelli statistici considerati, le funzioni utilizzate per ottenere stime puntuali. La tabella che segue riporta le principali statistiche campionarie univariate.

Statistica	$T(\mathbf{X}_n)$	Funzione
Media campionaria	$\overline{X}_n$	$\frac{1}{n}\sum_{i=1}^n X_i$
Somma campionaria	Y_n	$\sum_{i=1}^n X_i = n\overline{X}_n$
Varianza campionaria	$\widehat{\sigma}^2_n$	$\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2$
Varianza campionaria corretta	S^2_n	$(n-1)^{-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$
Minimo campionario	$X_{(1)}$	$\min\{X_1, \ldots, X_n\}$
Massimo campionario	$X_{(n)}$	$\max\{X_1, \ldots, X_n\}$

Esempi di statistiche bidimensionali sono i vettori

$$\left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2\right) \qquad \text{e} \qquad \left(X_{(1)}, X_{(n)}\right).$$

Un esempio di statistica n-dimensionale è il campione ordinato

$$\mathbf{X}_{(n)} = (X_{(1)}, X_{(2)}, \ldots, X_{(n)}).$$

Anche se la definizione che abbiamo dato di statistica è del tutto generale e non dipende quindi dal processo di acquisizione dei dati, ovvero dalle particolari forme di dipendenza delle v.a. $X_1, \ldots, X_n$, qui ci concentriamo sulle proprietà delle statistiche campionarie nel caso di campioni casuali, in cui le v.a. X_i sono indipendenti e somiglianti. Prima di passare all'esame di alcune proprietà delle statistiche considerate, illustriamo con un esempio concreto cosa si intende per distribuzione campionaria di una statistica.

3.2 Esempio (Distribuzione campionaria di $\overline{X}_n$ nel modello bernoulliano). Consideriamo un campione casuale per un modello bernoulliano. Siamo interessati alla distribuzione campionaria della statistica media campionaria, $\overline{X}_n$. A scopo illustrativo, consideriamo il caso $n = 4$.

X_1	X_2	X_3	X_4	$f_4(\mathbf{x}_n;\theta)$	Valore di $\overline{X}_4$	$f_{\overline{X}_4}(\cdot;\theta)$
0	0	0	0	$(1-\theta)^4$	0	$\binom{4}{0}(1-\theta)^4$
0	0	0	1	$\theta(1-\theta)^3$		
0	0	1	0	$\theta(1-\theta)^3$	$\frac{1}{4}$	$\binom{4}{1}\theta(1-\theta)^3$
0	1	0	0	$\theta(1-\theta)^3$		
1	0	0	0	$\theta(1-\theta)^3$		
0	0	1	1	$\theta^2(1-\theta)^2$		
0	1	0	1	$\theta^2(1-\theta)^2$		
0	1	1	0	$\theta^2(1-\theta)^2$	$\frac{1}{2}$	$\binom{4}{2}\theta^2(1-\theta)^2$
1	0	0	1	$\theta^2(1-\theta)^2$		
1	0	1	0	$\theta^2(1-\theta)^2$		
1	1	0	0	$\theta^2(1-\theta)^2$		
0	1	1	1	$\theta^3(1-\theta)$		
1	0	1	1	$\theta^3(1-\theta)$	$\frac{3}{4}$	$\binom{4}{3}\theta^3(1-\theta)$
1	1	0	1	$\theta^3(1-\theta)$		
1	1	1	0	$\theta^3(1-\theta)$		
1	1	1	1	θ^4	1	$\binom{4}{4}\,\theta^4$

Lo spazio campionario $\mathcal{X}^4 = \{0,1\}^4$ è riportato nella prima colonna della tabella. La seconda colonna riporta i valori della funzione di massa di probabilità $f_4(\cdot;\theta)$ per i 16 campioni che costituiscono $\mathcal{X}^4$. La terza e la quarta colonna riportano i valori che può assumere $\overline{X}_4$ e le probabilità con cui si realizzano. Complessivamente queste due colonne rappresentano la distribuzione campionaria di $\overline{X}_4$. Si noti che la funzione $f_{\overline{X}_4}(\cdot;\theta)$ dipende dal parametro θ. □

Nei paragrafi che seguono, sempre con riferimento a campioni casuali, consideriamo:

1. alcune proprietà del valore atteso e della varianza delle statistiche $\overline{X}_n$, $\widehat{\sigma}^2_n$ e S^2_n che sono valide (nel caso di campioni casuali) indipendentemente dalla distribuzione della v.a. di base X (paragrafo 3.3.1);
2. la distribuzione di $\overline{X}_n$ e $\sum_{i=1}^n X_i$ per alcuni modelli rilevanti, escluso il modello normale (paragrafo 3.3.2);
3. le distribuzioni delle statistiche $\overline{X}_n$, $\widehat{\sigma}^2_n$ e S^2_n nel caso di campionamento da popolazioni normali (paragrafo 3.4);
4. le approssimazioni asintotiche per le distribuzioni di somme e medie campionarie e alcuni risultati collegati (paragrafo 3.5);
5. le distribuzioni delle statistiche minimo e massimo campionario (paragrafo 3.6).

3.3 Somma, media e varianza campionarie

3.3.1 Proprietà generali

Nel caso di campioni casuali, le statistiche $\overline{X}_n$, $\widehat{\sigma}^2_n$ e S^2_n presentano alcune proprietà che risultano valide qualunque sia la distribuzione di probabilità delle v.a. $X_1, \ldots, X_n$.

3.3 Teorema. Sia $\mathbf{X}_n = (X_1, \ldots, X_n)$ un campione casuale proveniente da una popolazione X con valore atteso[1] $\mathbb{E}[X] = \mu$ e varianza $\mathbb{V}[X] = \sigma^2 < \infty$. Per le statistiche campionarie $\overline{X}_n$, $\widehat{\sigma}^2_n$ e S^2_n si ha:

[1] A rigore dovremmo usare i simboli μ_θ e σ^2_θ per indicare che valore atteso e varianza di X sono funzioni del parametro incognito del modello. Si utilizza una notazione più leggera per semplicità.

1. $\mathbb{E}[\overline{X}_n] = \mathbb{E}[X] = \mu, \quad \mathbb{V}[\overline{X}_n] = \frac{1}{n}\mathbb{V}[X] = \frac{\sigma^2}{n}$
   e quindi
   $\mathbb{E}[\overline{X}_n{}^2] = \mathbb{V}[\overline{X}_n] + (\mathbb{E}[\overline{X}_n])^2 = \frac{\sigma^2}{n} + \mu^2$;
2. $\mathbb{E}[\widehat{\sigma}_n^2] = \frac{n-1}{n}\mathbb{V}[X] = \frac{n-1}{n}\sigma^2$;
3. $\mathbb{E}[S_n^2] = \mathbb{V}[X] = \sigma^2$.

□

Dimostrazione.

1. Per il valore atteso di $\overline{X}_n$

$$\mathbb{E}[\,\overline{X}_n\,] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\,\mathbb{E}\,[\sum_{i=1}^{n} X_i] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\,[\,X_i\,] = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu,$$

   dove la seconda uguaglianza è dovuta alla proprietà di omogeneità dell'operatore $\mathbb{E}$, la terza alla sua linearità e la quarta al fatto che le v.a. X_i sono somiglianti. Per la varianza di $\overline{X}_n$ si ha

$$\mathbb{V}[\,\overline{X}_n\,] = \mathbb{V}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\,\mathbb{V}\,[\sum_{i=1}^{n} X_i] = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}[X_i] = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{\sigma^2}{n},$$

   dove la seconda uguaglianza si ottiene per le proprietà dell'operatore $\mathbb{V}$, la terza per l'indipendenza delle X_i e la quarta per la somiglianza delle X_i.

2. Si noti per prima cosa che:

$$\begin{aligned} n\,\widehat{\sigma}_n^2 = (n-1)S_n^2 &= \sum_{i=1}^{n}(X_i - \overline{X}_n)^2 = \sum_{i=1}^{n}\big(X_i^2 + \overline{X}_n{}^2 - 2X_i\,\overline{X}_n\big) \\ &= \sum_{i=1}^{n} X_i^2 + n\overline{X}_n{}^2 - 2\overline{X}_n\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} X_i^2 - n\,\overline{X}_n{}^2. \end{aligned}$$

   Dalle precedenti uguaglianze si ottiene

$$\begin{aligned} \mathbb{E}[\,\widehat{\sigma}_n^2\,] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n{}^2\right] = \frac{1}{n}\,\mathbb{E}[\sum_{i=1}^{n} X_i^2\,] - \mathbb{E}\left[\,\overline{X}_n{}^2\,\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[X_i^2\right] - \mathbb{E}\left[\,\overline{X}_n{}^2\,\right] \\ &= \frac{1}{n}\,n\,\mathbb{E}[\,X^2\,] - \mathbb{E}\left[\,\overline{X}_n{}^2\,\right] = \sigma^2 + \mu^2 - \left[\frac{\sigma^2}{n} + \mu^2\right] = \frac{n-1}{n}\,\sigma^2, \end{aligned}$$

   dove la seconda e la terza uguaglianza si derivano dalle proprietà dell'operatore $\mathbb{E}$, la quarta dalla somiglianza delle X_i e la quinta si ottiene utilizzando due volte la relazione che lega momento primo, momento secondo e varianza di una variabile aleatoria.

3. Osservando che $S_n^2 = n\widehat{\sigma}_n^2/(n-1)$, si ha che

$$\mathbb{E}[S_n^2] = \mathbb{E}\left[\frac{n}{n-1}\widehat{\sigma}_n^2\right] = \frac{n}{n-1}\mathbb{E}[\widehat{\sigma}_n^2] = \frac{n}{n-1}\cdot\frac{n-1}{n}\sigma^2 = \sigma^2.$$

3.4 Osservazione. Dal punto (1) del precedente teorema, osservando che $Y_n = n\overline{X}_n$, discende che

$$\mathbb{E}[Y_n] = n\mathbb{E}[\overline{X}_n] = n\mu \qquad \text{e} \qquad \mathbb{V}[Y_n] = n^2\mathbb{V}[\overline{X}_n] = n^2\frac{\mathbb{V}[X]}{n} = n\sigma^2.$$

Allo stesso risultato si giunge considerando direttamente valore atteso e varianza della somma aleatoria $\sum_{i=1}^n X_i$. □

3.5 Osservazione. Il valore statistico-inferenziale della proprietà (1) del precedente teorema di $\overline{X}_n$ consiste nel fatto che la sua distribuzione campionaria ha lo stesso valore atteso della v.a. di base X, ma la sua varianza è ridotta per un fattore $1/n$, dove n è la dimensione campionaria. Intuitivamente ciò vuol dire che, maggiore è la dimensione campionaria, minore la dispersione della distribuzione di $\overline{X}_n$. □

3.6 Osservazione. La proprietà (3), in base alla quale il valore atteso di S_n^2 è esattamente uguale a $\mathbb{V}_\theta[X]$, giustifica il nome di varianza campionaria *corretta*. □

3.3.2 Distribuzione di somma e media campionaria

Il valore del Teorema (3.3) consiste nel fatto che, nel caso di campioni casuali (variabili aleatorie i.i.d.), è valido per qualsiasi modello statistico in cui valore atteso e varianza sono definiti. Senza assunzioni sulla specifica distribuzione delle v.a. X_i, non possiamo tuttavia fare affermazioni generali sulla distribuzione delle statistiche Y_n e $\overline{X}_n$. Per molti tra i più comuni modelli risulta comunque semplice derivare la distribuzione di queste statistiche campionarie, utilizzando la funzione generatrice dei momenti.

Funzione generatrice dei momenti

3.7 Definizione. Sia X una v.a. con legge di probabilità (funzione di massa o di densità di probabilità) f_X. Si chiama *funzione generatrice dei momenti di* X (fgm) la funzione

$$M_X(t) = \mathbb{E}[e^{tX}],$$

supposto che il valore atteso precedente esista per t appartenente a qualche intorno di $t = 0$. Nel caso di v.a. assolutamente continua con funzione di densità f_X, la fgm è

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)dx;$$

nel caso di v.a. discrete con funzione di massa di probabilità f_X,

$$M_X(t) = \sum_i e^{tx_i} f_X(x_i).$$

□

3.8 Esempio (Modello gamma). Per una v.a. $X \sim \text{Gamma}(\theta_1, \texttt{rate} = \theta_2)$ si ha

$$\begin{aligned} M_X(t) = \mathbb{E}[e^{tx}] &= \frac{\theta_2^{\theta_1}}{\Gamma(\theta_1)} \int_0^\infty e^{tx} x^{\theta_1 - 1} e^{-x\theta_2} dx \\ &= \frac{\theta_2^{\theta_1}}{\Gamma(\theta_1)} \int_0^\infty x^{\theta_1 - 1} e^{-x(\theta_2 - t)} dx \\ &= \frac{\theta_2^{\theta_1}}{\Gamma(\theta_1)} \times \frac{\Gamma(\theta_1)}{(\theta_2 - t)^{\theta_1}} = \left(\frac{\theta_2}{\theta_2 - t}\right)^{\theta_1} \end{aligned}$$

dove la penultima uguaglianza si ottiene ricordando che, per $a > 0$ e $b > 0$ si ha in generale $\int_0^\infty y^{a-1}e^{-by}dy = \Gamma(a)/b^a$. Ricordando che il modello EN(θ) coincide con il modello Ga($1, \texttt{rate} = \theta$), la fgm per la v.a. esponenziale risulta essere $M_X(t) = \theta/(\theta - t),\ t < \theta$. In modo del tutto analogo, se si considera la parametrizzazione `scale`, si verifica che

$$M_X(t) = \left(\frac{1}{1-\theta_2 t}\right)^{\theta_1}$$

□

La tabella seguente riporta la fgm per i principali modelli parametrici.

Distribuzione	$M_X(t)$
Bern(θ)	$\theta e^t + (1-\theta)$
Binom(n, θ)	$[\theta e^t + (1-\theta)]^n$
Pois(θ)	$e^{\theta(e^t-1)}$
N(θ_1, θ_2)	$\exp\{\theta_1 t + \theta_2 t^2/2\}$
Gamma($\theta_1, \texttt{rate} = \theta_2$)	$\left(\frac{\theta_2}{\theta_2 - t}\right)^{\theta_1}\ (t < \theta_2)$
Gamma($\theta_1, \texttt{scale} = \theta_2$)	$\left(\frac{1}{1-\theta_2 t}\right)^{\theta_1}\ (t < 1/\theta_2)$
EN(θ) =Gamma($1, \texttt{rate} = \theta$)	$\frac{\theta}{\theta - t}\ (t < \theta)$
Esp(θ) =Gamma($1, \texttt{scale} = \theta$)	$\frac{1}{1-\theta t}\ (t < 1/\theta)$
Chi quadrato (θ) = Gamma($\frac{\theta}{2}, \texttt{scale} = 2$)	$\left(\frac{1}{1-2t}\right)^{\theta/2}$

La fgm presenta molte utili proprietà, alcune delle quali sono di seguito dimostrate. La prima di queste proprietà giustifica il nome di fgm.

3.9 Teorema. Se una v.a. X possiede momento di ordine r, $\mathbb{E}[X^r]$, questo coincide con la derivata r-esima rispetto a t della funzione $M_X(t)$ calcolata in $t = 0$:

$$\mathbb{E}[X^r] = \frac{d^r}{dt^r} M_X(t)|_{t=0}.$$

□

Dimostrazione. Per dimostrare questo risultato dobbiamo assumere che sia possibile scambiare le operazioni di integrale e derivata (operazione lecita praticamente in tutti i modelli considerati in questo testo). In tal caso si ha:

$$\begin{aligned}
\frac{d^r}{dt^r} M_X(t) &= \frac{d^r}{dt^r} \int_{-\infty}^{\infty} e^{tx} f_X(x) dx \\
&= \int_{-\infty}^{\infty} \frac{d^r}{dt^r} e^{tx} f_X(x) dx \\
&= \int_{-\infty}^{\infty} x^r e^{tx} f_X(x) dx
\end{aligned}$$

In $t = 0$ l'ultimo integrale diventa

$$\int_{-\infty}^{\infty} x^r f_X(x) dx = \mathbb{E}[X^r].$$

Nel caso di campioni casuali, è semplice determinare la funzione generatrice dei momenti di Y_n e di $\overline{X}_n$ in funzione della fgm della v.a. di base, M_X. Vale infatti il seguente risultato.

3.10 Teorema. Sia $X_1, \ldots, X_n$ un campione casuale proveniente da una popolazione in cui la v.a. di base X ha fgm $M_X(t)$. Si ha allora che, per $Y_n = \sum_{i=1}^n X_i$ e $\overline{X}_n = Y_n/n$,

$$M_{Y_n}(t) = [M_X(t)]^n, \quad M_{\overline{X}_n}(t) = \left[M_X\left(\frac{t}{n}\right)\right]^n \quad \text{e} \quad M_{\overline{X}_n}(t) = M_{Y_n}\left(\frac{t}{n}\right).$$

□

Dimostrazione. Per Y_n, tenendo conto del fatto che le v.a. X_i sono indipendenti e tutte con distribuzione uguale a quella di X, si ha

$$M_{Y_n}(t) = \mathbb{E}[e^{tY_n}] = \mathbb{E}[e^{t(X_1+\ldots+X_n)}] = \prod_{i=1}^n \mathbb{E}[e^{tX_i}] = \prod_{i=1}^n M_X(t) = [M_X(t)]^n.$$

Analogamente, per $\overline{X}_n$ si ha

$$M_{\overline{X}_n}(t) = \mathbb{E}[e^{t\overline{X}_n}] = \mathbb{E}[e^{t(X_1+\ldots+X_n)/n}] = \prod_{i=1}^n \mathbb{E}[e^{tX_i/n}] = \prod_{i=1}^n M_X(t/n) = [M_X(t/n)]^n.$$

Da quanto scritto discende anche che $M_{\overline{X}_n}(t) = M_{Y_n}(t/n)$.

Il precedente risultato può essere generalizzato al caso di combinazioni lineari di v.a. indipendenti ma non identicamente distribuite.

3.11 Teorema. Siano $X_1, \ldots, X_n$ n v.a. indipendenti e $a_1, \ldots, a_n$ e b delle costanti reali, e sia $Z = \sum_{i=1}^n a_i X_i + b$. Si ha allora che

$$M_Z(t) = e^{tb} \prod_{i=1}^n M_{X_i}(a_i t).$$

□

Dimostrazione. Per definizione di fgm abbiamo che

$$\begin{aligned} M_Z(t) &= \mathbb{E}[e^{tZ}] = \mathbb{E}\left[e^{t\sum_{i=1}^n a_i X_i + tb}\right] \\ &= e^{tb}\mathbb{E}\left[\prod_{i=1}^n e^{(a_i t)X_i}\right] = e^{tb}\prod_{i=1}^n \mathbb{E}\left[e^{(a_i t)X_i}\right] \\ &= e^{tb}\prod_{i=1}^n M_{X_i}(a_i t), \end{aligned}$$

dove la penultima uguaglianza si giustifica con l'indipendenza delle X_i.

3.12 Osservazione. Si noti che, ponendo $a_i = 1$, $i = 1, \ldots, n$ e $b = 0$ si ottiene M_{Y_n}; ponendo $b = 0$ e $a_i = 1/n$, $i = 1 \ldots, n$, si ottiene $M_{\overline{X}_n}$. Inoltre, dal Teorema (3.11) discende, come caso particolare, anche che

$$M_{aX+b}(t) = e^{tb} M_X(at).$$

□

Il seguente teorema consente di utilizzare la fgm della v.a. X per identificarne la legge di probabilità. (Per la dimostrazione si rimanda, ad esempio a Casella-Berger (2002), Teorema 2.3.11, pp. 65-66).

3.13 Teorema. Siano X e Y due v.a. che assumono valori in $\mathcal{U}$ con funzioni di ripartizione F_X e F_Y e con fgm M_X e M_Y. Se, indicato con I_0 un qualsiasi intorno aperto di $t = 0$,

$$M_X(t) = M_Y(t), \qquad \forall t \in I_0,$$

allora le v.a. X e Y sono uguali in distribuzione ($X \stackrel{d}{=} Y$), ovvero

$$F_X(u) = F_Y(u), \qquad \forall u \in \mathcal{U}.$$

□

3.14 Esempio (Modello normale). Sia $X_1, \ldots, X_n$ un campione casuale proveniente da una popolazione $\mathrm{N}(\theta_1, \theta_2)$. La fgm di ciascuna X_i è quindi $M_{X_i}(t) = \exp\{\theta_1 t + \theta_2 t^2/2\}$. Per la v.a. $\overline{X}_n$ si ha quindi che

$$\begin{aligned} M_{\overline{X}_n}(t) = [M_{X_i}(t/n)]^n &= \left[\exp\left(\theta_1 \frac{t}{n} + \frac{\theta_2 (t/n)^2}{2}\right)\right]^n \\ &= \exp\left(n\left(\theta_1 \frac{t}{n} + \frac{\theta_2 (t/n)^2}{2}\right)\right) = \exp\left(\theta_1 t + \frac{(\theta_2/n) t^2}{2}\right). \end{aligned}$$

L'ultima espressione trovata è la fgm di una v.a. $\mathrm{N}(\theta_1, \theta_2/n)$ e quindi, per il Teorema (3.13) possiamo affermare che $\overline{X}_n$ ha distribuzione $\mathrm{N}(\theta_1, \theta_2/n)$. Analogamente si dimostra che $M_{Y_n}(t) = \exp\{n\theta_1 t + n\theta_2 t^2/2\}$ e quindi che $Y_n \sim \mathrm{N}(n\theta_1, n\theta_2)$.

□

3.15 Esempio (Modello normale). Se $X_i \sim \mathrm{N}(\theta_1, \theta_2)$ i.i.d. e $Z = \sum_{i=1}^n a_i X_i + b$, per il Teorema 3.11 si ha che

$$\begin{aligned} M_Z(t) &= e^{tb} \prod_{i=1}^n \exp\{(\theta_1 a_i) t + (\theta_2 a_i^2) t^2/2\} \\ &= \exp\{(\theta_1 \sum a_i + b) t + (\theta_2 \sum a_i^2) t^2/2\} \end{aligned}$$

e quindi, per il Teorema 3.13, $Z \sim N\left(\theta_1 \sum a_i + b, \theta_2 \sum a_i^2\right)$. □

3.16 Esempio (Distribuzione di $\overline{X}_n$ e Y_n per il modello gamma). Sia $X_1, \ldots, X_n$ un campione casuale proveniente da una popolazione $\mathrm{Gamma}(\theta_1, \texttt{rate} = \theta_2)$. La fgm di ciascuna X_i è quindi $M_{X_i}(t) = [\theta_2/(\theta_2 - t)]^{\theta_1}$. Per la v.a. $\overline{X}_n$ si ha quindi che

$$M_{\overline{X}_n}(t) = [M_{X_i}(t/n)]^n = \left[\left(\frac{\theta_2}{\theta_2 - t/n}\right)^{\theta_1}\right]^n = \left(\frac{n\theta_2}{n\theta_2 - t}\right)^{n\theta_1}.$$

L'ultima espressione coincide con la fgm di una v.a. $\mathrm{Gamma}(n\theta_1, \texttt{rate} = n\theta_2)$. Analogamente si dimostra che $M_{Y_n}(t) = [\theta_2/(\theta_2 - t)]^{n\theta_1}$ e quindi che $Y_n \sim \mathrm{Gamma}(n\theta_1, \texttt{rate} = \theta_2)$. È anche semplice verificare che se le v.a. X_i sono indipendenti ma non identicamente distribuite, con $X_i \sim \mathrm{Gamma}(\theta_{1_i}, \texttt{rate} = \theta_2)$, $i = 1, \ldots, n$, allora $M_{Y_n} = [\theta_2/(\theta_2 - t)]^{\sum \theta_{1_i}}$, e $Y_n \sim \mathrm{Gamma}(\sum \theta_{1_i}, \texttt{rate} = \theta_2)$.
In modo del tutto analogo si verifica che, se le X_i sono v.a. $\mathrm{Gamma}(\theta_1 = \texttt{scale} = \theta_2)$ i.i.d, allora $Y_n|\boldsymbol{\theta} \sim \mathrm{Gamma}(n\theta_1, \texttt{scale} = \theta_2)$ e $\overline{X}_n|\boldsymbol{\theta} \sim \mathrm{Gamma}(n\theta_1, \texttt{scale} = \theta_2/n)$. Di conseguenza, se $X_i|\theta \sim \mathrm{Esp}(\theta) = \mathrm{Gamma}(1, \texttt{scale} = \theta)$, allora $Y_n|\boldsymbol{\theta} \sim \mathrm{Gamma}(n, \texttt{scale} = \theta_2)$ e $\overline{X}_n|\boldsymbol{\theta} \sim \mathrm{Gamma}(n, \texttt{scale} = \theta_2/n)$.

□

La seguente tabella riporta le fgm della somma campionaria, Y_n, e le corrispondenti distribuzioni di probabilità per le principali variabili aleatorie.

X_i	$M_{Y_n}(t)$	Y_n
Bern(θ)	$[\theta e^t + (1-\theta)]^n$	Binom(n, θ)
Pois(θ)	$e^{n\theta(e^t-1)}$	Pois($n\theta$)
N(θ_1, θ_2)	$\exp\{n\theta_1 t + n\theta_2 t^2/2\}$	N($n\theta_1, n\theta_2$)
Gamma (θ_1, `rate` $= \theta_2$)	$\left(\frac{\theta_2}{\theta_2 - t}\right)^{n\theta_1}$ $(t < \theta_2)$	Gamma($n\theta_1$, `rate` $= \theta_2$)
EN(θ) = Gamma(1, `rate` $= \theta$)	$\left(\frac{\theta}{\theta - t}\right)^{n}$ $t < \theta$	Gamma(n, `rate` $= \theta$)

Risulta utile osservare che, per le v.a. somma e media campionaria, nota la distribuzione dell'una, è semplice ottenere quella dell'altra. Nel caso di v.a. discrete, per le funzioni di massa di probabilità si ha che

$$f_{\overline{X}_n}(\overline{x}_n; \theta) = f_{Y_n}(n\overline{x}_n; \theta) \quad \text{e} \quad f_{Y_n}(y_n; \theta) = f_{\overline{X}_n}\left(\frac{y_n}{n}\right).$$

Nel caso di v.a. assolutamente continue, per le funzioni di densità, con una semplice trasformazione di variabili, si ha quanto segue:

$$f_{\overline{X}_n}(\overline{x}_n; \theta) = n f_{Y_n}(n\overline{x}_n; \theta) \quad \text{e} \quad f_{Y_n}(y_n; \theta) = \frac{1}{n} f_{\overline{X}_n}\left(\frac{y_n}{n}\right).$$

Consideriamo alcuni esempi rilevanti.

3.17 Esempio (Modello bernoulliano). Se $X_i \sim \text{Ber}(\theta), i = 1, \ldots, n$, i.i.d., allora [2] $Y_n = \sum_{i=1}^n X_i \sim \text{Bin}(n, \theta)$, $f_{Y_n}(y) = \binom{n}{y}\theta^y(1-\theta)^{n-y}$, $\quad y = 0, 1, \ldots, n$ e quindi:

$$\begin{aligned} f_{\overline{X}_n}(\overline{x}_n; \theta) &= \mathbb{P}\big(\overline{X}_n = \overline{x}_n;\ \theta\big) = \mathbb{P}\big(Y_n = n\overline{x}_n; \theta\big) \\ &= \binom{n}{n\overline{x}_n}\theta^{n\overline{x}_n}(1-\theta)^{n-n\overline{x}_n} \\ &= f_{Y_n}(n\overline{x}_n; \theta), \qquad \overline{x}_n = 0, \frac{1}{n}, \frac{2}{n}, \ldots 1. \end{aligned}$$

□

3.18 Esempio (Modello di Poisson). Se $X_i \sim \text{Pois}(\theta), i = 1, \ldots, n$, i.i.d., allora $Y_n = \sum_{i=1}^n X_i \sim \text{Pois}(n\theta)$, $f_{Y_n}(y) = e^{-n\theta}(n\theta)^y/y!$, $\quad y = 0, 1, 2, \ldots$ e quindi:

$$\begin{aligned} f_{\overline{X}_n}(\overline{x}_n; \theta) &= \mathbb{P}\big(\overline{X}_n = \overline{x}_n;\ \theta\big) = \mathbb{P}\big(Y_n = n\overline{x}_n; \theta\big) \\ &= \frac{e^{-n\theta}}{(n\overline{x}_n)!}(n\theta)^{n\overline{x}_n} \\ &= f_{Y_n}(n\overline{x}_n; \theta), \qquad \overline{x}_n = 0, \frac{1}{n}, \frac{2}{n}, \ldots \end{aligned}$$

□

3.19 Esempio(Modello gamma). Nel precedente Esempio 3.16 abbiamo visto che, se $X_i \sim \text{Gamma}(\theta_{1_i}, \theta_2), i = 1, \ldots, n$, indipendenti, allora (sia per la parametrizzazione `scale` che per quella `rate`) $Y_n \sim \text{Gamma}(\sum_{i=1}^n \theta_{1_i}, \theta_2)$. Ricordiamo che, in generale, se $Z \sim \text{Gamma}(\alpha, \texttt{rate} = \beta)$ e a è una costante positiva, la v.a. $aZ \sim \text{Gamma}(\alpha, \texttt{rate} = \beta/a)$.

[2] Si veda Tabella sopra.

Pertanto, $\overline{X}_n = Y_n/n \sim \text{Gamma}(\sum_{i=1}^n \theta_{1_i}, \texttt{rate} = n\theta_2)$ Se le v.a. sono anche identicamente distribuite (e quindi i.i.d.), $\overline{X}_n \sim \text{Gamma}(n\theta_1, \texttt{rate} = n\theta_2)$.
Risultati analoghi si trovano anche nel caso di parametrizzazione `scale`. Infatti, se $Z \sim \text{Gamma}(\alpha, \texttt{scale} = \beta)$ e a è una costante positiva, la v.a. $aZ \sim \text{Gamma}(\alpha, \texttt{scale} = a\beta)$. Pertanto, analogamente a quanto visto sopra, se le X_i sono v.a. $\text{Gamma}(\theta_1 = \texttt{scale} = \theta_2)$ i.i.d, allora $Y_n|\boldsymbol{\theta} \sim \text{Gamma}(n\theta_1, \texttt{scale} = \theta_2)$ e $\overline{X}_n|\boldsymbol{\theta} \sim \text{Gamma}(n\theta_1, \texttt{scale} = \theta_2/n)$. □

3.20 Esempio (Modello esponenziale negativo ed esponenziale). La variabile aleatoria esponenziale negativa si ottiene come caso particolare dalla v.a. $\text{Gamma}(\theta_1, \texttt{rate} = \theta_2)$ ponendo $\theta_1 = 1$ e $\theta_2 = \theta$. Si ha quindi che se $X_i \sim \text{EN}(\theta), i = 1, \ldots, n$, i.i.d., allora $Y_n = \sum_{i=1}^n X_i \sim \text{Gamma}(n, \texttt{rate} = \theta)$ e $\overline{X}_n \sim \text{Gamma}(n, \texttt{rate} = n\theta)$.
Nel caso del modello esponenziale, $X_i|\theta \sim \text{Esp}(\theta) = \text{Gamma}(1, \texttt{scale} = \theta)$, si ha quindi che $Y_n|\theta \sim \text{Gamma}(n, \texttt{scale} = \theta_2)$ e $\overline{X}_n|\boldsymbol{\theta} \sim \text{Gamma}(n, \texttt{scale} = \theta_2/n)$.

Ricapitoliamo nella seguente tabella i risultati relativi ai modelli gamma ed esponenziale per le due diverse parametrizzazioni.

X_i	Y_n	$\overline{X}_n$
$\text{Gamma}\ (\theta_1, \texttt{rate} = \theta_2)$	$\text{Gamma}(n\theta_1, \texttt{rate} = \theta_2)$	$\text{Gamma}(n\theta_1, \texttt{rate} = n\theta_2)$
$\text{EN}(\theta) = \text{Gamma}(1, \texttt{rate} = \theta)$	$\text{Gamma}(n, \texttt{rate} = \theta)$	$\text{Gamma}(n, \texttt{rate} = n\theta)$
$\text{Gamma}\ (\theta_1, \texttt{scale} = \theta_2)$	$\text{Gamma}(n\theta_1, \texttt{scale} = \theta_2)$	$\text{Gamma}(n\theta_1, \texttt{scale} = \theta_2/n)$
$\text{Esp}(\theta) = \text{Gamma}(1, \texttt{scale} = \theta)$	$\text{Gamma}(n, \texttt{scale} = \theta)$	$\text{Gamma}(n, \texttt{scale} = \theta/n)$

□

3.4 Campionamento da popolazioni normali

In questa sezione vengono presentati alcuni importanti risultati che riguardano statistiche campionarie nel caso di campioni casuali provenienti da popolazioni normali. Come vedremo nel Capitolo 4, questi risultati sono indispensabili per affrontare i principali problemi inferenziali per i parametri del modello normale.

3.4.1 Distribuzione di somma, media e varianza campionaria

In questa sezione (Teorema 3.23) vengono derivate le distribuzioni campionarie delle v.a. media e varianza campionaria per campioni casuali da un modello normale. A tal fine è necessario introdurre la v.a. chi-quadrato e i suoi legami con la v.a. normale.

3.21 Definizione (Chi quadrato). Una v.a. Chi quadrato con $p > 0$ gradi di libertà (in simboli $Y \sim \chi^2_p$) è una v.a. assolutamente continua con funzione di densità:

$$f_Y(y) = \frac{1}{2^{\frac{p}{2}}\Gamma\left(\frac{p}{2}\right)} y^{\frac{p}{2}-1} e^{-\frac{1}{2}y} I_{(0,+\infty)}(y). \tag{3.1}$$

□

Si ha quindi che $Y \sim \chi^2_p = \text{Ga}\left(\frac{p}{2}, \texttt{rate} = \frac{1}{2}\right) = \text{Ga}\left(\frac{p}{2}, \texttt{scale} = 2\right)$. Inoltre $\mathbb{E}[Y] = p$ e $\mathbb{V}_\theta[X] = 2p$.

3.22 Teorema (Proprietà del Chi quadrato).
(a) Data una v.a. normale standard $Z \sim \text{N}(0, 1)$, la variabile aleatoria $Y = Z^2$ è una v.a.

Chi quadrato con $p = 1$ gradi di libertà.
(b) Date $Y_1, ..., Y_n$ v.a. indipendenti e tali che $Y_i \sim \chi^2_{p_i}$, la v.a. somma è ancora una v.a. Chi quadrato con un numero di gradi di libertà che è pari alla somma dei gradi di libertà delle singole v.a., ovvero:

$$\sum_{i=1}^{n} Y_i = Y_1 + ... + Y_n \sim \chi^2_{\sum_{i=1}^n p_i}.$$

(c) Date n v.a. normali standardizzate $Z_1, ...Z_n$ indipendenti, la v.a. $Y_n = \sum_{i=1}^n Z_i^2$ è una v.a. Chi quadrato con n gradi di libertà.

Dimostrazione.
(a) Ricordando che una v.a. Chi quadrato con p gradi di libertà è una Gamma di parametri $\left(\frac{p}{2}; \frac{1}{2}\right)$ (con parametro `rate` uguale a 1/2), si deve dimostrare che $Y = Z^2$ ha la funzione di densità di una v.a. Gamma $\left(\frac{1}{2}, \frac{1}{2}\right)$ (sempre con parametrizzazione `rate`):

$$f(y) = \frac{1}{2^{\frac{1}{2}}\Gamma\left(\frac{1}{2}\right)} y^{-\frac{1}{2}} e^{-\frac{1}{2}y} I_{(0,+\infty)}(y), \tag{3.2}$$

Per effettuare la trasformazione conviene utilizzare la funzione di ripartizione. Per $y > 0$ si ha

$$F_Y(y) = P(Y \leq y) = P(Z^2 \leq y) = P\left(-\sqrt{y} \leq Z \leq \sqrt{y}\right) = F_Z(\sqrt{y}) - F_Z(-\sqrt{y}).$$

La funzione di densità si ottiene derivando la funzione di ripartizione rispetto a y, ovvero

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) = \frac{d}{dy} P(Y \leq y) = \frac{d}{dy}\left(F_Z(\sqrt{y}) - F_Z(-\sqrt{y})\right) \\ &= \frac{1}{2\sqrt{y}} f_Z(\sqrt{y}) + \frac{1}{2\sqrt{y}} f_Z(-\sqrt{y}), \end{aligned}$$

e, ricordando che la funzione di densità della v.a. normale standard è una funzione pari e che $\Gamma(1/2) = \sqrt{\pi}$,

$$f_Y(y) = \frac{1}{\sqrt{y}} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}y\right\} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y} y^{-\frac{1}{2}} \qquad y > 0$$

cioè proprio la (3.2).
(b) Per dimostrare la seconda parte del teorema, si ricordi innanzi tutto che la fgm di una v.a. $X \sim \text{Gamma}(\theta_1, \texttt{rate} = \theta_2)$ è

$$M_X(t) = \left(\frac{\theta_2}{\theta_2 - t}\right)^{\theta_1}$$

e quindi, ponendo $\theta_1 = \frac{p}{2}$ e $\theta_2 = \frac{1}{2}$, si ottiene agevolmente per Y che

$$M_Y(t) = \left(\frac{1}{1 - 2t}\right)^{\frac{p}{2}}$$

ovvero la fgm della v.a. χ^2_p. Per le note proprietà della fgm e l'indipendenza delle v.a. Y_i abbiamo quindi che

$$M_{\sum_{i=1}^n Y_i}(t) = M_{Y_1}(t) \cdot ... \cdot M_{Y_n}(t) = \prod_{i=1}^{n} \left(\frac{1}{1 - 2t}\right)^{\frac{p_i}{2}} = \left(\frac{1}{1 - 2t}\right)^{\frac{\sum_{i=1}^n p_i}{2}}$$

ed è quindi immediato riconoscere la fgm di una v.a. Chi quadrato con $\sum_{i=1}^{n} p_i$ gradi di libertà.
(c) Discende dai punti precedenti.

□

Il teorema che segue fornisce la distribuzione della media campionaria e di una funzione della devianza campionaria e costituisce un risultato fondamentale dell'inferenza statistica.

3.23 Teorema (Distribuzione di media e varianza campionaria). Sia $X_1, \ldots, X_n$ un campione casuale proveniente da una popolazione $\mathrm{N}(\mu, \sigma^2)$. Per le statistiche campionarie $\overline{X}_n = \sum_{i=1}^{n} X_i/n$, $S_n^2 = \sum_{i=1}^{n}(X_i - \overline{X}_n)^2/(n-1)$ e $\widehat{\sigma}_n^2 = \sum_{i=1}^{n}(X_i - \overline{X}_n)^2/n$ si ha:

1. $\overline{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$;
2. $\overline{X}_n$ e S_n^2 sono v.a. indipendenti;
3. $\frac{(n-1)S_n^2}{\sigma^2} = \frac{n\widehat{\sigma}_n^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \sim \chi^2_{n-1}$.

□

Dimostrazione. Vedi Appendice 3.7.1.

3.24 Osservazione

1. Il teorema fornisce le distribuzioni esatte di statistiche campionarie fondamentali. In particolare, come vedremo, la conoscenza della distribuzione di S_n^2 consente di calcolarne agevolmente la varianza.
2. In secondo punto del teorema è di fondamentale importanza. L'indipendenza di media e varianza campionaria è infatti una *caratteristica* esclusiva del modello normale. Ovvero: quando si considerano campioni casuali, $\overline{X}_n$ e S_n^2 sono indipendenti se e solo se le variabili X_i hanno distribuzione normale.

□

Il teorema che segue fornisce la distribuzione di una funzione della statistica $S_0^2 = \sum_{i=1}^{n}(X_i - \mu_0)^2/n$ che utilizzeremo per affrontare i problemi inferenziali che riguardano il parametro σ^2 quando il valore atteso della v.a. normale è noto e uguale a μ_0.

3.25 Teorema Sia $X_1, \ldots, X_n$ un campione casuale proveniente da una popolazione $\mathrm{N}(\mu_0, \sigma^2)$ (con $\mu_0 \in \mathbb{R}$ noto). Per la statistica campionaria $S_0^2 = \sum_{i=1}^{n}(X_i - \mu_0)^2/n$ si ha:

$$\frac{nS_0^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu_0)^2 \sim \chi^2_n;$$

□

Dimostrazione. Poichè $X_i \sim \mathrm{N}(\mu, \sigma^2)$, si ha che

$$Z_i = \frac{X_i - \mu}{\sigma} \sim \mathrm{N}(0,1) \quad \Rightarrow \quad Z_i^2 = \frac{(X_i - \mu)^2}{\sigma^2} \sim \chi^2_1, \quad i = 1, \ldots, n.$$

Dall'indipendenza delle X_i discende quella delle Z_i^2. Per la proprietà di additività delle v.a. Chi quadrato indipendenti si ha quindi che

$$\frac{nS_0^2}{\sigma^2} = \sum_{i=1}^{n} Z_i^2 \sim \chi^2_n.$$

I risultati dei Teoremi 3.23 e 3.25 possono essere utilizzati per determinare le distribuzioni delle statistiche S_n^2, $\widehat{\sigma}_n^2$ e S_0^2.

Corollario. Per le v.a. S_n^2, $\widehat{\sigma}_n^2$ e S_0^2 si ha

$$S_n^2 \sim \text{Gamma}\left(\frac{n-1}{2}, \texttt{rate} = \frac{n-1}{2\sigma^2}\right), \qquad \mathbb{E}[S_n^2] = \sigma^2 \qquad \text{e} \qquad \mathbb{V}[S_n^2] = \frac{2\sigma^4}{n-1},$$

$$\widehat{\sigma}_n^2 \sim \text{Gamma}\left(\frac{n-1}{2}, \texttt{rate} = \frac{n}{2\sigma^2}\right), \qquad \mathbb{E}[\widehat{\sigma}_n^2] = \frac{n-1}{n}\sigma^2 \qquad \text{e} \qquad \mathbb{V}[\widehat{\sigma}_n^2] = \frac{2(n-1)\sigma^4}{n^2},$$

e

$$S_0^2 \sim \text{Gamma}\left(\frac{n}{2}, \texttt{rate} = \frac{n}{2\sigma^2}\right), \qquad \mathbb{E}[S_0^2] = \sigma^2 \qquad \text{e} \qquad \mathbb{V}[S_0^2] = \frac{2\sigma^4}{n}$$

Dimostrazione. Ricordiamo che, in generale, una v.a. χ_p^2 coincide con una v.a. Gamma$(p/2, \texttt{rate} = 1/2)$ e che, per $a > 0$, se $X \sim \text{Gamma}(\alpha, \texttt{rate} = \beta)$ allora $aX \sim \text{Gamma}(\alpha, \texttt{rate} = \beta/a)$. Per il teorema 3.23 abbiamo che $(n-1)S_n^2/\sigma^2 \sim X_{n-1}^2 = \text{Gamma}[(n-1)/2, \texttt{rate} = 1/2]$ e quindi

$$S_n^2 = \frac{\sigma^2}{n-1}\left[\frac{n-1}{\sigma^2}S_n^2\right] \sim \text{Gamma}\left(\frac{n-1}{2}, \texttt{rate} = \frac{n-1}{2\sigma^2}\right)$$

Ricordando infine che per una v.a. Gamma$(\alpha, \texttt{rate} = \beta)$ si ha che $\mathbb{E}[X] = \alpha/\beta$ e $\mathbb{V}[X] = \alpha/\beta^2$, si ottiene facilmente che

$$\mathbb{E}[S_n^2] = \frac{n-1}{2} \times \frac{2\sigma^2}{n-1} = \sigma^2 \quad \text{e} \quad \mathbb{V}[S_n^2] = \frac{n-1}{2} \times \frac{4\sigma^4}{(n-1)^2} = \frac{2\sigma^4}{n-1}.$$

I risultati per $\widehat{\sigma}_n^2$ e per S_0^2 si verificano nello stesso modo, ricordando che $n\widehat{\sigma}_n^2/\sigma^2 \sim \chi_{n-1}^2$ e che $nS_0^2/\sigma^2 \sim \chi_n^2$.

3.4.2 Distribuzioni derivate: t di Student e F di Fisher

Consideriamo ora due distribuzioni utili per risolvere problemi inferenziali nel campionamento da popolazioni normali. La prima distribuzione si utilizza nei problemi inferenziali riguardanti il valore atteso μ, quando la varianza è incognita; la seconda nei problemi inferenziali riguardanti il rapporto tra varianze di popolazioni normali indipendenti.

T di Student

Nel prossimo capitolo vedremo che per risolvere problemi inferenziali per il valore atteso μ di una v.a. normale $\text{N}(\mu, \sigma^2)$ faremo uso della quantità aleatoria

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$$

la cui distribuzione è, per il Teorema 3.23, $\text{N}(0, 1)$. Tuttavia, nei casi in cui σ^2 è incognito, questa quantità diventa non utilizzabile. Una soluzione intuitiva è di sostituire σ al denominatore della precedente espressione con la statistica campionaria S_n,

$$\frac{\overline{X}_n - \mu}{S_n/\sqrt{n}}.$$

In questo modo l'oggetto aleatorio risultante, denominato t di Student (dallo pseudonimo di W.S. Gosset, che ne determinò la distribuzione di probabilità) non dipende dal parametro di disturbo σ^2.

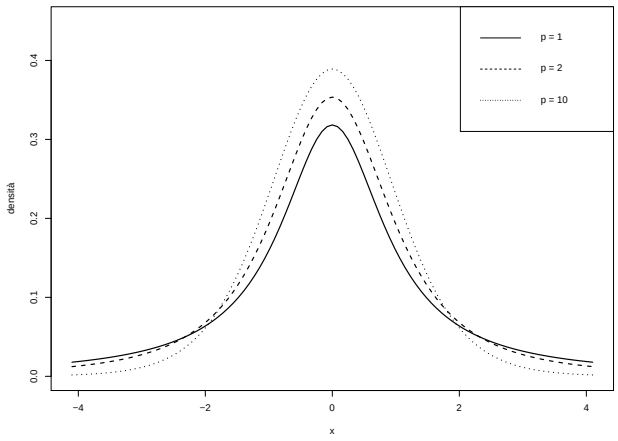

Figura 3.1: *Grafico della funzione di densità di una v.a. T di Student per alcuni valori dei gradi di libertà.*

3.26 Definizione (T di Student). Una v.a. T assolutamente continua che assume valori in $\mathbb{R}$ ha distribuzione *t di Student* con parametro $p > 0$, denominato *gradi di libertà* (notazione sintetica: $T \sim t_p$) se la sua funzione di densità è

$$f_T(t) = f_T(t;p) = \frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)} \frac{1}{(p\pi)^{1/2}} \frac{1}{(1+t^2/p)^{(p+1)/2}}, \qquad t \in \mathbb{R}. \tag{3.3}$$

□

in generale, se $T \sim t_p$

$$\mathbb{E}[T] = 0 \quad \text{se} \quad p > 1 \quad \text{e} \quad \mathbb{V}[T] = \frac{p}{p-2}, \quad \text{se} \quad p > 2.$$

Se $T \sim t_1$, ovvero se $p = 1$, T ha distribuzione di Cauchy di parametri $(0, 1)$. In questo caso sappiamo che il valore atteso e la varianza di T non esistono.
La Figura (3.1) riporta il grafico della funzione di densità della v.a. in esame per alcuni valori del parametro p.
Il teorema che segue illustra come si "costruisce" una v.a. T con distribuzione t_p e i suoi legami con le v.a. normali e Chi quadrato.

3.27 Teorema. Siano U e V due v.a. indipendenti tali che

$$U \sim \mathrm{N}(0,1) \qquad \text{e} \qquad V \sim \chi^2_p.$$

Allora

$$T = \frac{U}{\sqrt{\frac{V}{p}}} \sim t_p.$$

□

Dimostrazione. Si deve dimostrare che la v.a. T ha funzione di densità espressa dalla (3.3). Per l'ipotesi di indipendenza tra U e V la funzione di densità congiunta della v.a.

doppia (U, V) si fattorizza come segue:

$$\begin{aligned} f_{U,V}(u,v) &= f_V(v) \cdot f_U(u) \\ &= \left(2^{\frac{p}{2}}\Gamma\left(\frac{p}{2}\right)\right)^{-1} v^{\frac{p-2}{2}} e^{-\frac{1}{2}v} \cdot (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}u^2} \\ &= \left(2^{\frac{p}{2}}\Gamma\left(\frac{p}{2}\right)\sqrt{2\pi}\right)^{-1} v^{\frac{p}{2}-1} \exp\left\{-\frac{1}{2}(v+u^2)\right\}. \end{aligned} \tag{3.4}$$

È opportuno effettuare ora la seguente trasformazione doppia

$$\begin{cases} t = \frac{u}{\sqrt{v/p}} \\ x = v \end{cases} \Rightarrow \begin{cases} u = t\sqrt{x/p} \\ v = x \end{cases}$$

il cui Jacobiano è

$$J = \begin{vmatrix} \frac{\partial u}{\partial t} & \frac{\partial u}{\partial x} \\ \frac{\partial v}{\partial t} & \frac{\partial v}{\partial x} \end{vmatrix} = \begin{vmatrix} \sqrt{\frac{v}{p}} & \frac{t}{2\sqrt{px}} \\ 0 & 1 \end{vmatrix} = \sqrt{\frac{v}{p}}$$

La densità congiunta (3.4) diventa quindi:

$$f(t,x) = \left(2^{\frac{p}{2}}\Gamma\left(\frac{p}{2}\right)\sqrt{2\pi p}\right)^{-1} x^{\frac{p+1}{2}-1} \exp\left\{-\frac{1}{2}x\left(1+\frac{t^2}{p}\right)\right\},$$

che va ora integrata rispetto a x per ottenene la densità marginale di interesse

$$f_T(t) = \int_0^{+\infty} f(t,x)dx = \left(2^{\frac{p}{2}}\Gamma\left(\frac{p}{2}\right)\sqrt{2\pi p}\right)^{-1} \int_0^{+\infty} x^{\frac{p+1}{2}-1} \exp\left\{-\frac{1}{2}x\left(1+\frac{t^2}{p}\right)\right\} dx.$$

È immediato riconoscere nella precedente funzione integranda il nucleo di una Gamma di parametri $\alpha = \frac{p+1}{2}$ e $\beta = \frac{1}{2}(1+\frac{t^2}{p})$ e risolvere quindi l'integrale come segue:

$$\begin{aligned} f_T(t) &= \left(2^{\frac{p}{2}}\Gamma\left(\frac{p}{2}\right)\sqrt{2\pi p}\right)^{-1} \cdot \frac{\Gamma\left(\frac{p+1}{2}\right)}{\left[\frac{1}{2}\left(1+\frac{t^2}{p}\right)\right]^{\frac{p+1}{2}}} \\ &= \frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)} \frac{1}{(p\pi)^{1/2}} \frac{1}{(1+t^2/p)^{(p+1)/2}} \end{aligned}$$

che è in effetti la densità di una v.a. t di Student definita dalla (3.3) con p gradi di libertà.

Corollario. Sia $X_1, \ldots, X_n$ un campione casuale proveniente da una popolazione $\mathrm{N}(\mu, \sigma^2)$. Allora

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{S_n} \sim t_{n-1}.$$

Dimostrazione. Definiamo U e V nel modo seguente

$$U = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \qquad \text{e} \qquad V = \frac{(n-1)S_n^2}{\sigma^2}.$$

Dal Teorema (3.23) discende che $U \sim \mathrm{N}(0,1)$ [parte (1)], $V \sim \chi^2_{n-1}$ [parte (3)] e che U e V sono indipendenti [parte (2)]. Pertanto, per il teorema precedente

$$\frac{U}{\sqrt{V/(n-1)}} = \frac{\sqrt{n}(\overline{X}_n - \mu)/\sigma}{\sqrt{\frac{(n-1)S_n^2/\sigma^2}{n-1}}} = \frac{\sqrt{n}(\overline{X}_n - \mu)}{S_n} \sim t_{n-1}.$$

F di Fisher

3.28 Definizione (F di Fisher). Una v.a. W assolutamente continua che assume valori in $\mathbb{R}^+$ ha distribuzione *F di Fisher* (o Fisher-Snedecor) con parametri (gradi di libertà) (p, q), (notazione sintetica: $W \sim F_{p,q}$) se la sua funzione di densità è

$$f_W(w) = \frac{\Gamma\left(\frac{p+q}{2}\right)}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{q}{2}\right)} \left(\frac{p}{q}\right)^{p/2} \frac{w^{p/2-1}}{[1+(p/q)w]^{(p+q)/2}}, \qquad w \in \mathbb{R}^+, \quad p, q \in \mathbb{R}^+. \tag{3.5}$$

□

Per la v.a. $W \sim F_{p,q}$ si ha che

$$\mathbb{E}[W] = \frac{q}{q-2}, \quad q > 2 \qquad \text{e} \qquad \mathbb{V}[W] = 2\left(\frac{q}{q-2}\right)^2 \frac{p+q-2}{p(q-4)}, \quad q > 4.$$

Inoltre, se $T \sim t_q$, allora $T^2 \sim F_{1,q}$.

Il teorema che segue illustra come si "costruisce" una v.a. W con distribuzione $F_{p,q}$ e quali sono i suoi legami con le v.a. normale e Chi quadrato.

3.29 Teorema. Siano U e V due v.a. indipendenti tali che

$$U \sim \chi^2_p \qquad \text{e} \qquad V \sim \chi^2_q.$$

Allora

$$W = \frac{U/p}{V/q} \sim F_{p,q}.$$

□

Dimostrazione. Per dimostrare che la v.a. $W = \frac{U/p}{V/q}$ ha funzione di densità espressa dalla (3.5), si parte dalla funzione di densità congiunta della v.a. doppia (U, V):

$$f_{U,V}(u,v) = \left[2^{\frac{p+q}{2}}\Gamma(\frac{p}{2})\Gamma(\frac{q}{2})\right]^{-1} u^{\frac{p}{2}-1} v^{\frac{q}{2}-1} \exp\left\{-\frac{1}{2}(u+v)\right\} \tag{3.6}$$

Si effettua ora la trasformazione doppia

$$\begin{cases} w = \frac{\frac{u}{p}}{\frac{v}{q}} \\ x = \frac{v}{q} \end{cases} \Rightarrow \begin{cases} v = qx \\ u = pwx \end{cases}$$

con $u, v, w, x > 0$, il cui Jacobiano è

$$J = \begin{vmatrix} \frac{\partial u}{\partial w} & \frac{\partial u}{\partial x} \\ \frac{\partial v}{\partial w} & \frac{\partial v}{\partial x} \end{vmatrix} = \begin{vmatrix} px & pw \\ 0 & q \end{vmatrix} = pqx$$

Sostituendo nella (3.6) e moltiplicando per lo Jacobiano si ha

$$\begin{aligned} f(w,x) &= \left[2^{\frac{p+q}{2}}\Gamma(\frac{p}{2})\Gamma(\frac{q}{2})\right]^{-1} (pwx)^{\frac{p}{2}-1}(qx)^{\frac{q}{2}-1} \exp\left\{-\frac{1}{2}(qx+pwx)\right\} pqx = \\ &= \left[2^{\frac{p+q}{2}}\Gamma(\frac{p}{2})\Gamma(\frac{q}{2})\right]^{-1} p^{\frac{p}{2}} q^{\frac{q}{2}} w^{\frac{p}{2}-1} x^{\frac{p+q}{2}-1} \exp\left\{-\frac{1}{2}x(q+pw)\right\} \end{aligned}$$

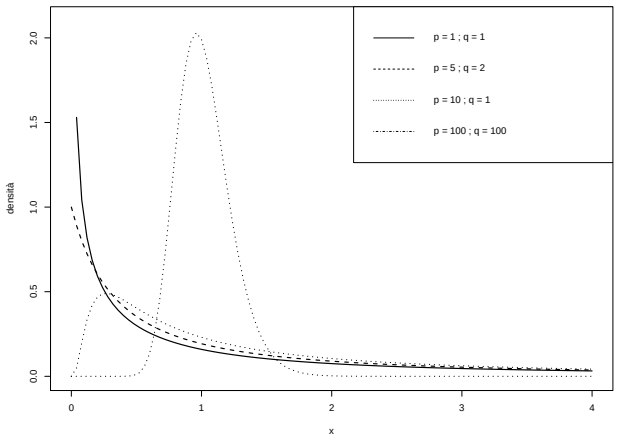

Figura 3.2: *Grafico della funzione di densità di una v.a. F di Fisher per alcuni valori dei parametri p e q.*

A questo punto per ricavare la densità marginale $f_F(w)$ si integra la densità congiunta rispetto a x, e si risolve l'integrale per sostituzione ponendo $t = x(pw+q) \Rightarrow dt = dx(pw+q) \Rightarrow x = \frac{t}{pw+q}$:

$$\begin{aligned}
f_W(w) &= \int_0^\infty f(w,x)dx = \\
&= \left[2^{\frac{p+q}{2}}\Gamma(\frac{p}{2})\Gamma(\frac{q}{2})\right]^{-1} p^{\frac{p}{2}} q^{\frac{q}{2}} w^{\frac{p}{2}-1} \int_0^\infty x^{\frac{p+q}{2}-1} \exp\left\{-\frac{1}{2}x(q+pw)\right\} dx \\
&= \left[2^{\frac{p+q}{2}}\Gamma(\frac{p}{2})\Gamma(\frac{q}{2})\right]^{-1} p^{\frac{p}{2}} q^{\frac{q}{2}} w^{\frac{p}{2}-1} \int_0^\infty (\frac{t}{pw+q})^{\frac{p+q}{2}-1} \exp\left\{-\frac{1}{2}t\right\} d\frac{t}{pw+q} = \\
&= \left[2^{\frac{p+q}{2}}\Gamma(\frac{p}{2})\Gamma(\frac{q}{2})\right]^{-1} p^{\frac{p}{2}} q^{\frac{q}{2}} w^{\frac{p}{2}-1} (\frac{1}{pw+q})^{\frac{p+q}{2}} \int_0^\infty t^{\frac{p+q}{2}-1} \exp\left\{-\frac{1}{2}t\right\} dt = \\
&= \left[2^{\frac{p+q}{2}}\Gamma(\frac{p}{2})\Gamma(\frac{q}{2})\right]^{-1} p^{\frac{p}{2}} q^{\frac{q}{2}} w^{\frac{p}{2}-1} (\frac{1}{pw+q})^{\frac{p+q}{2}} \left[2^{\frac{p+q}{2}}\Gamma(\frac{p+q}{2})\right] = \\
&= \frac{\Gamma(\frac{p+q}{2})}{\Gamma(\frac{p}{2})\Gamma(\frac{q}{2})} p \frac{(pw)^{\frac{p}{2}-1}}{(pw+q)^{\frac{p+q}{2}}} q^{\frac{q}{2}}
\end{aligned}$$

che con un semplice passaggio algebrico (basta aggiungere e togliere $\frac{p}{2}+1$ all'esponente di q) coincide con la (3.5).
La Figura (3.2) riporta il grafico della funzione di densità della v.a. in esame per alcuni valori dei parametri p e q.

Possiamo utilizzare il precedente risultato per determinare la funzione di densità di una funzione del rapporto delle varianze campionarie relative a campioni indipendenti provenienti da due popolazioni normali. Vale il seguente teorema.

Corollario. Sia $X_1, \ldots, X_n$ un campione casuale proviente da una popolazione $\mathrm{N}(\mu_x, \sigma_x^2)$ e $Y_1, \ldots, Y_m$ un campione casuale, indipendente dal primo, proveniente da una popolazione $\mathrm{N}(\mu_y, \sigma_y^2)$. Siano inoltre S_x^2 e S_y^2 le varianze campionarie corrette ottenute dai due campioni.

Si ha allora che la variabile aleatoria

$$\frac{S_x^2/\sigma_x^2}{S_y^2/\sigma_y^2} \sim F_{n-1,m-1}.$$

Alcune proprietà utili

- $F_{n_1,\infty} \to \frac{\chi^2_{n_1}}{n_1}$
- $F_{\infty,n_2} \to \frac{n_2}{\chi^2_{n_2}}$
- $X \sim F_{n_1,n_2} \Rightarrow \frac{1}{X} \sim F_{n_2,n_1}$
- $X \sim t_n \Rightarrow X^2 \sim F_{1,n}$

3.5 Approssimazioni asintotiche

Non sempre è agevole determinare la distribuzione di probabilità delle v.a. somma e media campionaria, Y_n e $\overline{X}_n$. Inoltre, spesso, anche quando queste distribuzioni sono note, il loro uso può risultare analiticamente oneroso. In questi casi, sotto opportuno condizioni, è possibile utilizzare delle approssimazioni delle distribuzioni di probabilità di Y_n e $\overline{X}_n$ che possiamo ottenere in virtù delle proprietà di convergenza delle successioni di v.a. $(Y_n,\ n \in \mathbb{N})$ e $(\overline{X}_n,\ n \in \mathbb{N})$. Parliamo in questo caso di approssimazioni asintotiche delle distribuzioni delle statistiche somma e media campionarie. Il risultato di convergenza che sfruttiamo è il teorema del limite centrale. Premettiamo alla sua enunciazione la definizione di convergenza in distribuzione a una variabile aleatoria di una successione di v.a.

3.30 Definizione (Convergenza in distribuzione). Sia $(X_n,\ n \in \mathbb{N})$ una successione di v.a. e siano $(F_n(\cdot),\ n \in \mathbb{N})$ e $F_X(\cdot)$ la successione delle funzioni di ripartizioni delle v.a. X_n e la funzione di ripartizione di X. La successione delle v.a. X_n *converge in distribuzione* alla v.a. X (sinteticamente: $X_n \xrightarrow{d} X$) se

$$\lim_{n\to\infty} F_n(x) = F_X(x)$$

in ogni punto x in cui $F_X(\cdot)$ è continua. □

Per una successione di v.a. convergente a X, è ragionevole assumere che, a partire da un valore di n adeguatamente elevato,

$$F_{X_n}(x) \simeq F_X(x),$$

ovvero che il valore della f.r. di una specifica v.a. X_n della successione, calcolato in x, sia abbastanza vicino al valore di F_X calcolato nello stesso punto. Grazie al teorema del limite centrale possiamo applicare questo modo di ragionare alla successione delle statistiche campionarie Y_n e $\overline{X}_n$.

3.31 Teorema (Limite centrale). Sia $(X_n,\ n \in \mathbb{N})$ una successione di variabili aleatorie i.i.d. con valore atteso $\mathbb{E}[X]$ e $0 < \mathbb{V}[X] < \infty$. Si ha allora che la successione

$$\frac{\sqrt{n}(\overline{X}_n - \mathbb{E}[X])}{\sqrt{\mathbb{V}[X]}} \xrightarrow{d} \mathrm{N}(0,1).$$

□

Il precedente è un risultato limite. Per n sufficientemente grande (ma finito) possiamo però aspettarci che

$$\frac{\sqrt{n}(\overline{X}_n - \mathbb{E}[X])}{\sqrt{\mathbb{V}[X]}} \dot{\sim} \mathrm{N}(0,1),$$

dove, in generale con la notazione $U \dot{\sim} V$ indichiamo che la distribuzione della v.a. V è approssimativamente uguale a quella della v.a. U. Dalla precedente relazione discende che

$$\overline{X}_n \dot{\sim} N\left(\mathbb{E}[X], \mathbb{V}[X]/n\right), \qquad \text{e} \qquad Y_n \dot{\sim} N(n\mathbb{E}[X], n\mathbb{V}[X]).$$

Questo vuol dire che, per n sufficientemente elevato, per ogni $x \in \mathbb{R}$ si ha che

$$F_{\overline{X}_n}(x) \simeq \Phi\left(\frac{\sqrt{n}(x - \mathbb{E}[X])}{\sqrt{\mathbb{V}[X]}}\right) \qquad \text{e} \qquad F_{Y_n}(x) \simeq \Phi\left(\frac{x - n\mathbb{E}[X]}{\sqrt{n\mathbb{V}[X]}}\right),$$

dove $\Phi(\cdot)$ indica la f.r. della v.a. $\mathrm{N}(0,1)$ e $F_{\overline{X}_n}(\cdot)$ e $F_{Y_n}(\cdot)$ sono le f.r. delle v.a. $\overline{X}_n$ e Y_n. Questo risultato consente ad esempio una agevole approssimazione della probabilità che le v.a. Y_n e $\overline{X}_n$ assumano valori in un intervallo reale (a, b). Si ha infatti che

$$\mathbb{P}(a \leq \overline{X}_n \leq b) = F_{\overline{X}_n}(b) - F_{\overline{X}_n}(a) \simeq \Phi\left(\frac{\sqrt{n}(b - \mathbb{E}[X])}{\sqrt{\mathbb{V}[X]}}\right) - \Phi\left(\frac{\sqrt{n}(a - \mathbb{E}[X])}{\sqrt{\mathbb{V}[X]}}\right),$$

e

$$\mathbb{P}(a \leq Y_n \leq b) = F_{Y_n}(b) - F_{Y_n}(a) \simeq \Phi\left(\frac{b - n\mathbb{E}[X]}{\sqrt{n\mathbb{V}[X]}}\right) - \Phi\left(\frac{a - n\mathbb{E}[X]}{\sqrt{n\mathbb{V}[X]}}\right).$$

Si noti che le approssimazioni per le probabilità del tipo

$$\mathbb{P}(Y_n \leq b) \qquad \text{e} \qquad \mathbb{P}(Y_n \geq a)$$

si ottengono facilmente dalle precedenti espressioni rispettivamente per $a \to -\infty$ e per $b \to +\infty$. Per le proprietà della funzione di ripartizione si ha

$$\mathbb{P}(Y_n \leq b) \simeq \Phi\left(\frac{b - n\mathbb{E}[X]}{\sqrt{n\mathbb{V}[X]}}\right) \qquad \text{e} \qquad \mathbb{P}(Y_n \geq a) \simeq 1 - \Phi\left(\frac{a - n\mathbb{E}[X]}{\sqrt{n\mathbb{V}[X]}}\right)$$

3.32 Esempio (Approssimazione normale della binomiale). Sia $X_1, \ldots, X_n$ un campione casuale con $X_i \sim \mathrm{Ber}(\theta)$, $i = 1, \ldots, n$, $\theta \in [0,1]$. Sappiamo già che la v.a. $Y_n = \sum_{i=1}^n X_i \sim \mathrm{Binom}(n, \theta)$. Per il Teorema (3.31)

$$Y_n \dot{\sim} \mathrm{N}(n\theta, n\theta(1-\theta)) \qquad \text{e} \qquad \overline{X}_n \dot{\sim} \mathrm{N}(\theta, \theta(1-\theta)/n).$$

Pertanto, per $a, b \in \mathbb{R}^+$,

$$\mathbb{P}(a \leq Y_n \leq b) \simeq \Phi\left(\frac{b - n\theta}{\sqrt{n\theta(1-\theta)}}\right) - \Phi\left(\frac{a - n\theta}{\sqrt{n\theta(1-\theta)}}\right)$$

Analogamente, per $\overline{X}_n$ si ottiene

$$\mathbb{P}(a \leq \overline{X}_n \leq b) \simeq \Phi\left(\frac{\sqrt{n}(b - \theta)}{\sqrt{\theta(1-\theta)}}\right) - \Phi\left(\frac{\sqrt{n}(a - \theta)}{\sqrt{\theta(1-\theta)}}\right)$$

Per valutare l'accuratezza dell'approssimazione normale con un esempio, determiniamo il valore di $\mathbb{P}(Y_n \geq 30)$, assumendo $n = 100$ e $\theta = 0.4$. Utilizzando ad esempio il software `R`[3] si ottiene

$$\begin{aligned}\mathbb{P}(Y_n \geq 30) &= \sum_{i=30}^{100} \binom{100}{i} (0.4)^i (1-0.4)^{100-i} \\ &= 1 - \texttt{pbinom}(29, \texttt{size} = 100, \texttt{prob} = 0.4) = 0.985.\end{aligned}$$

Sfruttando l'approssimazione normale si trova

$$\mathbb{P}(Y_n \geq 30) \simeq 1 - \Phi\left(\frac{30 - 100 \cdot (0.4)}{\sqrt{100 \cdot (0.4) \cdot (0.6)}}\right) = 1 - \Phi(-2.041) = 0.979.$$

L'accuratezza dell'approssimazioni dipende dalla dimensione campionaria e dal valore del parametro θ (che però nei problemi inferenziali non è noto). □

3.33 Esempio (Modello di Poisson). Sia $X_1, \ldots, X_n$ un campione casuale con $X_i \sim \text{Pois}(\theta),\ i = 1, \ldots, n,\ \theta \in \mathbb{R}^+$. Sappiamo che in questo caso $Y_n \sim \text{Pois}(n\theta)$ Per il Teorema (3.31)

$$Y_n \dot{\sim} \text{N}(n\theta, n\theta) \qquad \text{e} \qquad \overline{X}_n \dot{\sim} \text{N}(\theta, \theta/n).$$

□

3.34 Esempio (Modello esponenziale). Sia $X_1, \ldots, X_n$ un campione casuale con $X_i \sim \text{Esp}(\theta),\ i = 1, \ldots, n,\ \theta \in \mathbb{R}^+$. La distribuzione esatta di Y_n e $\overline{X}_n$ è stata determinata in un esempio precedente. Per il Teorema (3.31)

$$Y_n \dot{\sim} \text{N}(n\theta, n\theta^2) \qquad \text{e} \qquad \overline{X}_n \dot{\sim} \text{N}(\theta, \theta^2/n).$$

□

3.35 Esempio (Modello uniforme). Sia $X_1, \ldots, X_n$ un campione casuale con $X_i \sim \text{Unif}(0, \theta),\ i = 1, \ldots, n,\ \theta \in \mathbb{R}^+$. Si ha quindi che

$$Y_n \dot{\sim} \text{N}\left(\frac{n\theta}{2}, \frac{n\theta^2}{12}\right) \qquad \text{e} \qquad \overline{X}_n \dot{\sim} \text{N}\left(\frac{\theta}{2}, \frac{\theta^2}{12n}\right).$$

□

Enunciamo ora un teorema utile per determinare una approssimazione normale per la distribuzione di alcune funzioni di successioni di v.a.

3.36 Teorema (Slutsky). Siano $(X_n,\ n \in \mathbb{N})$ e $(Z_n,\ n \in \mathbb{N})$ due successioni[4] di v.a. e X una v.a. tali che $X_n \xrightarrow{d} X$ e $Z_n \xrightarrow{p} a$, con $a \in \mathbb{R}$. Allora

1. $Z_n X_n \xrightarrow{d} aX$.
2. $X_n + Z_n \xrightarrow{d} X + a$.

[3] Si veda il sito `http://cran.r-project.org/`

[4] Si ricordi che una successione di v.a. converge in probabilità a una costante a se, $\forall \epsilon > 0$ $\lim_{n\to\infty} \mathbb{P}\{|X_n - a| > \epsilon\} = 0$.

□

3.37 Esempio. Il Teorema di Slutsky consente, ad esempio, di determinare la distribuzione approssimata della funzione

$$\frac{\sqrt{n}(\overline{X}_n - \mathbb{E}[X])}{S_n}$$

quando vale il teorema del limite centrale. In questo caso, infatti, $\sqrt{n}(\overline{X}_n - \mathbb{E}[X])/\sqrt{\mathbb{V}[X]} \xrightarrow{d} \mathrm{N}(0,1)$ e se $\sqrt{\mathbb{V}[X]}/S_n \xrightarrow{p} 1$ (condizione verificata se $\lim_{n\to\infty} \mathbb{V}[S_n^2] = 0$), allora

$$\frac{\sqrt{n}(\overline{X}_n - \mathbb{E}[X])}{S_n} = \frac{\sqrt{\mathbb{V}[X]}}{S_n} \frac{\sqrt{n}(\overline{X}_n - \mathbb{E}[X])}{\sqrt{\mathbb{V}[X]}} \xrightarrow{d} \mathrm{N}(0,1).$$

Si noti che la precedente approssimazione vale per campioni provenienti da popolazioni diverse da quella normale. Nel caso di campionamento da popolazioni normali non è necessario ricorrere a una approssimazione asintotica per la distribuzione di $\sqrt{n}(\overline{X}_n - \mathbb{E}[X])/S_n$, dal momento che la distribuzione esatta di questa variabile aleatoria è nota per ogni possibile valore di $n > 2$ e coincide con quella di quella di una t_{n-1}. □

3.5.1 Metodo delta

Il metodo delta consente di trovare una approssimazione per la distribuzione di probabilità di una funzione di una v.a. (qui unidimensionale) che ha distribuzione asintotica normale. Si tratta, in buona sostanza, di una estensione del teorema del limite centrale.

3.38 Teorema Sia $(X_n,\ n \in \mathbb{N})$ una successione di v.a. tale che

$$\sqrt{n}(X_n - \theta) \xrightarrow{d} \mathrm{N}(0, v(\theta))$$

e g una funzione (di X_n e θ) tale che esiste la derivata prima $g'(\theta) \neq 0$. Si ha allora che

$$\sqrt{n}[g(X_n) - g(\theta)] \xrightarrow{d} \mathrm{N}(0, v(\theta)[g'(\theta)]^2).$$

□

Dimostrazione. Le ipotesi sulla funzione g consentono lo sviluppo in serie di Taylor, arrestato al primo ordine, della funzione $g(\cdot)$ per X_n in un intorno di θ:

$$g(X_n) = g(\theta) + g'(\theta)(X_n - \theta) + r_n$$

dove il resto dello sviluppo in serie r_n tende in probabilità a zero per $n \to +\infty$. Si ottiene quindi che

$$g(X_n) - g(\theta) \simeq g'(\theta)(X_n - \theta). \tag{3.7}$$

Premoltiplicando entrambe le precedenti quantità per $\sqrt{n}$ si ottiene

$$\sqrt{n}[g(X_n) - g(\theta)] \simeq g'(\theta)\sqrt{n}[(X_n - \theta)].$$

Poichè $\sqrt{n}[(X_n - \theta)]$ per ipotesi converge in distribuzione a $\mathrm{N}(0, v(\theta))$, per il Teorema di Slutsky,

$$g'(\theta)\sqrt{n}[(X_n - \theta)] \xrightarrow{d} \mathrm{N}(0, v(\theta)[g'(\theta)]^2)$$

e quindi, per la (3.7) si ha quindi che

$$\sqrt{n}[g(X_n) - g(\theta)] \xrightarrow{d} \mathrm{N}(0, v(\theta)[g'(\theta)]^2).$$

3.39 Osservazione. Grazie al precedente risultato, per n sufficientemente elevato abbiamo che,

$$X_n \dot{\sim} N\left(\theta, \frac{v(\theta)}{n}\right) \quad \Rightarrow \quad g(X_n) \dot{\sim} N\left(g(\theta), \frac{v(\theta)}{n}[g'(\theta)]^2\right), \tag{3.8}$$

e che:

$$\mathbb{E}_\theta[g(X_n)] \simeq g(\theta) \qquad \mathbb{V}_\theta[g(X_n)] \simeq v(\theta)[g'(\theta)]^2/n.$$

Le quantità $g(\theta)$ e $v(\theta)[g'(\theta)]^2/n$ sono il *valore atteso asintotico* e la *varianza asintotica* di $g(X_n)$. □

3.40 Esempio (Modello esponenziale). Sia $X_1, \ldots, X_n$ un campione casuale da una popolazione $\mathrm{Esp}(\theta)$. Poichè in questo caso $\mathbb{E}_\theta[X_i] = \theta$ e $\mathbb{V}_\theta[X_i] = \theta^2$, per il teorema del limite centrale si ha che

$$\sqrt{n}\left(\overline{X}_n - \theta\right) \xrightarrow{d} N\left(0, \theta^2\right)$$

Vogliamo determinare una approssimazione asintotica per la distribuzione di $g(\overline{X}_n) = 1/\overline{X}_n$. Per la funzione $g(\theta) = 1/\theta$ si ha che $g'(\theta) = -1/\theta^2$. Pertanto, per il teorema (3.38) e la (3.8), osservando che in questo caso $v(\theta) = \theta^2$, si ha che

$$\frac{1}{\overline{X}_n} \dot{\sim} \mathrm{N}\left(\frac{1}{\theta}, \frac{1}{n\theta^2}\right).$$

□

3.6 Distribuzione di $X_{(1)}$ e $X_{(n)}$

In alcuni modelli statistici la risoluzione dei problemi inferenziali coinvolge le statistiche ordinate, $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$. Ad esempio, ciò accade nei problemi in cui il supporto della v.a. di base X ha supporto $\mathcal{X}_\theta$ dipendente dal parametro incognito del modello. Un esempio tipico è il modello uniforme in $[0, \theta]$. Risulta pertanto utile disporre delle leggi di probabilità di queste variabili aleatorie. Il teorema che segue fornisce le espressioni generali delle funzioni di ripartizione e, nel caso di v.a. X_i assolutamente continue, delle funzioni di densità di probabilità di $X_{(1)}$ e $X_{(n)}$.

3.41 Teorema Siano $X_1, \ldots, X_n$ n variabili aleatorie i.i.d., ciascuna con funzione di ripartizione $F_{X_i}(\cdot; \theta) = F_X(\cdot; \theta)$ e siano

$$X_{(1)} = \min\{X_1, \ldots, X_n\}, \qquad X_{(n)} = \max\{X_1, \ldots, X_n\}.$$

1. Le funzioni di ripartizione di $X_{(1)}$ e $X_{(2)}$ sono rispettivamente

$$F_{X_{(1)}}(x; \theta) = 1 - [1 - F_X(x; \theta)]^n \qquad \text{e} \qquad F_{X_{(n)}}(x; \theta) = [F_X(x; \theta)]^n;$$

2. Se le v.a. X_i sono assolutamente continue con densità $f_X(\cdot; \theta)$, le funzioni di densità di $X_{(1)}$ e $X_{(n)}$ sono

$$f_{X_{(1)}}(x; \theta) = n[1 - F_X(x; \theta)]^{n-1} f_X(x, \theta) \qquad \text{e} \qquad f_{X_{(n)}}(x; \theta) = n[F_X(x; \theta)]^{n-1} f_X(x; \theta).$$

□

Dimostrazione.

1. Per $X_{(1)}$ si ha:

$$\begin{aligned} F_{X_{(1)}}(x;\theta) &= \mathbb{P}\big(\min\{X_1,\dots X_n\} \le x;\ \theta\big) = \mathbb{P}\big(X_1 \le x \cup X_2 \le x \cup \dots \cup X_n \le x\,;\,\theta\big) \\ &= \mathbb{P}\big(\cup_{i=1}^n \{X_i \le x\}\,;\,\theta\big) = 1 - \mathbb{P}\big(\,\big[\cup_{i=1}^n \{X_i \le x\}\big]^C;\,\theta\big) \\ &= 1 - \mathbb{P}\big(\,\big[\cap_{i=1}^n \{X_i \le x\}^C\big];\,\theta\big) \\ &= 1 - \mathbb{P}\big(\cap_{i=1}^n \{X_i > x\}\,;\,\theta\big) \\ &= 1 - \prod_{i=1}^n \mathbb{P}\big(X_i > x\,;\,\theta\big) \\ &= 1 - \big[1 - \mathbb{P}\big(X \le x\,;\,\theta\big)\big]^n \\ &= 1 - \big[1 - F_X\big(x;\,\theta\big)\big]^n, \end{aligned}$$

dove la quinta uguaglianza è dovuta alla legge di de Morgan, in base alla quale, per n insiemi $A_1,\dots,A_n$ si ha che $[\cup_{i=1}^n A_i]^C = \cap_{i=1}^n A_i^C$.

Per $X_{(n)}$ si ha:

$$\begin{aligned} F_{X_{(n)}}(x;\theta) &= \mathbb{P}\big(\max\{X_1,\dots X_n\} \le x;\ \theta\big) = \mathbb{P}\big(X_1 \le x \cap X_1 \le x \cap \dots \cap X_1 \le x\,;\,\theta\big) \\ &= \mathbb{P}\big(\cap_{i=1}^n \{X_i \le x\ ;\ \theta\}\big) = \prod_{i=1}^n \big[F_{X_i}(x\,;\,\theta)\big] = \big[F_X(x\,;\,\theta)\big]^n. \end{aligned}$$

2. Le funzioni di densità di $X_{(1)}$ e $X_{(n)}$ si ottengono immediatamente derivando le funzioni di ripartizione:

$$\begin{aligned} f_{X_{(1)}}(x;\theta) &= \frac{\partial}{\partial x} F_{X_{(1)}}(x;\theta) = n[1 - F_X(x;\theta)]^{n-1} f_X(x,\theta) \qquad \text{e} \\ f_{X_{(n)}}(x;\theta) &= \frac{\partial}{\partial x} F_{X_{(n)}}(x;\theta) = n[F_X(x;\theta)]^{n-1} f_X(x;\theta). \end{aligned}$$

3.42 Esempio (Modello uniforme). Se le variabili X_i hanno distribuzione uniforme in $[0,\theta]$, la funzione di densità è

$$f_X(x;\theta) = \frac{1}{\theta} I_{[0,\theta]}(x)$$

e la funzione di ripartizione è

$$F_X(x;\theta) = \frac{x}{\theta} I_{[0,\theta]}(x).$$

In questo caso la funzione di densità della v.a. $X_{(n)}$ risulta essere:

$$f_{X_{(n)}}(x;\theta) = n\left[\frac{x}{\theta}\right]^{n-1} \frac{1}{\theta} I_{[0,\theta]}(x) = n\frac{x^{n-1}}{\theta^n} I_{[0,\theta]}(x).$$

È immediato calcolare il valore atteso, il momento secondo e la varianza di $X_{(n)}$:

$$\mathbb{E}(X_{(n)}) = \int_0^\theta x \cdot n\frac{x^{n-1}}{\theta^n} dx = \frac{n}{\theta^n}\frac{x^{n+1}}{n+1}\Big|_0^\theta = \frac{n}{n+1}\theta$$

$$\mathbb{E}(X_{(n)}^2) = \int_0^\theta x^2 \cdot n\frac{x^{n-1}}{\theta^n} dx = \frac{n}{\theta^n}\frac{x^{n+2}}{n+2}\Big|_0^\theta = \frac{n}{n+2}\theta^2$$

$$\mathbb{V}(X_{(n)}) = \mathbb{E}(X_{(n)}^2) - [\mathbb{E}(X_{(n)})]^2 = \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\theta\right)^2 = \frac{n\theta^2}{(n+1)^2(n+2)}$$

□

3.7 Appendice

3.7.1 Dimostrazione del teorema 3.23

1. Vedi Esempio 3.14.

2. Per dimostrare l'indipendenza di S_n^2 e $\overline{X}_n$ mostriamo che: (a) S_n^2 è funzione delle v.a. $Y_2 = X_2 - \overline{X}_n, Y_3 = X_3 - \overline{X}_n, \ldots, Y_n = X_n - \overline{X}_n$; (b) $Y_2, \ldots, Y_n$ sono indipendenti dalla v.a. $Y_1 = \overline{X}_n$.

   Osservando che $\sum_{i=1}^n (X_i - \overline{X}_n) = 0$, si ottiene che

   $$(X_1 - \overline{X}_n)^2 = \left[\sum_{i=2}^n (X_i - \overline{X}_n)\right]^2$$

   e quindi che

   $$\begin{aligned}(n-1)S_n^2 &= \sum_{i=1}^n (X_i - \overline{X}_n)^2 \\ &= (X_1 - \overline{X}_n)^2 + \sum_{i=2}^n (X_i - \overline{X}_n)^2 \\ &= \left[\sum_{i=2}^n (X_i - \overline{X}_n)\right]^2 + \sum_{i=2}^n (X_i - \overline{X}_n)^2.\end{aligned}$$

   Il punto (a) è quindi verificato. Per dimostrare (b) si deve verificare che la funzione di densità congiunta delle v.a. $Y_1, Y_2, \ldots, Y_n$ si fattorizza come segue

   $$f_{Y_1,Y_2,\ldots,Y_n}(y_1, y_2, \ldots, y_n) = f_{Y_1}(y_1) \times f_{Y_2,\ldots,Y_n}(y_2, \ldots, y_n).$$

   La funzione di densità $f_{Y_1,Y_2,\ldots,Y_n}$ si determina considerando la formula di trasformazione per v.a. multiple. Considerando le trasformazioni $Y_i = g_i(\mathbf{X}_n)$ e le trasformazioni inverse $X_i = h_i(\mathbf{Y_n}),\ i = 1, \ldots, n$:

   $$\left\{\begin{array}{l} y_1 = g_1(\mathbf{x}_n) = \overline{x}_n \\ y_2 = g_2(\mathbf{x}_n) = x_2 - \overline{x}_n \\ \vdots \\ y_n = g_n(\mathbf{x}_n) = x_n - \overline{x}_n \end{array}\right. \quad \text{e} \quad \left\{\begin{array}{l} x_1 = h_1(\mathbf{y}_n) = y_1 - (y_2 + \ldots + y_n) \\ x_2 = h_2(\mathbf{y}_n) = y_2 + y_1 \\ \vdots \\ x_n = h_n(\mathbf{y}_n) = y_n + y_1 \end{array}\right.$$

   la funzione di denstità congiunta nelle nuove variabili è definita come segue

   $$f_{Y_1,Y_2,\ldots,Y_n}(y_1, y_2, \ldots, y_n) = f_{X_1,X_2,\ldots,X_n}[h_1(\mathbf{y}_n), h_2(\mathbf{y}_n), \ldots, h_n(\mathbf{y}_n)] \cdot |J|,$$

   dove $|J|$ è il determinante della matrice delle derivate parziali delle funzioni h_i rispetto alle variabili $y_i,\ i = 1, \ldots, n$. In questo caso si verifica facilmente che

   $$J = \begin{bmatrix} 1 & -1 & -1 & \ldots & -1 \\ 1 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \ldots & 1 \end{bmatrix}.$$

   Si può mostrare che il determinante di questa matrice è pari a n. Considerando che

   $$f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = \frac{1}{(2\pi)^{n/2}} e^{-1/2\sum_{i=1}^n x_i^2}$$

e sostitituendo le espressioni $x_i = h_i(\mathbf{y}_n),\ i = 1, \ldots, n$ prima determinate si ottiene
$f_{Y_1,Y_2,\ldots,Y_n}(y_1, y_2, \ldots, y_n) =$

$$\begin{aligned} &= f_{X_1}\left(y_1 - \sum_{i=2}^{n} y_i\right) \cdot f_{X_2}(y_2 + y_1) \cdot \ldots \cdot f_{X_n}(y_n + y_1) \cdot n = \\ &= n\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(y_1 - \sum_{i=2}^{n} y_i\right)^2\right\} \times \left(\frac{1}{\sqrt{2\pi}}\right)^{n-1} \exp\left\{-\frac{1}{2}\sum_{i=2}^{n}(y_i + y_1)^2\right\}. \end{aligned}$$

Osservando che per la somma dei quadrati contenuti negli esponenti si ha

$$\left(y_1 - \sum_{i=2}^{n} y_i\right)^2 + \sum_{i=2}^{n}(y_i + y_1)^2 = ny_1^2 + \sum_{i=2}^{n} y_i^2 + \left(\sum_{i=2}^{n} y_i\right)^2,$$

si ottiene che
$f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) =$

$$\begin{aligned} &= \left(\frac{n}{2\pi}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2}ny_1^2\right\} \times \left[\frac{n}{(2\pi)^{n-1}}\right]^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\sum_{i=2}^{n} y_i^2 + \left(\sum_{i=2}^{n} y_i\right)^2\right]\right\} \\ &= f_{Y_1(y_1)} \times f_{Y_2,\ldots,Y_n}(y_2, \ldots, y_n). \end{aligned}$$

Dalla precedente fattorizzazione segue il punto (b), ovvero l'indipendenza di $Y_2, \ldots, Y_n$ da Y_1 e quindi quella di S_n^2 da $\overline{X}_n$.

3. Per dimostrare questa parte del teorema verifichiamo innanzitutto che

$$(n-1)S_n^2 = (n-2)S_{n-1}^2 + \left(\frac{n-1}{n}\right)(X_n - \overline{X}_{n-1})^2. \tag{3.9}$$

Innanzitutto si ha che

$$\begin{aligned} (n-1)S_n^2 &= \sum_{i=1}^{n-1}(X_i - \overline{X}_n)^2 + (X_n - \overline{X}_n)^2 \\ &= \sum_{i=1}^{n-1}(X_i \pm \overline{X}_{n-1} - \overline{X}_n)^2 + (X_n - \overline{X}_n)^2 \\ &= \sum_{i=1}^{n-1}(X_i - \overline{X}_{n-1})^2 + (n-1)(\overline{X}_{n-1} - \overline{X}_n)^2 + (X_n - \overline{X}_n)^2 \\ &= (n-2)S_{n-1}^2 + (n-1)(\overline{X}_{n-1} - \overline{X}_n)^2 + (X_n - \overline{X}_n)^2. \end{aligned}$$

Dalla relazione

$$\overline{X}_n = \frac{X_n + (n-1)\overline{X}_{n-1}}{n}$$

che si verifica facilmente, discende che

$$(\overline{X}_{n-1} - X_n) = n(\overline{X}_{n-1} - \overline{X}_n) \quad \text{e} \quad (X_n - \overline{X}_n) = (n-1)(X_n - \overline{X}_{n-1}).$$

Sostituendo le precedenti espressioni nell'ultima espressione trovata per $(n-1)S_n^2$ si ottiene quindi

$$\begin{aligned} (n-1)S_n^2 &= (n-2)S_{n-1}^2 + \frac{n-1}{n^2}(X_n - \overline{X}_{n-1})^2 + \frac{(n-1)^2}{n^2}(X_n - \overline{X}_{n-1})^2 \\ &= (n-2)S_{n-1}^2 + \left(\frac{n-1}{n}\right)(X_n - \overline{X}_{n-1})^2. \end{aligned}$$

Per completare la dimostrazione procediamo per induzione. Per $n = 2$, (3.9) diventa

$$S_2^2 = \frac{1}{2}(X_2 - X_1)^2.$$

In questo caso

$$(X_1 - X_2) \sim \mathrm{N}(0, 2) \quad \Rightarrow \quad \frac{1}{\sqrt{2}}(X_1 - X_2) \sim \mathrm{N}(0, 1) \quad \Rightarrow \quad \frac{1}{2}(X_1 - X_2)^2 = S_2^2 \sim \chi_1^2.$$

Supponiamo ora che per un intero k generico si abbia che $(k-1)S_k^2 \sim \chi_{k-1}^2$ e mostriamo che ciò implica che $kS_{k+1}^2 \sim \chi_k^2$. Per la (3.9) possiamo scrivere che

$$kS_{k+1}^2 = (k-1)S_k^2 + \left(\frac{k}{k+1}\right)(X_{k+1} - \overline{X}_k)^2.$$

Si osservi che

$$\begin{aligned} X_{k+1} - \overline{X}_k \sim N\left(0, 1 + \frac{1}{k}\right) \quad &\Rightarrow \quad \sqrt{\frac{k}{k+1}}(X_{k+1} - \overline{X}_k) \sim \mathrm{N}(0, 1) \\ &\Rightarrow \quad \left(\frac{k}{k+1}\right)(X_{k+1} - \overline{X}_k)^2 \sim \chi_1^2. \end{aligned}$$

Ricordando che, per ipotesi di induzione, $(k-1)S_k^2 \sim \chi_k^2$, è sufficiente mostrare l'indipendenza dei due addendi che costituiscono kS_{k+1}^2 affinchè, in virtù dell'additività del χ^2, il punto (3) risulti verificato. Ma poichè S_k^2 è indipendente da $\overline{X}_k$ (per il punto (2)) e da X_{k+1} (le osservazioni iniziali sono i.i.d.), lo è anche dalla funzione $(X_{k+1} - \overline{X}_k)$ ed il teorema è così dimostrato.

Capitolo 4

Inferenza frequentista: stima puntuale

Questo capitolo, che apre la trattazione dei metodi inferenziali frequentisti, è dedicato al problema della stima puntuale del parametro incognito di un modello statistico. Dopo aver introdotto il concetto di stimatore puntuale, il capitolo esamina: (a) due metodi per determinare stimatori puntuali; (b) un metodo per valutare e confrontare gli stimatori. Si esamina poi il concetto di ottimalità degli stimatori puntuali e, a seguire, si definiscono le proprietà asintotiche delle procedure. Si considerano inoltre le principali proprietà delle due specifiche famiglie di stimatori introdotte: gli stimatori dei momenti e gli stimatori di massima verosimiglianza. Infine vengono illustrate alcune estensioni al caso multiparametrico.

4.1 Definizioni preliminari

Per i problemi di stima puntuale si considerano particolari statistiche denominate *stimatori puntuali*. Ricordiamo inoltre la definizione di *stima*, già introdotta in 2.6.

4.1 Definizione (Stimatore e stima puntuale di un parametro). Dato il modello statistico $\{\mathcal{X}^n, f_n(\mathbf{x}_n;\theta), \theta \in \Theta\}$, una statistica

$$T : \mathcal{X}^n \to \Theta$$

che assume valori nello spazio dei parametri Θ, si chiama *stimatore* del parametro θ. Il valore $T(\mathbf{x}_n) = t$ corrispondente ad un campione osservato, si chiama *stima* del parametro. □

Uno stimatore associa a ciascun punto $\mathbf{x}_n \in \mathcal{X}^n$ un valore in Θ: si tratta di uno scalare, nel caso $k = 1$, o di un vettore, nel caso $k > 1$. La stima è quindi il valore osservato (scalare o vettoriale) della v.a. stimatore.

La definizione di stimatore è estremamente vaga. A volte, stimatori scelti su base intuitiva possono risultare molto validi. In altri casi l'intuizione può essere ingannevole o non utile a suggerire procedure di stima. E' quindi necessario avere a disposizione dei *metodi di stima*, ovvero delle procedure che, dato un modello statistico, ci permettano di determinare uno stimatore ragionevole per il parametro incognito. Inoltre è necessario definire dei *metodi di valutazione degli stimatori* da utilizzare nel caso esistano più stimatori per lo stesso parametro.

4.2 Metodi per la stima puntuale

Esistono diversi metodi per ottenere stimatori puntuali per il parametro di un modello statistico. I principali sono due: il metodo degli stimatori di massima verosimiglianza e il metodo dei momenti.

4.2.1 Stimatore dei momenti

Prima di introdurre il metodo di generazione di stimatori puntuali, richiamiamo le definizioni di *momento di ordine r* e di *momento centrale di ordine r* di una variabile aleatoria e introduciamo quelle di *momenti campionari di ordine r* e di *momenti campionari centrali di ordine r*.

Momenti e momenti campionari

4.2 Definizione (Momenti di una v.a.). Dato un modello statistico parametrico, $\{\mathcal{X}, f_X(x;\theta), \Theta\}$ per la v.a. X e un numero intero r, si chiama *momento di ordine r* di X la seguente quantità:

$$\mu_r(\theta) = \mathbb{E}_\theta\left[X^r\right] = \begin{cases} \int_{\mathcal{X}} x^r f_X(x;\theta)dx & \text{se } X \text{ è assolutamente continua} \\ \sum_{x_i \in \mathcal{X}} x_i^r f_X(x_i;\theta) & \text{se } X \text{ è discreta} \end{cases}.$$

Si chiama *momento centrale* di ordine r della v.a. X la quantità:

$$\overline{\mu}_r(\theta) = \mathbb{E}_\theta\left([X - \mathbb{E}_\theta(X)]^r\right) = \begin{cases} \int_{\mathcal{X}} [x - \mu_1(\theta)]^r f_X(x;\theta)dx \\ \sum_{x_i \in \mathcal{X}} [x_i - \mu_1(\theta)]^r f_X(x_i;\theta) \end{cases}.$$

□

Il momento di ordine $r = 1$ di X coincide con il valore atteso di X, mentre quello di ordine $r = 2$ con la media quadratica:

$$\mu_1(\theta) = \mathbb{E}_\theta[X], \qquad \mu_2(\theta) = \mathbb{E}_\theta[X^2].$$

Il più usato tra i momenti centrali, è quello di ordine 2 che corrisponde alla varianza $\mathbb{V}_\theta[X]$ della v.a. X:

$$\overline{\mu}_2(\theta) = \mathbb{E}_\theta\left[\left(X - \mathbb{E}_\theta[X]\right)^2\right] = \mathbb{V}_\theta\left[X\right].$$

Ricordando che

$$\mathbb{V}_\theta[X] = \mathbb{E}_\theta[(X - \mathbb{E}_\theta(X))^2] = \mathbb{E}_\theta[X^2] - (\mathbb{E}_\theta[X])^2,$$

possiamo osservare che tra i momenti $\mu_1(\theta)$, $\mu_2(\theta)$ e $\overline{\mu}_2(\theta)$ vale la seguente relazione:

$$\overline{\mu}_2(\theta) = \mu_2(\theta) - \mu_1(\theta)^2. \tag{4.1}$$

4.3 Osservazione I momenti e i momenti centrali di X sono funzioni del parametro incognito del modello. Si è utilizzata la notazione $\mu_r(\theta)$, $\overline{\mu}_r(\theta)$, $\mathbb{E}_\theta[X]$ e $\mathbb{V}_\theta\left[X\right]$, proprio per mettere in evidenza la dipendenza dei momenti dal parametro θ, che indicizza la famiglia di distribuzioni per il modello $\{\mathcal{X}, f_X(x;\theta), \Theta\}$. □

4.4 Esempio (Modello uniforme). Sia X una v.a. con ha distribuzione uniforme nell'intervallo $[0, \theta]$. Il momento primo di X è:

$$\mu_1(\theta) = \mathbb{E}_\theta[X] = \frac{1}{\theta}\int_0^\theta x dx = \frac{1}{\theta}\left[\frac{x^2}{2}\right]_0^\theta = \frac{\theta}{2}.$$

Per $r > 1$,

$$\mu_r(\theta) = \frac{1}{\theta}\int_0^\theta x^r dx = \frac{1}{\theta}\left[\frac{x^{r+1}}{r+1}\right]_0^\theta = \frac{\theta^r}{r+1}.$$

Inoltre, ricordando che $\overline{\mu}_2(\theta) = \mu_2(\theta) - \mu_1(\theta)^2$, si ricava anche che

$$\overline{\mu}_2(\theta) = \mathbb{V}_\theta[X] = \frac{\theta^2}{3} - \frac{\theta^2}{4} = \frac{\theta^2}{12}.$$

□

4.5 Esempio (Modello beta). Si consideri il modello statistico in cui la variabile di base X ha distribuzione di tipo Beta con parametri $\theta_1 > 0$ e $\theta_2 > 0$ entrambi incogniti. Per questo modello $\boldsymbol{\theta} = (\theta_1, \theta_2)$ e la la funzione di densità è:

$$f_X(x;\theta) = \frac{\Gamma(\theta_1+\theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)}\, x^{\theta_1-1}(1-x)^{\theta_2-1} \quad 0 < x < 1.$$

In tal caso lo spazio parametrico $\Theta = (0,\infty)\times(0,\infty) \subset \mathbb{R}^2$ ha dimensione 2 e dunque il metodo dei momenti ha bisogno di due equazioni nelle due incognite $\theta = (\theta_1, \theta_2)$.
Si può dimostrare che in questo caso le espressioni per i momenti di ordine r sono:

$$\mu_r(\theta) = \mu_r(\theta_1,\theta_2) = \frac{\theta_1(\theta_1+1)\dots(\theta_1+r-1)}{(\theta_1+\theta_2)(\theta_1+\theta_2+1)\dots(\theta_1+\theta_2+r-1)}.$$

□

4.6 Definizione (Momenti campionari). Dato un campione $X_1, \dots, X_n$ di dimensione n e un numero intero r, si chiama *momento campionario di ordine r* la statistica campionaria

$$m_r(\mathbf{X}_n) = \frac{1}{n}\sum_{i=1}^n (X_i)^r$$

Si chiama *momento centrale campionario di ordine r* la statistica campionaria

$$\overline{m}_r(\mathbf{X}_n) = \frac{1}{n}\sum_{i=1}^n \left(X_i - \overline{X}_n\right)^r.$$

□

I momenti campionari di ordine $r = 1$ e $r = 2$ coincidono rispettivamente con la media aritmetica e la media quadratica campionaria:

$$m_1(\mathbf{X}_n) = \overline{X}_n, \qquad m_2(\mathbf{X}_n) = \frac{1}{n}\sum_{i=1}^n (X_i)^2.$$

Il momento secondo centrale campionario coincide con la varianza campionaria:

$$\overline{m}_2(\mathbf{X}_n) = \widehat{\sigma}^2_n = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2.$$

Inoltre, osservando che

$$\widehat{\sigma}^2_n = \frac{1}{n}\sum_{i=1}^n (X_i)^2 - \overline{X}_n{}^2,$$

vale la relazione

$$\overline{m}_2(\mathbf{X}_n) = m_2(\mathbf{X}_n) - m_1(\mathbf{X}_n)^2. \tag{4.2}$$

4.7 Osservazione In corrispondenza di un campione osservato $\mathbf{x}_n$, le quantità $m_r(\mathbf{x}_n)$ e $\overline{m}_r(\mathbf{x}_n)$ sono le realizzazioni delle v.a. $m_r(\mathbf{X}_n)$ e $\overline{m}_r(\mathbf{X}_n)$. A differenza dei momenti $\mu_r(\theta)$ e $\overline{\mu}_r(\theta)$ di X, i momenti campionari $m_r(\mathbf{X}_n)$ e $\overline{m}_r(\mathbf{X}_n)$ sono pertanto quantità osservabili. □

Metodo dei momenti

Possiamo ora illustrare il metodo dei momenti, introdotto da Karl Pearson a fine '800. L'idea è di ottenere uno stimatore per il parametro vettoriale $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ come soluzione di un sistema di k equazioni in k incognite ottenuto imponendo l'uguaglianza dei primi k momenti $\mu_r(\theta)$ con i primi k momenti campionari $m_r(\mathbf{X}_n)$.

4.8 Definizione (Stimatore dei momenti). Dato il modello statistico $\{\mathcal{X}^n, f_n(\cdot;\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$, con $\Theta \subseteq \mathbb{R}^k$, si chiama *sistema dei momenti* il seguente sistema di k equazioni

$$\left\{ \begin{array}{lcl} \mu_1(\theta_1, \ldots, \theta_k) & = & m_1(\mathbf{X}_n) = \frac{1}{n}\sum_{i=1}^n X_i \\ \ldots & & \ldots \\ \mu_r(\theta_1, \ldots, \theta_k) & = & m_r(\mathbf{X}_n) = \frac{1}{n}\sum_{i=1}^n X_i^r \\ \ldots & & \ldots \\ \mu_k(\theta_1, \ldots, \theta_k) & = & m_k(\mathbf{X}_n) = \frac{1}{n}\sum_{i=1}^n X_i^k \end{array} \right. \tag{4.3}$$

Se la soluzione (rispetto alle variabili $\theta_1, \ldots, \theta_k$) del sistema (4.3) esiste ed è unica, allora essa è una funzione (vettoriale) dei momenti campionari, $\mathbf{g}(m_1, \ldots, m_k)$, e viene indicata con

$$\widehat{\boldsymbol{\theta}}_m(\mathbf{X}_n) = \mathbf{g}(m_1, \ldots, m_k) = (\widehat{\theta}_{m,1}(\mathbf{X}_n), \ldots, \widehat{\theta}_{m,k}(\mathbf{X}_n))$$

e viene chiamata *stimatore dei momenti* del parametro $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$. □

Nel caso $k = 1$ il sistema (4.3) diventa l'equazione

$$\mu_1(\theta) = \overline{X}_n. \tag{4.4}$$

Nel caso $k = 2$ il sistema è

$$\left\{ \begin{array}{lcl} \mu_1(\theta_1, \theta_2) & = & \frac{1}{n}\sum_{i=1}^n X_i \\ \mu_2(\theta_1, \theta_2) & = & \frac{1}{n}\sum_{i=1}^n X_i^2 \end{array} \right. . \tag{4.5}$$

Ricordando che $\mu_1(\theta_1, \theta_2) = \mathbb{E}_\theta[X]$, $\mu_2(\theta_1, \theta_2) = \mathbb{E}_\theta[X^2] = \mathbb{V}_\theta[X] + \mathbb{E}_\theta[X]^2$, $m_1(\mathbf{X}_n) = \overline{X}_n$ e $m_2(\mathbf{X}_n) = \widehat{\sigma}_n^2 + \overline{X}_n{}^2$, il sistema diventa:

$$\left\{ \begin{array}{lcl} \mathbb{E}_\theta[X] & = & \overline{X}_n \\ \mathbb{V}_\theta[X] & = & \widehat{\sigma}_n^2 \end{array} \right. . \tag{4.6}$$

4.9 Esempio. Per i modelli Ber(θ), Pois(θ), Esp(θ), N(θ, σ_0^2) (varianza nota), per i quali $\mu_1(\theta) = \mathbb{E}_\theta[X] = \theta$, si ha banalmente che l'equazione (4.4) diventa

$$\theta = \overline{X}_n.$$

In tutti questi casi, $\widehat{\theta}_m(\mathbf{X}_n) = \overline{X}_n$. Si osservi che lo stimatore dei momenti coincide in tutti questi casi con lo stimatore di massima verosimiglianza di θ nei rispettivi modelli. □

4.10 Esempio (Modello uniforme). Se consideriamo un modello uniforme in $[0, \theta]$, l'equazione (4.4) diventa:

$$\frac{\theta}{2} = \overline{X}_n$$

e quindi lo stimatore dei momenti è $\widehat{\theta}_m(\mathbf{X}_n) = 2\overline{X}_n$. In questo caso lo stimatore dei momenti non coincide con lo stimatore di massima verosimiglianza che, come si è visto, risulta essere $\widehat{\theta}_{mv}(\mathbf{X}_n) = X_{(n)}$. Si noti che, in questo caso, lo stimatore dei momenti potrebbe assumere valori esterni all'intervallo $[0, \theta]$. □

4.11 Esempio (Modello beta). Sia $\mathbf{X}_n = (X_1, \ldots, X_n)$ un campione casuale proveniente da una popolazione Beta$(\theta, 1)$. In questo caso

$$f_X(x; \theta) = \theta\, x^{\theta-1}, \quad x \in [0, 1], \quad \theta > 0.$$

È semplice verificare che

$$\mathbb{E}_\theta[X] = \frac{\theta}{\theta + 1} \quad \Rightarrow \quad \widehat{\theta}_m(\mathbf{X}_n) = \frac{\overline{X}_n}{1 - \overline{X}_n}.$$

In questo caso si mostra che $\widehat{\theta}_{mv}(\mathbf{X}_n) = -n/(\sum_{i=1}^n \ln X_i) \neq \widehat{\theta}_m(\mathbf{X}_n)$

Riprendiamo in considerazione il caso generale, ovvero consideriamo il modello statistico in cui la variabile di base X abbia distribuzione di tipo Beta con parametri $\theta_1 > 0$ e $\theta_2 > 0$, entrambi incogniti. In tal caso lo spazio parametrico $\Theta = (0, \infty) \times (0, \infty) \subset \mathbb{R}^2$ ha dimensione 2 e dunque il metodo dei momenti ha bisogno di due equazioni nelle due incognite $\theta = (\theta_1, \theta_2)$.
Si può dimostrare che in questo caso le espressioni per i momenti di ordine r sono:

$$\mu_r(\theta) = \mu_r(\theta_1, \theta_2) = \frac{\theta_1(\theta_1 + 1)\ldots(\theta_1 + r - 1)}{(\theta_1 + \theta_2)(\theta_1 + \theta_2 + 1)\ldots(\theta_1 + \theta_2 + r - 1)}$$

e dunque il sistema da risolvere sarà:

$$\left\{ \begin{array}{rcccl} \frac{\theta_1}{\theta_1+\theta_2} & = & \mu_1(\theta) & = & m_1(x_1, ..., x_n) = \overline{x}_1 = \frac{1}{n}\sum_{i=1}^n x_i \\ \frac{\theta_1(\theta_1+1)}{(\theta_1+\theta_2)(\theta_1+\theta_2+1)} & = & \mu_2(\theta) & = & m_2(x_1, ..., x_n) = \overline{x}_2 = \frac{1}{n}\sum_{i=1}^n x_i^2 \end{array} \right.$$

Tale sistema può essere risolto attraverso un'opportuna successione sistemi equivalenti ottenuti attraverso manipolazioni algebriche

$$\left\{ \begin{array}{rcl} \theta_1 & = & (\theta_1 + \theta_2)\overline{x}_1 \\ \frac{\theta_1}{\theta_1+\theta_2}\frac{\theta_1+1}{(\theta_1+\theta_2)+1} & = & \overline{x}_2 \end{array} \right.$$

e semplici sostituzioni come ad esempio $\overline{x}_1$ al posto di $\theta_1/(\theta_1+\theta_2)$ e di $\frac{\theta_1}{\overline{x}_1}$ al posto di $(\theta_1+\theta_2)$ nella seconda equazione da cui

$$\left\{ \begin{array}{rcl} \theta_1(1 - \overline{x}_1) & = & \theta_2\overline{x}_1 \\ \overline{x}_1\frac{\theta_1+1}{\frac{\theta_1}{\overline{x}_1}+1} & = & \overline{x}_2 \end{array} \right. \iff \left\{ \begin{array}{rcl} \theta_2 & = & \theta_1\frac{1-\overline{x}_1}{\overline{x}_1} \\ \overline{x}_1\frac{\theta_1+1}{\frac{\theta_1+\overline{x}_1}{\overline{x}_1}} & = & \overline{x}_2 \end{array} \right.$$

e quindi

$$\left\{ \begin{array}{rcl} \theta_2 & = & \theta_1\frac{1-\overline{x}_1}{\overline{x}_1} \\ \overline{x}_1^2(\theta_1 + 1) & = & \overline{x}_2(\theta_1 + \overline{x}_1) \end{array} \right. \iff \left\{ \begin{array}{rcl} \theta_2 & = & \theta_1\frac{1-\overline{x}_1}{\overline{x}_1} \\ \theta_1(\overline{x}_1^2 - \overline{x}_2) & = & \overline{x}_1\overline{x}_2 - \overline{x}_1^2 \end{array} \right.$$

e infine

$$\left\{\begin{array}{lcl}\theta_2 &=& \frac{1-\overline{x}_1}{\overline{x}_1}\overline{x}_1\,\dfrac{\overline{x}_2-\overline{x}_1}{\overline{x}_1^2-\overline{x}_2} = \\ \theta_1 &=& \overline{x}_1\,\dfrac{\overline{x}_2-\overline{x}_1}{\overline{x}_1^2-\overline{x}_2} = \overline{x}_1\,\dfrac{\overline{x}_1-\overline{x}_2}{\overline{x}_2-\overline{x}_1^2}\end{array}\right. \quad\Longleftrightarrow\quad \left\{\begin{array}{lcl}{\theta_2}^* &=& (1-\overline{x}_1)\,\dfrac{\overline{x}_1-\overline{x}_2}{\overline{x}_2-\overline{x}_1^2} \\ {\theta_1}^* &=& \overline{x}_1\,\dfrac{\overline{x}_1-\overline{x}_2}{\overline{x}_2-\overline{x}_1^2}\end{array}\right.$$

Lo stimatore dei momenti per $\theta=(\theta_1,\theta_2)$ sarà dunque dato dalla statistica vettoriale

$$\widehat{\theta}_M=\widehat{\theta}_M(X_1,...,X_n)=(\widehat{\theta}_{M,1}(X_1,...,X_n),\widehat{\theta}_{M,2}(X_1,...,X_n))=\left(\overline{X}_1\,\frac{\overline{X}_1-\overline{X}_2}{\overline{X}_2-\overline{X}_1^2},(1-\overline{X}_1)\,\frac{\overline{X}_1-\overline{X}_2}{\overline{X}_2-\overline{X}_1^2}\right).$$

□

4.12 Esempio (Modello normale). Consideriamo il modello $\mathrm{N}(\theta_1,\theta_2)$. In questo caso $\mu_1(\theta)=\theta_1$ e $\mu_2(\theta)=\theta_2+\theta_1^2$. Inoltre, per la (4.2), $m_2(\mathbf{X}_n)=\widehat{\sigma}_n^2+{\overline{X}_n}^2$. Il sistema dei momenti diventa quindi

$$\left\{\begin{array}{lcl}\theta_1 &=& \overline{X}_n \\ \theta_2+\theta_1^2 &=& \widehat{\sigma}_n^2+{\overline{X}_n}^2\end{array}\right. . \tag{4.7}$$

La soluzione del sistema (4.6) è quindi il vettore $\widehat{\theta}_m(\mathbf{X}_n)=\left(\overline{X}_n,\widehat{\sigma}_n^2\right)$. □

4.13 Esempio (Modello gamma). Sia $X\sim\mathrm{Gamma}(\theta_1,\theta_2)$. In questo caso, ricordando che $\mathbb{E}_\theta[X]=\theta_1/\theta_2$ e $\mathbb{V}_\theta[X]=\theta_1/\theta_2^2$, il sistema dei momenti diventa

$$\left\{\begin{array}{lcl}\frac{\theta_1}{\theta_2} &=& \overline{X}_n \\ \frac{\theta_1}{\theta_2^2} &=& \widehat{\sigma}_n^2\end{array}\right. . \tag{4.8}$$

La soluzione del sistema dei momenti è quindi il vettore $\widehat{\boldsymbol{\theta}}_m(\mathbf{X}_n)=\left({\overline{X}_n}^2/\widehat{\sigma}_n^2,\overline{X}_n/\widehat{\sigma}_n^2\right)$.

□

4.14 Osservazione Il metodi dei momenti fornisce spesso stimatori piuttosto grossolani per i parametri del modello. A volte i valori di stima cadono al fuori dello spazio parametrico (violando cosí la definizione stessa di stimatore che abbiamo dato). □

4.15 Esempio (Modello binomiale). Assumiamo che $X_1,\ldots,X_n$ siano i.i.d $\mathrm{Binom}(k,p)$, ovvero

$$f_{X_i}(x|k,p)=\binom{k}{x}p^k(1-p)^{k-x},\qquad x=0,1,\ldots,k,\quad i=1,\ldots,n.$$

Supponiamo inoltre che siano incogniti sia k che p. Il questo caso $\boldsymbol{\theta}=(k,p)$, $\mu_1(\boldsymbol{\theta})=kp$ e $\mu_2(\boldsymbol{\theta})=kp(1-p)+k^2p^2$. Il sistema dei momenti diventa quindi

$$\left\{\begin{array}{lcl}p &=& \overline{X}_n/k \\ kp(1-p)+k^2p^2 &=& \widehat{\sigma}_n^2+{\overline{X}_n}^2\end{array}\right. \tag{4.9}$$

la cui soluzione è

$$\left\{\begin{array}{lcl}\widehat{p}_m &=& \overline{X}_n/\widehat{k}_m \\ \widehat{k}_m &=& {\overline{X}_n}^2/(\overline{X}_n-\widehat{\sigma}_n^2)\end{array}\right. . \tag{4.10}$$

Si noti che $\widehat{k}_m$ e $\widehat{p}_m$ assumono valori negativi quando $\overline{X}_n < \widehat{\sigma}^2_n$, ovvero in presenza di forte variabilità campionaria. □

4.16 Osservazione La definizione data del metodo dei momenti, che prevede di considerare i *primi* k momenti, può non condurre ad un sistema in cui siano presenti i k parametri incogniti θ_i del modello. In questo caso è necessario includere nel sistema equazioni del tipo $\mu_r(\theta) = m_r(\mathbf{X}_n),\ r > k$, in modo da ottenere un sistema con soluzione. Può anche succedere che, in un modello con un solo parametro incognito, l'equazione da considerare non sia la prima, ovvero $\mu_1(\theta) = m_1(\mathbf{X}_n)$. □

4.17 Esempio (Modello normale, valore atteso noto). Sia $X_1, \ldots, X_n$ i.i.d. $\mathrm{N}(\mu_0, \theta)$, in cui il valore atteso delle X_i è noto. In questo caso la prima equazione dei momenti coincide con l'uguaglianza

$$\mu_0 = \overline{X}_n$$

che non dipende dal parametro incognito ed è quindi inutile per trovare uno stimatore di θ. La seconda equazione $\theta + \mu_0^2 = \sum_{i=1}^n X_i^2/n$ porta allo stimatore

$$\widehat{\theta}_m = \frac{1}{n}\sum_{i=1}^n X_i^2 - \mu_0^2$$

che, come nell'esempio precedente, può essere negativo. Nel caso particolare in cui $\mu_0 = 0$, si ottiene che $\widehat{\theta}_m = \widehat{\theta}_{mv} = \sum_{i=1}^n X_i^2/n$. In questo esempio potremmo pensare di ottenere uno stimatore della varianza imponendo l'uguaglianza dei momenti centrali di ordine due, $\overline{\mu}_2(\theta) = \overline{m}_2(\mathbf{X}_n)$. In questo caso si ottiene $\widehat{\theta}'_m = \sum_{i=1}^n (X_i - \overline{X}_n)^2/n$. Neanche in questo modo si ottiene la stima piú intuitiva della varianza, che si è invece ottenuta con il metodo della massima verosimiglianza, ovvero la stima $S_0^2 = \sum_{i=1}^n (X_i - \mu_0)^2/n$. □

4.2.2 Stimatore di massima verosimiglianza

Nel capitolo 3 abbiamo introdotto il concetto di verosimiglianza e di stima di massima verosimiglianza. Lo stimatore di massima verosimiglianza è la v.a. il cui valore osservato in corrispondenza di un campione $\mathbf{x}_n$ è una stima di massima verosimiglianza. Ad esempio, per i modelli $\mathrm{Ber}(\theta)$, $\mathrm{Pois}(\theta)$, $\mathrm{N}(\theta, 1)$ abbiamo verificato che la stima di massima verosimiglianza di θ è, per ciascuno di questi modelli, la media aritmetica dei valori campionari osservati. In questi casi lo stimatore di massima verosimiglianza è la v.a $\overline{X}_n$. Piú formalmente, possiamo definire lo stimatore di massima verosimiglianza come segue.

4.18 Definizione (Stimatore di massima verosimiglianza). Sia $\{\mathcal{X}^n, f_n(\mathbf{x}_n; \theta), \theta \in \Theta\}$ un modello statistico per il campione $X_1, \ldots, X_n$. Per ogni $\mathbf{x}_n \in \mathcal{X}^n$ sia $L(\theta; \mathbf{x}_n)$ la funzione di verosimiglianza di θ associata al campione osservato e $\widehat{\theta}_{mv}(\mathbf{x}_n)$ una corrispondente stima di massima verosimiglianza

$$\widehat{\theta}_{mv}(\mathbf{x}_n) = \arg\sup_{\theta \in \Theta} L(\theta; \mathbf{x}_n).$$

Si chiama *stimatore di massima verosimiglianza di* θ la v.a. $\widehat{\theta}_{mv}(\mathbf{X}_n)$. □

Per la determinazione degli stimatori di mv per i parametri dei principali modelli, si rimanda al Capitolo 3. La discussione delle proprietà frequentiste è invece rinviata al Paragrafo 4.7.

4.3 Valutazione degli stimatori

Nel paragrafo precedente sono stati illustrati due tra i possibili metodi per la generazione di stimatori puntuali. La pluralità di metodi per trovare stimatori fa sí che siano disponibili stimatori distinti per lo stesso parametro. Ad esempio, nel caso del modello uniforme in $[0, \theta]$, abbiamo visto che $\widehat{\theta}_m = 2\overline{X}_n$ mentre $\widehat{\theta}_{mv} = X_{(n)}$. È evidente che questi due stimatori danno luogo, in generale, a stime diverse del parametro incognito. È quindi necessario definire dei criteri che consentano il confronto tra stimatori alternativi e, in ogni caso, la valutazione della qualità di un singolo stimatore.

L'approccio piú naturale per il confronto di diversi metodi per risolvere un problema inferenziale è quello offerto dalla *teoria statistica delle decisioni*. Nel caso della stima puntuale, l'idea alla base di questa teoria[1] è piuttosto semplice. Si assume che a ogni possibile stima $d(\mathbf{x}_n)$ di θ (detta *funzione di decisione*) sia associata una perdita, dovuta al fatto che il valore osservato della stima non coincide esattamente con quello del parametro da stimare. La funzione che misura la perdita associata alla stima $d(\mathbf{x}_n)$ di θ si chiama *funzione di perdita* di d:

$$\mathbb{L}(\theta, d(\mathbf{x}_n))$$

Si ipotizza che, se $d(\mathbf{x}_n) \neq \theta$, $\mathbb{L}(\theta, d(\mathbf{x}_n)) > 0$ e che $\mathbb{L}(\theta, d(\mathbf{x}_n)) = 0$ se $d(\mathbf{x}_n) = \theta$. Esempi classici di funzioni di perdita sono le funzioni

$$\begin{aligned} \mathbb{L}(\theta, d(\mathbf{x}_n)) &= |\theta - d(\mathbf{x}_n)| \quad \text{perdita assoluta,} \\ \mathbb{L}(\theta, d(\mathbf{x}_n)) &= [\theta - d(\mathbf{x}_n)]^2 \quad \text{perdita quadratica.} \end{aligned}$$

Per valutare lo stimatore $d(\mathbf{X}_n)$ di θ, che è una variabile aleatoria, possiamo allora usare la perdita aleatoria

$$\mathbb{L}(\theta, d(\mathbf{X}_n)).$$

Nell'ottica del Principio del campionamento ripetuto, la valutazione dello stimatore $d(\mathbf{X}_n)$ deve coinvolgere la distribuzione di probabilità della perdita aleatoria $\mathbb{L}(\theta, d(\mathbf{X}_n))$. Il criterio piú comune per valutare $d(\mathbf{X}_n)$ si basa sul valore atteso di $\mathbb{L}(\theta, d(\mathbf{X}_n))$. Questo valore atteso prende il nome di *funzione di rischio*:

$$R(\theta, d) = \mathbb{E}_\theta[\mathbb{L}(\theta, d(\mathbf{X}_n))],$$

dove il valore atteso si intende effettuato rispetto alla distribuzione campionaria, $f_n(\cdot, \theta)$. La funzione di rischio di uno stimatore d, per definizione, dipende dal parametro incognito. Per ogni stimatore d, si ha una funzione di θ che descrive la variazione del valore atteso della perdita associata a d, in funzione del parametro θ.

La funzione di rischio è quindi una sintesi (rispetto alla variabilità campionaria) della quantità aleatoria $\mathbb{L}(\theta, d(\mathbf{X}_n))$ e costituisce, in termini decisionali, un *criterio* per il confronto di stimatori alternativi e per la ricerca di eventuali decisioni ottime.

4.3.1 Errore quadratico medio di uno stimatore

La funzione di perdita piú comunemente utilizzata nei problemi di stima puntuale di un parametro è la perdita quadratica. In questo caso la funzione di rischio di d prende il nome di *errore quadratico medio* o, in inglese, *mean squared error*. Consideriamo il caso in cui θ è uno scalare.

[1] Diamo qui solo alcuni cenni di questa teoria, senza ulteriori approfondimenti che andrebbero aldilà degli obiettivi di questo testo.

4.19 Definizione (Errore quadratico medio). Dato il modello statistico $\{\mathcal{X}^n, f_n(\mathbf{x}_n;\theta), \theta \in \Theta\}$ e un campione $X_1, \ldots, X_n$, si chiama *errore quadratico medio* dello stimatore d di θ la funzione

$$\mathrm{MSE}_\theta(d) = \mathbb{E}_\theta[(d(\mathbf{X}_n) - \theta)^2].$$

□

Indicata con $f_d(\cdot;\theta)$ la legge di probabilità di $d(\mathbf{X}_n)$, la funzione MSE di d è

$$\mathrm{MSE}_\theta(d) = \sum_i (t_i - \theta)^2 f_d(t_i;\theta)$$

nel caso in cui $d(\mathbf{X}_n)$ è una v.a. discreta che assume valori t_i. Nel caso in cui $d(\mathbf{X}_n)$ è una v.a. assolutamente continua abbiamo

$$\mathrm{MSE}_\theta(d) = \int_{\mathbb{R}} (t - \theta)^2 f_d(t;\theta) dt.$$

L'interpretazione e il calcolo di $\mathrm{MSE}_\theta(d)$ sono agevolati dalla seguente proprietà.

4.20 Teorema (Scomposizione dell'errore quadratico medio). Dato il modello statistico $\{\mathcal{X}^n, f_n(\mathbf{x}_n;\theta), \theta \in \Theta\}$ e un generico stimatore d di θ, si ha che

$$\mathrm{MSE}_\theta(d) = \mathbb{E}_\theta[(d(\mathbf{X}_n) - \theta)^2] = \mathbb{V}_\theta[d(\mathbf{X}_n)] + B_\theta[d(\mathbf{X}_n)]^2,$$

dove la funzione

$$B_\theta[d(\mathbf{X}_n)] = \mathbb{E}_\theta[d(\mathbf{X}_n)] - \theta$$

è denominata *distorsione* (*bias*, in inglese) dello stimatore d. □

Dimostrazione. Per semplicità di notazione indichiamo solo con d la v.a. $d(\mathbf{X}_n)$. Il risultato si ottiene osservando che

$$\begin{aligned} \mathrm{MSE}_\theta(d) &= \mathbb{E}_\theta[(d-\theta)^2] = \mathbb{E}_\theta[(d \pm \mathbb{E}_\theta(d) - \theta)^2] \\ &= \mathbb{E}_\theta[(d - \mathbb{E}_\theta(d))^2] + (\mathbb{E}_\theta(d) - \theta)^2 + 2(\mathbb{E}_\theta(d) - \theta)\mathbb{E}_\theta[d - \mathbb{E}_\theta(d)] \\ &= \mathbb{V}_\theta(d) + [B_\theta(d)]^2 \end{aligned}$$

dove l'ultima uguaglianza si ottiene osservando che $\mathbb{E}_\theta[d - \mathbb{E}_\theta(d)] = 0$. □

La funzione MSE_θ incorpora quindi due componenti: la varianza dello stimatore e il quadrato della distorsione. Per i valori di θ in cui $\mathrm{MSE}_\theta(d)$ assume valori non elevati, lo stimatore d presenta bassi valori di varianza e distorsione. Il controllo dell'errore quadratico medio coincide quindi con il controllo congiunto di una misura di variabilità e di una misura di accuratezza (vicinanza in media a θ) dello stimatore d.

4.21 Definizione (Stimatore non distorto). Uno stimatore $d(\mathbf{X}_n)$ del parametro θ si definisce *non distorto* (oppure *corretto*) se

$$\mathbb{E}_\theta[d(\mathbf{X}_n)] = \theta, \qquad \forall\ \theta \in \Theta.$$

ovvero se $B_\theta(d) = 0$ per ogni $\theta \in \Theta$. □

4.22 Osservazione. Si noti che la definizione di non distorsione richiede che l'uguaglianza tra il valore atteso di $d(\mathbf{X}_n)$ e θ valga *per ogni* valore del parametro stesso. □

4.23 Osservazione. Il concetto di non distorsione (o correttezza) è centrale nell'impostazione inferenziale frequentista. L'idea (un po' imprecisa ma pragmatica) è la seguente: uno stimatore è non distorto per θ se, supponendo di poterne calcolare il valore in corrispondenza di tutti i campioni dello spazio dei campioni, la media aritmetica dei valori di stima ottenuti coincide con il valore vero del parametro. È evidente che tale proprietà di uno stimatore ha rilievo autonomo particolare solo se si adotta il Principio del campionamento ripetuto. A differenza di quanto avviene in ottica frequentista, nelle impostazioni basate sulla funzione di verosimiglianza e in quella bayesiana la non distorsione di uno stimatore non è di per sè una proprietà (rilevante) della procedura di stima puntuale. □

4.24 Esempio (Media campionaria stimatore non distorto). La statistica $\overline{X}_n$ è uno stimatore non distorto del parametro θ per molti dei modelli che abbiamo considerato. Infatti, per i modelli Ber(θ), Pois(θ), Exp(θ), N(θ, σ^2), si ha che $\widehat{\theta}_m(\mathbf{X}_n) = \widehat{\theta}_{mv}(\mathbf{X}_n) = \overline{X}_n$ e che

$$\mathbb{E}_\theta[\overline{X}_n] = \theta, \quad \forall \theta \in \Theta.$$

In tutti questi casi $\text{MSE}_\theta[\overline{X}_n]$ coincide con la varianza di $\overline{X}_n$. Per il modello Ber(θ) abbiamo che $\text{MSE}_\theta[\overline{X}_n] = \theta(1-\theta)/n$, per il modello Pois($\theta$), $\text{MSE}_\theta[\overline{X}_n] = \theta/n$, per il modello Exp($\theta$), $\text{MSE}_\theta[\overline{X}_n] = \theta^2/n$. Per il modello N($\theta, \sigma^2$), l'errore quadratico medio è σ^2/n e quindi non dipende dal parametro θ. In tutti i casi considerati la funzione MSE dipende dalla numerosità campionaria, n. Per ogni valore fissato di θ, MSE_θ è funzione decrescente della numerosità campionaria (in particolare, il valore di MSE_θ tende a zero per $n \to \infty$, qualunque sia il valore di θ).

Si noti che, nel caso bernoulliano, indipendentemente dalla numerosità campionaria, la funzione MSE_θ ha un punto di massimo in $\theta = 1/2$ e vale zero in $\theta = 0$ e $\theta = 1$.

□

4.25 Esempio (Modello uniforme). Nel caso di n v.a. X_i con distribuzione uniforme in $[0, \theta]$, $\theta > 0$ la media campionaria è uno stimatore distorto. Si ha infatti che $\mathbb{E}_\theta[\overline{X}_n] = \mathbb{E}_\theta[X_i] = \theta/2$, $\forall \theta \in [0, 1]$. È invece non distorto lo stimatore dei momenti, $\widehat{\theta}_m(\mathbf{X}_n) = 2\overline{X}_n$. Per questo stimatore si ha

$$\text{MSE}_\theta[\widehat{\theta}_m] = \mathbb{V}_\theta[2\overline{X}_n] = 4\frac{\mathbb{V}_\theta[X_i]}{n} = \frac{\theta^2}{3n}.$$

Risulta invece distorto lo stimatore di massima verosimiglianza, in quanto $\mathbb{E}_\theta[X_{(n)}] = \frac{n}{n+1}\theta$, mentre è non distorto lo stimatore $\frac{n+1}{n}X_{(n)}$:

$$\mathbb{E}_\theta\left[\frac{n+1}{n}X_{(n)}\right] = \frac{n+1}{n}\mathbb{E}_\theta[X_{(n)}] = \frac{n+1}{n}\frac{n}{n+1}\theta = \theta, \qquad \forall \theta > 0.$$

□

4.3.2 Confronto tra stimatori

Come detto sopra, la funzione di rischio $R(\theta, d)$ e quindi, nel caso di perdita quadratica, l'errore quadratico medio, consentono:

a) il confronto relativo tra coppie di stimatori;

b) la definizione di stimatore *ottimo*.

4.26 Definizione (Stimatore migliore). Dati due stimatori d_1 e d_2 del parametro θ di un modello statistico, lo stimatore d_1 è *migliore* dello stimatore d_2 ($d_1 \succ d_2$) se

$$\begin{aligned} R(\theta, d_1) &\leq R(\theta, d_2), \qquad \forall\ \theta \in \Theta \\ R(\theta, d_1) &\neq R(\theta, d_2). \end{aligned}$$

Se si utilizza la funzione di perdita quadratica, d_1 è migliore di d_2 se

$$\begin{aligned} \mathrm{MSE}_\theta(d_1) &\leq \mathrm{MSE}_\theta(d_2), \qquad \forall\ \theta \in \Theta \\ \mathrm{MSE}_\theta(d_1) &\neq \mathrm{MSE}_\theta(d_2). \end{aligned}$$

□

4.27 Osservazione Nella definizione appena data si intende che uno stimatore è migliore di un altro se la funzione di rischio del primo risulta minore o uguale di quella del secondo per ogni valore di θ e se esiste almeno un valore di θ per il quale la disuguaglianza vale in senso stretto.

□

4.28 Esempio (Confronto tra stimatori, modello uniforme). Sia $X_1, \ldots, X_n$ un campione casuale da una popolazione uniforme in $[0, \theta]$, $\theta > 0$. In questo caso lo stimatore di massima verosimiglianza e lo stimatore dei momenti non coincidono. Abbiamo infatti che

$$\widehat{\theta}_{mv}(\mathbf{X}_n) = X_{(n)} \quad \text{e} \quad \widehat{\theta}_m(\mathbf{X}_n) = 2\overline{X}_n.$$

Possiamo ora confrontarli con il criterio MSE. Abbiamo verificato in precedenza che

$$\mathbb{E}_\theta[X_{(n)}] = \frac{n\theta}{n+1}, \quad \text{e} \quad \mathbb{V}_\theta[X_{(n)}] = \frac{n\theta^2}{(n+1)^2(n+2)}.$$

Lo stimatore $\widehat{\theta}_{mv}$ è quindi distorto con errore quadratico medio

$$\mathrm{MSE}_\theta[\widehat{\theta}_{mv}] = \mathbb{V}_\theta[X_{(n)}] + B_\theta[X_{(n)}]^2 = \frac{n\theta^2}{(n+1)^2(n+2)} + \left(\frac{n\theta}{n+1} - \theta\right)^2 = \frac{2\theta^2}{(n+1)(n+2)}.$$

Lo stimatore dei momenti è invece non distorto con

$$\mathrm{MSE}_\theta[\widehat{\theta}_m] = \mathbb{V}_\theta[2\overline{X}_n] = 4\frac{\mathbb{V}_\theta[X_i]}{n} = \frac{\theta^2}{3n}.$$

Si ha quindi che

$$\mathrm{MSE}_\theta(X_{(n)}) < \mathrm{MSE}_\theta(2\overline{X}_n) \quad \Leftrightarrow \quad \frac{2\theta^2}{(n+1)(n+2)} < \frac{\theta^2}{3n}.$$

La precedente disequazione è verificata per ogni $\theta > 0$ per i valori di n tali che $n^2 + 4 > 0$, ovvero per ogni $n \in \mathbb{N}$. □

4.4 Stimatori ottimi

L'obiettivo della teoria frequentista della stima puntuale è trovare stimatori *ottimi* per il parametro di un modello.

4.29 Definizione (Stimatore ottimo). Dato il modello statistico $\{\mathcal{X}^n, f_n(\mathbf{x}_n;\theta), \theta \in \Theta\}$ e la funzione di perdita $\mathbb{L}(\theta, d)$, lo stimatore d^* è uno *stimatore ottimo* rispetto alla funzione di rischio $R(\theta,\cdot)$ se se $\forall d \in D$,

$$R(\theta, d^*) \leq R(\theta, d), \quad \forall \theta \in \Theta,$$

dove D rappresenta la classe di tutti gli stimatori di θ. □

Si noti che, in base alla definizione data, lo stimatore ottimo può non esistere e, se esiste, può non essere unico. Inoltre, nel caso specifico in cui la $\mathbb{L}(d, \theta)$ è la funzione di perdita quadratica, uno stimatore ottimo minimizza l'errore quadratico medio:

$$\mathrm{MSE}_\theta(d^*) \leq \mathrm{MSE}_\theta(d), \quad \forall \theta \in \Theta, \quad \forall d \in D.$$

Si ha quindi che

$$d^* \succeq d, \qquad \forall d \in \mathcal{D}.$$

Uno stimatore ottimo in MSE è pertanto uno stimatore per il quale, per ogni valore possibile di θ, la somma di varianza e distorsione (al quadrato) non supera quella di nessun altro stimatore. Ricordando che, per ogni stimatore d, $\mathrm{MSE}_\theta(d)$ è una funzione di θ, lo stimatore d^* è ottimo se la curva $\mathrm{MSE}_\theta(d^*)$ non supera mai (ovvero per ogni $\theta \in \Theta$) la curva MSE_θ di nessun altro stimatore $d \in D$. Si tratta di un requisito molto forte, che difficilmente uno stimatore riesce a soddisfare, se ci si riferisce all'intera classe degli stimatori di un parametro. Ciò che accade, in genere, è che le curve MSE_θ dei diversi stimatori si intersecano, ovvero che uno stimatore è migliore di un altro solo per alcuni valori di θ e non *uniformemente* in Θ (ovvero non $\forall \theta \in \Theta$).

4.30 Esempio (Modello bernoulliano). Sia $X_1, \ldots, X_n$ un campione casuale da una popolazione $\mathrm{Ber}(\theta)$. Si considerino i due stimatori:

$$d_1(\mathbf{X}_n) = \overline{X}_n \qquad \text{e} \qquad d_2(\mathbf{X}_n) = \frac{\sum_{i=1}^n X_i + \sqrt{n/4}}{n + \sqrt{n}}.$$

In questo caso si può verificare che nessuno dei due stimatori risulta migliore dell'altro. Si ha infatti che

$$\mathrm{MSE}_\theta(d_1) = \frac{\theta(1-\theta)}{n}, \qquad \mathrm{MSE}_\theta(d_2) = \frac{n}{4(n+\sqrt{n})^2}$$

e che $d_1 \succeq d_2$ se e solo se $\theta \notin [1/2 \pm 1/2\sqrt{1 - n^2/(n+\sqrt{n})^2}]$. □

In generale, il motivo per cui non si riesce a determinare uno stimatore ottimo in D risiede nell'ampiezza della classe D. La teoria frequentista affronta questo problema operando restrizioni dell'insieme D, e cercando stimatori ottimi in tali sottoclassi. La soluzione piú comunemente utilizzata è quella restringere la ricerca degli stimatori ottimi nella sottoclasse $D_\mathcal{U}$ di D costituita dagli stimatori non distorti di θ:

$$D_\mathcal{U} = \{d \in D : \mathbb{E}_\theta[d(\mathbf{X}_n)] = \theta, \forall \theta \in \Theta\}.$$

4.31 Definizione (Stimatore ottimo non distorto: UMVUE) Dato il modello statistico $\{\mathcal{X}^n, f_n(\mathbf{x}_n;\theta), \theta \in \Theta\}$, d^* è uno *stimatore ottimo non distorto* di θ se

$$d^* \in D_\mathcal{U} \quad \text{e} \quad R(\theta, d^*) \leq R(\theta, d), \quad \forall \theta \in \Theta, \quad \forall d \in D_\mathcal{U}.$$

Nel caso in cui $R(\theta, d) = \mathrm{MSE}_\theta(d)$, d^* si dice *stimatore non distorto di varianza minima* di θ (in inglese *uniformly minimum variance unbiased estimator*, o UMVUE) se

$$\mathbb{E}_\theta[d^*(\mathbf{X}_n)] = \theta, \quad \forall \theta \in \Theta \quad \text{e} \quad \mathbb{V}_\theta[d^*(\mathbf{X}_n)] \leq \mathbb{V}_\theta[d(\mathbf{X}_n)], \quad \forall \theta \in \Theta, \quad \forall d \in D_\mathcal{U}.$$

□

Esistono due strategie per l'individuazione di stimatori UMVUE.

a) Verifica di ottimalità basata sul limite inferiore di Cramer-Rao. Per un dato modello statistico, si definisce il limite inferiore per la varianza degli stimatori non distorti di θ (o di una funzione $g(\theta)$ di interesse). Se uno stimatore di d^* non distorto di θ ha varianza coincidente con tale quantità, allora d^* è UMVUE di θ.

b) Procedura costruttiva di Rao-Blackwell e Lehmann-Scheffé. Partendo da uno stimatore d non distorto e da una statistica T sufficiente per il modello, si costruisce uno stimatore d' non distorto, funzione di T e migliore di d. Sotto opportune condizioni si può mostrare che d' è UMVUE ed è unico.

Prima di trattare il primo dei metodi è necessario introdurre un concetto fondamentale nella teoria dell'inferenza statistica.

4.4.1 Informazione di Fisher

In questo paragrafo introduciamo una delle quantità più importanti della teoria dell'inferenza statistica: l'informazione attesa di Fisher. Il concetto è strettamente legato a quello di funzione di informazione di Fisher (o informazione di Fisher), informazione di Fisher osservata e a quello di funzione di punteggio, introdotti nel Capitolo 3. In quei casi si trattava di funzioni del campione osservato, in linea con l'ottica post-sperimentale su cui si basa il Principio di Verosimiglianza. L' informazione attesa di Fisher è invece definita in linea con l'ottica pre-sperimentale del Principio del Campionamento Ripetuto. Consideriamo per il momento il caso uniparametrico. L'estensione al caso multiparametrico è trattata nel Paragrafo 4.8.

4.32 Definizione (Informazione attesa di Fisher). Dato il modello statistico $\{\mathcal{X}^n, f_n(\mathbf{x}_n;\theta), \theta \in \Theta\}$, con $\Theta \subseteq \mathbb{R}$ e $f_n(\cdot;\theta)$ differenziabile rispetto a θ, e dato un campione[2] $X_1, \ldots, X_n$, si chiama *informazione attesa di Fisher* la seguente funzione di θ:

$$I_n(\theta) = \mathbb{E}_\theta \left[\left(\frac{\partial}{\partial \theta} \ln f_n(\mathbf{X}_n;\theta) \right)^2 \right]. \tag{4.11}$$

□

Nel caso di v.a. assolutamente continue si ha che:

$$I_n(\theta) = \int_{\mathcal{X}^n} \left(\frac{\partial}{\partial \theta} \ln f_n(\mathbf{x}_n;\theta) \right)^2 f_n(\mathbf{x}_n;\theta) d\mathbf{x}_n.$$

Nel caso di v.a. discrete l'integrale (che, ricordiamo, è multidimensionale) viene sostituito da una sommatoria.
Una semplificazione per l'espressione di $I_n(\theta)$ si può ottenere per la seguente importante classe di modelli.

4.33 Definizione (Problema regolare di stima - II). Si ha un *problema regolare di stima*[3] quando per il modello statistico $\{\mathcal{X}^n, f_n(\mathbf{x}_n;\theta), \theta \in \Theta\}$ valgono le seguenti due proprietà[4]:

1. il supporto delle v.a. X_i non dipende dal parametro incognito θ;

[2] Si noti che non si richiede che il campione sia i.i.d.
[3] Integriamo qui la definizione di problema regolare di stima data nel Capitolo 3.
[4] Notare la distinzione tra derivata parziale ($\partial/\partial\theta$) e derivata ($d/d\theta$).

2. è possibile derivare rispetto a θ due volte sotto il segno di integrale (è possibile quindi scambiare il segno di integrale con quello di derivata parziale prima e seconda rispetto a θ). In particolare, valgono le seguenti due identità:

$$\int_{\mathcal{X}^n} \frac{\partial}{\partial\theta} f_n(\mathbf{x}_n;\theta)d\mathbf{x}_n = \frac{d}{d\theta}\int_{\mathcal{X}^n} f_n(\mathbf{x}_n;\theta)d\mathbf{x}_n = 0 \tag{4.12}$$

$$\int_{\mathcal{X}^n} \frac{\partial^2}{\partial\theta^2} f_n(\mathbf{x}_n;\theta)d\mathbf{x}_n = \frac{d^2}{d\theta^2}\int_{\mathcal{X}^n} f_n(\mathbf{x}_n;\theta)d\mathbf{x}_n = 0 \tag{4.13}$$

□

4.34 Osservazione. Nel seguito assumeremo sempre garantite le condizioni di regolarità del modello. Tali proprietà valgono per modelli che sono famiglie esponenziali. □

La Definizione 4.32 vale per campioni generici. Nel caso di campioni casuali, il calcolo di $I_n(\theta)$ si semplifica consideravolmente.

4.35 Teorema. Si consideri il modello statistico $\{\mathcal{X}^n, f_n(\mathbf{x}_n;\theta), \theta \in \Theta\}$ con $\mathcal{X}^n$ indipendente da θ e un campione casuale $X_1, \ldots, X_n$. Si assuma inoltre che il problema di stima sia regolare. Valgono allora i seguenti risultati.

a) L'informazione di Fisher per il modello $\{\mathcal{X}^n, f_n(\mathbf{x}_n;\theta), \theta \in \Theta\}$ diventa

$$I_n(\theta) = n\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f_X(X;\theta)\right)^2\right] = nI_1(\theta).$$

dove $I_1(\theta)$ indica l'informazione di Fisher associate alla singola v.a. X.

b) Per $I_1(\theta)$ si ha che

$$I_1(\theta) = \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f_X(X;\theta)\right)^2\right] = -\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2}\ln f_X(X;\theta)\right]$$

e quindi

$$I_n(\theta) = -n\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2}\ln f_X(X;\theta)\right].$$

□

4.36 Osservazione

1. Nel caso dei problemi regolari di stima e di campioni casuali si ha quindi che

$$I_n(\theta) = \mathbb{E}_\theta[\mathcal{I}_n(\theta; \mathbf{X}_n)],$$

dove $\mathcal{I}_n(\theta; \mathbf{X}_n)$ rappresenta l'oggetto aleatorio la cui realizzazione in corrispondenza di un campione osservato, $\mathcal{I}_n(\theta; \mathbf{x}_n)$ è la funzione di informazione di Fisher, definita nella 2.17.

2. La determinazione di I_n attraverso I_1 consente una notevole semplificazione nei calcoli (che diventano unidimensionali).

□

Dimostrazione.
Parte a).

$$\begin{aligned} I_n(\theta) &= \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f_n(\mathbf{X}_n;\theta)\right)^2\right] = \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln\prod_{i=1}^n f_X(X_i;\theta)\right)^2\right] \\ &= \mathbb{E}_\theta\left[\left(\sum_{i=1}^n \frac{\partial}{\partial\theta}\ln f_X(X_i;\theta)\right)^2\right] = \sum_{i=1}^n \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f_X(X_i;\theta)\right)^2\right] + \\ &\quad + \sum_{i\neq j}\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f_X(X_i;\theta)\frac{\partial}{\partial\theta}\ln f_X(X_j;\theta)\right)\right] \\ &= nI_1(\theta) + 0. \end{aligned}$$

L'ultima uguaglianza si giustifica osservando che, per $i \neq j$, le v.a. X_i e X_j (e quindi anche loro funzioni) sono indipendenti e che, per la (4.12),

$$\begin{aligned} \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f_X(X_i;\theta)\right)\right] &= \int_{\mathcal{X}} \frac{\partial}{\partial\theta}\ln f_X(x;\theta)\times f_X(x;\theta)dx \\ &= \int_{\mathcal{X}} \frac{\frac{\partial}{\partial\theta}f_X(x;\theta)}{f_X(x;\theta)} f_X(x;\theta)dx \\ &= \frac{d}{d\theta}\int_{\mathcal{X}} f_X(x;\theta)dx = \frac{d}{d\theta}1 = 0. \end{aligned}$$

Pertanto

$$\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f_X(X_i;\theta)\frac{\partial}{\partial\theta}\ln f_X(X_j;\theta)\right)\right] = \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f_X(X_i;\theta)\right)\right]\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f_X(X_j;\theta)\right)\right] = 0.$$

Si osservi che per questa prima parte del risultato risulta essenziale l'ipotesi (4.12).
Parte b). Per verificare la seconda parte del teorema, usiamo la seguente notazione semplificata:

$$f = f_X(x;\theta),\quad f' = \frac{\partial}{\partial\theta}f_X(x;\theta),\quad f'' = \frac{\partial^2}{\partial\theta^2}f_X(x;\theta),\quad \ln f = \ln f_X(x;\theta),\quad \int = \int_{\mathcal{X}}.$$

Si ha quindi che

$$\begin{aligned}\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2}\ln f_X(X;\theta)\right] &= \int\left(\frac{\partial}{\partial\theta}\frac{f'}{f}\right)f dx = \int\frac{f''f-(f')^2}{f^2}f dx\\ &= \int f''dx - \int\left(\frac{f'}{f}\right)^2 f dx = 0 - \int\left(\frac{\partial}{\partial\theta}\ln f\right)^2 f dx\\ &= -\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f_X(X;\theta)\right)^2\right] = -I_1(\theta).\end{aligned}$$

dove la penultima uguaglianza si giustifica ricordando che, per la (4.13), $\int f''\ dx = 0$. Ricordando infine che, per campioni casuali, $I_n(\theta) = nI_1(\theta)$, si ottiene la parte b) del teorema.

Esempi rilevanti

Consideriamo il calcolo di $I_n(\theta)$ nel caso di campioni casuali per alcuni esempi rilevanti di modelli uniparametrici che sono famiglie esponenziali e per i quali sono pertanto valide le ipotesi del precedente teorema.

4.37 Esempio (Modello bernoulliano). Nel caso di un campione casuale di dimensione n da una Ber(θ) si ha:

$$\begin{aligned}\ln f_X(X;\theta) &= X\ln\theta + (1-X)\ln(1-\theta)\\ \frac{\partial}{\partial\theta}\ln f_X(X;\theta) &= \frac{X}{\theta} - \frac{1-X}{1-\theta}\\ \frac{\partial^2}{\partial\theta^2}\ln f_X(X;\theta) &= -\frac{X}{\theta^2} - \frac{1-X}{(1-\theta)^2}\\ \mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2}\ln f_X(X;\theta)\right] &= -I_1(\theta) = -\frac{\mathbb{E}_\theta[X]}{\theta^2} - \frac{1-\mathbb{E}_\theta[X]}{(1-\theta)^2} = -\frac{\theta}{\theta^2} - \frac{1-\theta}{(1-\theta)^2} = -\frac{1}{\theta(1-\theta)}\\ I_n(\theta) &= nI_1(\theta) = \frac{n}{\theta(1-\theta)}.\end{aligned}$$

□

4.38 Esempio (Modello di Poisson). Nel caso di un campione casuale di dimensione n da una Pois(θ) si ha:

$$\begin{aligned}\ln f_X(X;\theta) &= -\theta + X\ln\theta - \ln X!\\ \frac{\partial}{\partial\theta}\ln f_X(X;\theta) &= -1 + \frac{X}{\theta}\\ \frac{\partial^2}{\partial\theta^2}\ln f_X(X;\theta) &= -\frac{X}{\theta^2}\\ \mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2}\ln f_X(X;\theta)\right] &= -I_1(\theta) = -\frac{\mathbb{E}_\theta[X]}{\theta^2} = -\frac{\theta}{\theta^2} = -\frac{1}{\theta}\\ I_n(\theta) &= nI_1(\theta) = \frac{n}{\theta}.\end{aligned}$$

□

4.39 Esempio (Modello Esponenziale). Nel caso di un campione casuale di dimensione n da una $\text{Esp}(\theta)$ si ha:

$$\begin{aligned}
\ln f_X(x;\theta) &= -\ln\theta - \frac{X}{\theta} \\
\frac{\partial}{\partial\theta}\ln f_X(X;\theta) &= -\frac{1}{\theta} + \frac{X}{\theta^2} \\
\frac{\partial^2}{\partial\theta^2}\ln f_X(X;\theta) &= \frac{1}{\theta^2} - \frac{2X}{\theta^3} = \frac{\theta - 2X}{\theta^3} \\
\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2}\ln f_X(X;\theta)\right] &= -I_1(\theta) = -\frac{\theta - 2\mathbb{E}_\theta[X]}{\theta^3} = -\frac{1}{\theta^2} \\
I_n(\theta) &= nI_1(\theta) = \frac{n}{\theta^2}.
\end{aligned}$$

□

4.40 Esempio (Modello normale, varianza nota). Nel caso di un campione casuale di dimensione n da una $\text{N}(\theta, \sigma^2)$ (con σ^2 nota) si ha:

$$\begin{aligned}
\ln f_X(x;\theta) &= -\ln(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2}(X-\theta)^2 \\
\frac{\partial}{\partial\theta}\ln f_X(X;\theta) &= \frac{1}{\sigma^2}(X-\theta) \\
\frac{\partial^2}{\partial\theta^2}\ln f_X(X;\theta) &= -\frac{1}{\sigma^2} \\
\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2}\ln f_X(X;\theta)\right] &= -\frac{1}{\sigma^2} \\
I_1(\theta) &= \frac{1}{\sigma^2} \\
I_n(\theta) &= nI_1(\theta) = \frac{n}{\sigma^2}.
\end{aligned}$$

□

4.41 Esempio (Modello normale, valore atteso noto). Nel caso di un campione casuale di dimensione n da una $\text{N}(\mu_0, \theta)$ si ha:

$$\begin{aligned}
\ln f_X(x;\theta) &= -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\theta - \frac{1}{2\theta}(X-\mu_0)^2 \\
\frac{\partial}{\partial\theta}\ln f_X(x;\theta) &= -\frac{1}{2\theta} + \frac{1}{2\theta^2}(X-\mu_0)^2 \\
\frac{\partial^2}{\partial\theta^2}\ln f_X(x;\theta) &= \frac{1}{2\theta^2} - \frac{1}{\theta^3}(X-\mu_0)^2 \\
\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2}\ln f_X(X;\theta)\right] &= \frac{1}{2\theta^2} - \frac{\mathbb{E}_\theta(X-\mu_0)^2}{\theta^3} = \frac{1}{2\theta^2} - \frac{\theta}{\theta^3} = -\frac{1}{2\theta^2} \\
I_1(\theta) &= \frac{1}{2\theta^2} \\
I_n(\theta) &= nI_1(\theta) = \frac{n}{2\theta^2}.
\end{aligned}$$

□

Introduciamo una importante funzione collegata all'informazione di Fisher.

4.42 Definizione (Funzione di punteggio). Dato il modello statistico $\{\mathcal{X}^n, f_n(\mathbf{x}_n;\theta), \theta \in \Theta\}$ (con $\Theta \in \mathbb{R}$ e $f_n(\cdot;\theta)$ derivabile rispetto a θ) ed un campione osservato $\mathbf{x}_n$ (di dimensione n), si chiama *funzione di punteggio* (score function) la seguente quantità:

$$S_n(\mathbf{x}_n;\theta) = \frac{\partial}{\partial\theta} \ln f_n(\mathbf{x}_n;\theta). \tag{4.14}$$

Per $n = 1$ la funzione diventa

$$S_1(x;\theta) = \frac{\partial}{\partial\theta} \ln f_X(x;\theta).$$

□

Dalla definizione data si evince che, se consideriamo l'oggetto aleatorio $S_n(\mathbf{X}_n, \theta)$,

$$I_n(\theta) = \mathbb{E}_\theta \left[S_n(\mathbf{X}_n;\theta)^2 \right].$$

Nel caso di campioni casuali,

$$S_n(\mathbf{x}_n;\theta) = \frac{\partial}{\partial\theta} \ln \prod_{i=1}^n f_X(x_i;\theta) = \sum_{i=1}^n \frac{\partial}{\partial\theta} \ln f_X(x_i;\theta) = \sum_{i=1}^n S_1(x_i;\theta).$$

Per la funzione di punteggio vale il seguente importante risultato.

4.43 Teorema Dato il modello statistico $\{\mathcal{X}^n, f_n(\mathbf{x}_n;\theta), \theta \in \Theta\}$, nel caso di un problema regolare di stima (vedi Definizione 4.33) si ha che, per un campione casuale,

$$\mathbb{E}_\theta[S_n(\mathbf{X}_n;\theta)] = 0, \qquad \text{e} \qquad \mathbb{V}_\theta[S_n(\mathbf{X}_n;\theta)] = I_n(\theta).$$

□

Dimostrazione. Ricordando che $S_n(\mathbf{x}_n;\theta) = \sum_{i=1}^n S_1(x_i;\theta)$, per mostrare che $\mathbb{E}_\theta[S_n(\mathbf{X}_n;\theta)] = 0$ è sufficiente mostrare che $\mathbb{E}_\theta[S_1(X;\theta)] = 0$. Dalla definizione di funzione di punteggio si ha che

$$\begin{aligned}
\mathbb{E}_\theta[S_1(X;\theta)] &= \int_{\mathcal{X}} \frac{\partial}{\partial\theta} \ln f_X(x;\theta) f_X(x;\theta) dx \\
&= \int_{\mathcal{X}} \frac{\frac{\partial}{\partial\theta} f_X(x;\theta)}{f_X(x;\theta)} f_X(x;\theta) dx \\
&= \frac{d}{d\theta} \int_{\mathcal{X}} f_X(x;\theta) dx = 0.
\end{aligned}$$

Pertanto

$$\begin{aligned}
\mathbb{V}_\theta[S_n(\mathbf{X}_n;\theta)] &= n\mathbb{V}_\theta[S_1(X;\theta)] \\
&= n\mathbb{E}_\theta[S_1(X;\theta)^2] \\
&= -n\mathbb{E}_\theta\left[\left(\frac{\partial^2}{\partial\theta^2} \ln f_X(X;\theta)\right)\right] \\
&= I_n(\theta),
\end{aligned}$$

dove la penultima uguaglianza si ha in virtú del Teorema 4.35.

4.4.2 Limite inferiore di Cramer-Rao

In alcuni problemi inferenziali è possibile determinare un valore (o meglio, una funzione di θ), che indichiamo con $\text{cr}(\theta)$, al di sotto del quale non può andare la varianza di tutti gli stimatori non distorti di un parametro di interesse. In questi casi, dato uno stimatore non distorto del parametro, se la sua varianza uguaglia $\text{cr}(\theta)$, lo stimatore è un UMVUE per il parametro. La piú nota e utilizzata limitazione inferiore per la varianza di stimatori non distorti si ottiene dalla seguente disuguaglianza. Nel seguito si adotta la notazione del caso assolutamente continuo, senza perdere tuttavia in generalità.

4.44 Teorema (Disuguaglianza di Cramer-Rao). Sia $X_1, \ldots, X_n$ un campione con funzione di densità $f_n(\cdot;\theta)$ e $d(\mathbf{X}_n)$ uno stimatore non distorto di $\tau(\theta)$ (dove $\tau(\cdot)$ è una qualsiasi funzione di θ). Per un problema regolare di stima e, in particolare, se

$$\tau'(\theta) = \frac{d}{d\theta}\mathbb{E}_\theta[d(\mathbf{X}_n)] = \int_{\mathcal{X}^n} \frac{\partial}{\partial\theta}[d(\mathbf{x}_n) f_n(\mathbf{x}_n;\theta)]d\mathbf{x}_n$$

e se

$$\mathbb{V}_\theta[d(\mathbf{X}_n)] < \infty$$

allora, per ogni $\theta \in \Theta$,

$$\mathbb{V}_\theta[d(\mathbf{X}_n)] \geq \frac{\left(\frac{d}{d\theta}\mathbb{E}_\theta[d(\mathbf{X}_n)]\right)^2}{I_n(\theta)} = \frac{[\tau'(\theta)]^2}{I_n(\theta)}.$$

□

Dimostrazione. La dimostrazione consiste in una applicazione della disuguaglianza di Cauchy-Schwarz, in base alla quale, date due variabili aleatorie Z e Y,

$$[\text{Cov}_\theta(Z,Y)]^2 \leq \mathbb{V}_\theta(Z)\mathbb{V}_\theta(Y).$$

Da questa discende che

$$\mathbb{V}_\theta(Z) \geq \frac{[\text{Cov}_\theta(Z,Y)]^2}{\mathbb{V}_\theta(Y)}. \tag{4.15}$$

Poniamo ora:

$$Z = d(\mathbf{X}_n), \qquad \text{e} \qquad Y = S_n(\mathbf{X}_n,\theta) = \frac{\partial}{\partial\theta}\ln f_n(\mathbf{X}_n;\theta).$$

Si ha quindi che

$$\mathbb{E}_\theta[Z] = \mathbb{E}_\theta[d(\mathbf{X}_n)] = \tau(\theta), \qquad \text{e} \qquad \mathbb{E}_\theta[Y] = \mathbb{E}_\theta\left[\frac{\partial}{\partial\theta}\ln f_n(\mathbf{x}_n;\theta)\right].$$

Per definizione di covarianza si ha che

$$\text{Cov}_\theta(Z,Y) = \mathbb{E}_\theta[ZY] - \mathbb{E}_\theta[Z]\mathbb{E}_\theta[Y].$$

Poiché però, in questo caso, per il Teorema 4.43 (scambiabilità tra integrale e derivata parziale prima rispetto a θ),

$$\mathbb{E}_\theta[Y] = \mathbb{E}_\theta\left[\frac{\partial}{\partial\theta}\ln f_n(\mathbf{x}_n;\theta)\right] = 0,$$

si ha allora che

$$\begin{aligned}\mathrm{Cov}_\theta(Z,Y) &= \mathbb{E}_\theta[ZY] \\ &= \mathbb{E}_\theta\left[d(\mathbf{X}_n)\frac{\partial}{\partial\theta}\ln f_n(\mathbf{X}_n;\theta)\right] \\ &= \int_{\mathcal{X}^n} d(\mathbf{X}_n)\frac{\frac{\partial}{\partial\theta}f_n(\mathbf{x}_n;\theta)}{f_n(\mathbf{x}_n;\theta)}f_n(\mathbf{x}_n;\theta)d\mathbf{x}_n \\ &= \frac{d}{d\theta}\mathbb{E}_\theta[d(\mathbf{X}_n)] = \tau'(\theta),\end{aligned}$$

dove l'ultima uguaglianza si giustifica ricordando la possibilità di scambiare integrale e derivata. Dalla (4.15), ricordando che $\mathbb{V}_\theta[Z] = \mathbb{V}_\theta[d(\mathbf{X}_n)]$ e che $\mathbb{V}_\theta[Y] = \mathbb{V}_\theta[S_n(\mathbf{X}_n;\theta)] = I_n(\theta)$, discende quindi che

$$\mathbb{V}_\theta[d(\mathbf{X}_n)] \geq \frac{\left(\frac{d}{d\theta}\mathbb{E}_\theta[d(\mathbf{X}_n)]\right)^2}{I_n(\theta)}.$$

4.45 Osservazione *a*) La quantità

$$\mathrm{cr}[\tau(\theta)] = \frac{[\tau'(\theta)]^2}{I_n(\theta)}$$

viene di solito denominata *limite inferiore di Cramer-Rao* (`licr`) per gli stimatori non distorti di $\tau(\theta)$. Se, nelle ipotesi del teorema (problema regolare di stima), uno stimatore non distorto di $\tau(\theta)$ ha varianza coincidente con $\mathrm{cr}[\tau(\theta)]$, questo è certamente un UMVUE.

b) Nel caso di modelli uniparametrici qui trattati ($\theta \subseteq \mathbb{R}$), il `licr` è raggiunto dalla varianza di uno stimatore non distorto $d(\mathbf{X}_n)$ (ovviamente nel caso di problemi regolari di stima) se e solo se: (a) la distribuzione di probabilità della v.a. di base X costituisce una famiglia esponenziale; (b) lo stimatore in esame è in relazione lineare con la funzione di punteggio $S_n(\mathbf{X}_n;\theta)$[5].

c) Se $d(\mathbf{X}_n)$ è uno stimatore non distorto di θ, la disuguaglianza diventa

$$\mathbb{V}_\theta[d(\mathbf{X}_n)] \geq \frac{1}{I_n(\theta)} = \mathrm{cr}(\theta).$$

□

4.46 Definizione (Efficienza). Sia $d(\cdot)$ uno stimatore non distorto di θ. La quantità

$$\mathrm{eff}(d,\theta) = \frac{\mathrm{cr}(\theta)}{\mathbb{V}_\theta[d(\mathbf{X}_n)]}$$

si chiama *efficienza dello stimatore d*. □

Consideriamo ora alcuni esempi di stimatori efficienti.

4.47 Esempio (Modello bernoulliano.) Per un campione casuale da $\mathrm{Ber}(\theta)$ si ha che

$$\mathrm{cr}(\theta) = [I_n(\theta)]^{-1} = \frac{\theta(1-\theta)}{n}.$$

[5]Si vedano: Wijsman, R.A. (1973). On the attainment of the Cramer Rao lower bound. *The Annals of Statistics*, vol. 1, n. 3, pp. 538-542 e Schervish, M.J. (1995). *Theory of Statistics*, pp. 301-302, Springer.

Osservando che lo stimatore $d(\mathbf{X}_n) = \overline{X}_n$ è non distorto per θ e ha varianza coincidente con $\text{cr}(\theta)$, si ha che $\overline{X}_n$ è un UMVUE di θ nel modello bernoulliano. □

4.48 Esempio (Modello di Poisson). Per un campione casuale da Pois(θ) si ha che

$$\text{cr}(\theta) = [I_n(\theta)]^{-1} = \frac{\theta}{n}.$$

Lo stimatore $d(\mathbf{X}_n) = \overline{X}_n$ è non distorto per θ, ha varianza coincidente con $\text{cr}(\theta)$ ed è quindi un UMVUE di θ nel modello di Poisson. □

4.49 Esempio (Modello esponenziale). Per un campione casuale da Esp(θ) si ha che

$$\text{cr}(\theta) = [I_n(\theta)]^{-1} = \frac{\theta^2}{n}.$$

Lo stimatore $d(\mathbf{X}_n) = \overline{X}_n$ è non distorto per θ, ha varianza coincidente con $\text{cr}(\theta)$ ed è quindi un UMVUE di θ nel modello esponenziale. □

4.50 Esempio (Modello normale, valore atteso noto). Per un campione casuale da $\text{N}(\mu_0, \theta)$ si ha che

$$\text{cr}(\theta) = [I_n(\theta)]^{-1} = \frac{2\theta^2}{n}.$$

Lo stimatore $d(\mathbf{X}_n) = S_0^2 = \sum_{i=1}^n (X_i - \mu_0)^2/n$ è non distorto per θ, ha varianza coincidente con $\text{cr}(\theta)$ ed è quindi un UMVUE di θ.

□

Limiti applicativi della disuguaglianza di Cramer-Rao

Il `licr` consente di verificare se uno stimatore $d \in D_{\mathcal{U}}$ è UMVUE o di verificare lo scarto tra la varianza di uno stimatore dato e la varianza minima potenzialmente raggiungibile dagli stimatori nella casse $D_{\mathcal{U}}$. Tra l'altro (vedi Osservazione 4.45), abbiamo sottolineato che tale valore minimo è uguagliato dalla varianza di stimatori non distorti solo nel caso in cui si considerino famiglie esponenziali uniparametriche. La procedura di Cramer-Rao presenta quindi due limitazioni rilevanti.

a. Applicabilità limitata. Le ipotesi del Teorema 4.44 sono abbastanza restrittive. Ad esempio queste ipotesi non sono soddisfatte dai modelli con supporto dipendente da θ e dai modelli in cui non è valido lo scambio tra derivata ed integrale.

b. Impossibilità di stabilire se uno stimatore non distorto con varianza superiore a $\text{cr}(\theta)$ sia o meno un UMVUE.

Consideriamo due esempi comuni in cui si presentano i suddetti problemi.

4.51 Esempio (Modello uniforme)[6]. Nel modello uniforme in $[0, \theta]$ il supporto di X dipende da θ. Inoltre non è soddisfatta la prima ipotesi richiesta dal Teorema 4.44, ovvero non è possibile scambiare il segno di derivata ed integrale. Si ha infatti che[7], per una funzione

[6] Esempio tratto da Casella-Berger (op. cit., pp. 339-340).

[7] La seconda uguaglianza si giustifica applicando la Regola di Leibnitz (vedi Casella-Berger (2002), p. 69), in base alla quale, se $u(x, \theta)$, $a(\theta)$ e $b(\theta)$ sono funzioni differenziabili rispetto a θ, allora

$$\frac{d}{d\theta} \int_{a(\theta)}^{b(\theta)} u(x, \theta) dx = u[b(\theta), \theta] \frac{d}{d\theta} b(\theta) - u[a(\theta), \theta] \frac{d}{d\theta} a(\theta) + \int_{a(\theta)}^{b(\theta)} \frac{\partial}{\partial \theta} u(x, \theta) dx.$$

Nel nostro caso il risultato si ottiene ponendo $a(\theta) = 0$, $b(\theta) = \theta$ e $u(x, \theta) = h(x)/\theta$.

$h(x)$,

$$\begin{aligned}\frac{d}{d\theta}\int_{\mathcal{X}} h(x)f_X(x;\theta)dx &= \frac{d}{d\theta}\int_0^\theta h(x)\frac{1}{\theta}dx \\ &= \frac{h(\theta)}{\theta}+\int_0^\theta h(x)\frac{\partial}{\partial\theta}\left(\frac{1}{\theta}\right)dx \\ &\neq \int_0^\theta h(x)\frac{\partial}{\partial\theta}\left(\frac{1}{\theta}\right)dx\end{aligned}$$

a meno che $h(\theta)/\theta = 0$ per ogni $\theta > 0$. Se si calcola l'informazione attesa, assumendo erroneamente la derivabilità rispetto a θ di $\ln f_X(X;\theta) = -1/\theta$, si otterrebbe

$$n\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f_X(X;\theta)\right)^2\right] = \frac{n}{\theta^2}.$$

Tuttavia questo valore non costituisce il limite inferiore degli stimatori non distorti di θ. Se infatti consideriamo lo stimatore non distorto

$$d(\mathbf{X}_n) = \frac{n+1}{n}X_{(n)},$$

si ha che (vedi Esempio 4.25) per ogni $n \in \mathbb{N}$ e per ogni $\theta > 0$,

$$\mathbb{V}_\theta[d(\mathbf{X}_n)] = \frac{\theta^2}{n(n+2)} < \frac{\theta^2}{n}.$$

Quanto abbiamo trovato costituisce però solo una apparente violazione del `licr`, dal momento che il modello considerato come già osservato, non soddisfa le ipotesi richieste dal Teorema 4.44. Mostreremo nell'Esempio 4.66 che lo stimatore considerato è UMVUE. □

4.52 Esempio (Modello normale). Si consideri un campione casuale di dimensione n da $\mathrm{N}(\theta_1, \theta_2)$, in cui $\boldsymbol{\theta} = (\theta_1, \theta_2)$ ha entrambe le componenti incognite e consideriamo il problema di stima di θ_2. Il modello normale soddisfa le ipotesi del Teorema 4.44. In questo caso, tuttavia, trattandosi di un problema con due parametri incogniti, non è garantito (e infatti, come vedremo meglio in seguito, non si realizza) il raggiungimento del `licr` per stimatori non distorti di θ_2. Per calcolare $\mathrm{cr}(\theta_2)$ consideriamo i seguenti calcoli riferiti alla coppia θ_1, θ_2:

$$\begin{aligned}\ln f_X(X;\boldsymbol{\theta}) &= -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\theta_2 - \frac{1}{2\theta_2}(X-\theta_1)^2 \\ \frac{\partial}{\partial\theta_2}\ln f_X(X;\boldsymbol{\theta}) &= -\frac{1}{2\theta_2} + \frac{1}{2\theta_2^2}(X-\theta_1)^2 \\ \frac{\partial^2}{\partial\theta_2^2}\ln f_X(X;\boldsymbol{\theta}) &= \frac{1}{2\theta_2^2} - \frac{1}{\theta_2^3}(X-\theta_1)^2 \\ \mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta_2^2}\ln f_X(X;\boldsymbol{\theta})\right] &= \frac{1}{2\theta_2^2} - \frac{1}{\theta_2^3}\mathbb{E}_\theta[(X-\theta_1)^2] = \frac{1}{2\theta_2^2} - \frac{1}{\theta_2^3}\theta_2 = -\frac{1}{2\theta_2^2} = -I_1(\theta_2) \\ I_n(\theta_2) &= nI_1(\theta_2) = \frac{n}{2\theta_2^2}\end{aligned}$$

Il `licr` per la varianza degli stimatori non distorti di θ_2 è quindi

$$\mathrm{cr}(\theta_2) = \frac{2\theta_2^2}{n}.$$

Se consideriamo lo stimatore non distorto di θ_2, $S_n^2 = \sum_{i=1}^n (X_i - \overline{X}_n)/(n-1)$, abbiamo che

$$\mathbb{V}_\theta[S_n^2] = \frac{2\theta_2^2}{n-1} > \frac{2\theta_2^2}{n} = \text{cr}(\theta_2).$$

Non possiamo quindi concludere che S_n^2 sia UMVUE per θ_2, ma non possiamo neanche escluderlo (mostreremo che è UMVUE nell'Esempio 4.65). La procedura di Cramer-Rao non è quindi utile in questo caso ad individuare uno stimatore UMVUE per θ_2.

In questo stesso esempio, se il valore atteso di X è noto e pari a μ_0, possiamo considerare lo stimatore non distorto di θ_2, $S_0^2 = \sum_{i=1}^n (X_i - \mu_0)^2/n$, per il quale si ha che

$$\mathbb{V}_\theta[S_0^2] = \frac{2\theta_2^2}{n} = \text{cr}(\theta_2).$$

In questo caso, possiamo concludere che S_0^2 è UMVUE di θ_2. L'esistenza di uno stimatore UMVUE la cui varianza uguaglia il `licr` era del resto prevedibile dal momento che, quando il valore atteso è noto, il modello $\text{N}(\mu_0.\theta_2)$ è una famiglia esponenziale uniparametrica.

□

4.4.3 Procedura di Rao-Blackwell

La procedura di Rao-Blacwell consente di *migliorare* uno stimatore dato e, sotto opportune condizioni, di *costruire* uno stimatore ottimo per un parametro di interesse. La procedura poggia su due teoremi fondamentali della teoria dell'inferenza statistica: il Teorema di Rao-Blackwell (R-B) e il Teorema di Lehmann-Scheffé (L-S)[8].

Rispetto alla procedura basata sul `licr`, quella che viene introdotta in questo paragrafo presenta i seguenti vantaggi:

1. si tratta di una procedura *costruttiva* (e non di uno strumento di verifica) per il miglioramento di uno stimatore dato e, sotto opportune condizioni, consente la costruzione di uno stimatore ottimo;

2. le ipotesi per la validità dei teoremi richiesti sono molto meno restrittive di quelle del Teorema 4.44.

Il Teorema di Rao-Blackwell: miglioramento di stimatori

Diamo qui una versione piuttosto generale del Teorema di R-B. La restrizione al caso di funzione di perdita quadratica e di stimatori non distorti viene considerata nell'Osservazione 4.55.

Prima di esporre risultati formali, ricordiamo che, nell'ambito dell'inferenza statistica, le statistiche sufficienti hanno un ruolo fondamentale, dal momento che riassumono l'intera informazione contenuta nel campione. Risulta pertanto intuitivo considerare procedure inferenziali (stime puntuali, nel caso specifico) basate su statistiche sufficienti. Ci aspettiamo infatti che, in generale, gli stimatori basati su statistiche sufficienti siano *migliori* di quelli costruiti con statistiche non sufficienti. Questo ragionamento è formalizzato nei risultati che seguono.

[8] Si veda:
Rao C.R. (1945) Information and accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, 37 (3). pp. 81-91.
Blackwell D. (1947). Conditional expectation and unbiased sequential estimation. *Annals of Mathematical Statistics*, 18, pp. 105-110.
Lehmann E.L. e Scheffé H. (1950). Completeness, similar regions, and unbiased estimation. I., *Sankhya*, 10 (4), pp. 305-340.

4.53 Definizione (Classe degli stimatori funzioni di statistiche sufficienti). Dato un modello statistico $\{\mathcal{X}^n, f_n(\mathbf{x}_n;\theta), \theta \in \Theta\}$, indichiamo con D_S la classe degli stimatori di θ che sono funzioni di una statistica T sufficiente per il modello:

$$D_S = \{d \in D :\ d(\mathbf{x}_n) = h[T(\mathbf{x}_n)]\},$$

dove h è una qualunque funzione definita in $\mathcal{T}$. □

4.54 Teorema (Rao-Blackwell). Sia $\{\mathcal{X}^n, f_n(\mathbf{x}_n;\theta), \theta \in \Theta\}$ un modello statistico, T una statistica sufficiente e $d \in D$ un qualsiasi stimatore di θ. Se la funzione di perdita $\mathbb{L}(\theta, d)$ è convessa[9] in d per ogni $\theta \in \Theta$, la funzione

$$d'(\mathbf{x}_n) = \mathbb{E}_\theta[d(\mathbf{X}_n)|T = t] \tag{4.16}$$

(se esiste) è uno stimatore, funzione di T, tale che

$$R(\theta, d') \leq R(\theta, d), \qquad \forall \theta \in \Theta.$$

Inoltre

$$\mathbb{E}_\theta[d(\mathbf{X}_n)] = \mathbb{E}_\theta[d'(\mathbf{X}_n)]. \tag{4.17}$$

□

Dimostrazione Dobbiamo mostrare che:
a) d' è uno stimatore (ovvero non dipende da θ);
b) d' è funzione di $T(\mathbf{x}_n)$ (ovvero è funzione di statistica sufficiente);
c) $R(\theta, d') \leq R(\theta, d), \quad \forall \theta \in \Theta, \quad \forall d \neq d'$ (ovvero d' è migliore di d);
d) $\mathbb{E}_\theta[d(\mathbf{X}_n)] = \mathbb{E}_\theta[d'(\mathbf{X}_n)]$.

Per i punti a) e b) osserviamo che il valore atteso che definisce $d'(\mathbf{x}_n)$ va inteso rispetto alla distribuzione di $\mathbf{X}_n$ condizionata a T, ovvero $f_{\mathbf{X}_n|T}(\mathbf{x}_n|T = t)$. Tale distribuzione, per la sufficienza di $T(\cdot)$ non dipende da θ (vedi proprietà statistiche sufficienti) e il risultante valore atteso è funzione dei dati campionari (attraverso $T(\mathbf{x}_n) = t$) e non del parametro.

Consideriamo il punto c). Indichiamo, per chiarezza, con $\mathbb{E}_\theta^{\mathbf{X}_n}$, $\mathbb{E}_\theta^{T}$, $\mathbb{E}_\theta^{\mathbf{X}_n|T}$ le funzioni valore atteso con riferimento alle distribuzioni $f_{\mathbf{X}_n}(\cdot;\theta)$, $f_T(\cdot;\theta)$ e $f_{\mathbf{X}_n|T}(\cdot|T = t)$. Per la proprietà del valore atteso iterato[10]

$$R(\theta, d) = \mathbb{E}_\theta^{\mathbf{X}_n}[\mathbb{L}(\theta, d(\mathbf{X}_n))] = \mathbb{E}_\theta^{T}\left[\mathbb{E}_\theta^{\mathbf{X}_n|T}\left[\mathbb{L}(\theta, d(\mathbf{X}_n))\right]\right]. \tag{4.18}$$

Per la Disuguaglianza di Jensen[11] e la convessità della funzione $\mathbb{L}$ in d, si ha che

$$\mathbb{E}_\theta^{\mathbf{X}_n|T}\left[\mathbb{L}(\theta, d(\mathbf{X}_n))\right] \geq \mathbb{L}\left(\theta, \mathbb{E}_\theta^{\mathbf{X}_n|T}[d(\mathbf{X}_n)]\right) = \mathbb{L}\left(\theta, d'(\mathbf{X}_n)\right).$$

[9] **Funzione convessa**. Una funzione reale di variabile reale f è convessa se, comunque presi due punti x_1, x_2 (con $x_1 < x_2$) in $\mathbb{R}$ e una costante $\lambda \in (0,1)$, si ha:

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2).$$

[10] **Valore atteso iterato.** Data la v.a. doppia (X, Y) si ha che $\mathbb{E}_\theta[X] = \mathbb{E}_\theta\left[\mathbb{E}_\theta(X|Y)\right]$ dove il valore atteso interno è rispetto alla distribuzione di X condizionata a Y e quello esterno rispetto alla distribuzione di Y.

[11] **Disuguaglianza di Jensen**. Sia X una v.a. tale che $\mathbb{P}(X \in C) = 1$ con C insieme convesso di $\mathbb{R}$ e sia $f : C \to \mathbb{R}$ una funzione convessa. Se esistono i valori attesi $\mathbb{E}_\theta(X)$ e $\mathbb{E}_\theta[f(X)]$, allora

$$\mathbb{E}[f(X)] \geq f(\mathbb{E}[X]).$$

Sostituendo in (4.18) si ottiene quindi

$$R(\theta, d(\mathbf{X}_n)) \geq \mathbb{E}_\theta^T \left[\mathbb{L}\left(\theta, d'(\mathbf{X}_n)\right)\right] = R(\theta, d'(\mathbf{X}_n)), \qquad \forall \theta \in \Theta.$$

Consideriamo infine il punto d). Per mostrare che d e d' hanno stesso valore atteso, si osservi che

$$\mathbb{E}_\theta[d(\mathbf{X}_n)] = \mathbb{E}_\theta^{\mathbf{X}_n}[d(\mathbf{X}_n)] = \mathbb{E}_\theta^T\left[\mathbb{E}_\theta^{\mathbf{X}_n|T}[d(\mathbf{X}_n)]\right] = \mathbb{E}_\theta^T\left[d'(\mathbf{X}_n)\right].$$

Corollario. Nelle ipotesi del Teorema 4.54, se $\mathbb{L}(\theta, d) = [\theta - d(\mathbf{x}_n)]^2$ (perdita quadratica) e se $d \in D_{\mathcal{U}}$ (stimatore non distorto di θ), allora $d' \in D_S \cap D_{\mathcal{U}}$ e

$$\mathbb{V}_\theta(d') \leq \mathbb{V}_\theta(d), \qquad \forall \theta \in \Theta.$$

Dimostrazione. Utilizzando la perdita quadratica si ha che $R(\theta; \cdot) = \mathrm{MSE}_\theta(\cdot)$. Inoltre, per la (4.17), $d \in D_{\mathcal{U}} \Rightarrow d' \in D_{\mathcal{U}}$ e , in questo caso, $R(\theta, d) = \mathrm{MSE}_\theta(d) = \mathbb{V}_\theta(d)$ e $R(\theta, d') = \mathrm{MSE}_\theta(d') = \mathbb{V}_\theta(d')$.

4.55 Osservazione

1. Il Teorema 4.54 si applica anche nel caso in cui $\dim(\theta) = k > 1$.

2. Il teorema è valido anche per modelli con supporto dipendente da θ e anche per problemi di stima non regolari. La procedura ha quindi una applicabilita piú vasta del Teorema di Cramer-Rao.

3. Il teorema afferma sostanzialmente che gli unici stimatori da prendere in esame sono quelli in D_S. Infatti, comunque scelto lo stimatore iniziale d, lo stimatore risultante d', migliore di quello di partenza, è funzione della statistica sufficiente T con cui si definisce d'. Questo vuol dire che, per determinare uno stimatore ottimo di θ, possiamo concentrare la ricerca nella classe D_S degli stimatori che sono funzioni di statistiche sufficienti per il modello.

4. Il teorema di R-B non garantisce che d' sia esso stesso una statistica sufficiente (non è richiesta l'invertibilità della funzione h nella definizione di D_S).

5. Il teorema non garantisce che d' sia lo stimatore ottimo (o, nel caso $d \in D_{\mathcal{U}}$, che sia UMVUE). A questa esigenza risponde il Teorema di L-S.

□

Il calcolo della (4.16) (la cosidetta *rao-blackwellizzazione* di d) previsto dal Teo. 4.54 non è in genere semplice. Consideriamo un caso in cui questa operazione risulta invece molto agevole

4.56 Esempio (Modello bernoulliano). Consideriamo un campione casuale da $\mathrm{Ber}(\theta)$. In questo modello sappiamo che $T(\mathbf{X}_n) = \sum_{i=1}^n X_i$ è una statistica sufficiente minimale. Come stimatore di partenza consideriamo $d(\mathbf{X}_n) = X_1$, ovvero la prima osservazione campionaria. Si tratta, evidentemente, di uno stimatore pessimo, in quanto trascura $n-1$ osservazioni disponibili e stima $\theta \in [0,1]$ con il valore $x_1 = 0$ oppure con il valore $x_1 = 1$. Lo stimatore d è tuttavia non distorto, poiché $\mathbb{E}_\theta[X_1] = \theta, \forall \theta \in [0,1]$, e può essere migliorato

con la procedura di R-B.

$$\begin{aligned} d'(\mathbf{x}_n) &= \mathbb{E}_\theta[X_1|\sum_{i=1}^n X_i = t] = \mathbb{P}[X_1 = 1|\sum_{i=1}^n X_i = t] \\ &= \frac{\mathbb{P}[X_1 = 1, \sum_{i=1}^n X_i = t]}{\mathbb{P}[\sum_{i=1}^n X_i = t]} \\ &= \frac{\mathbb{P}[\sum_{i=1}^n X_i = t|X_1 = 1]\mathbb{P}[X_1 = 1]}{\mathbb{P}[\sum_{i=1}^n X_i = t]} \\ &= \frac{\mathbb{P}[\sum_{i=2}^n X_i = t-1]\mathbb{P}[X_1 = 1]}{\mathbb{P}[\sum_{i=1}^n X_i = t]} \\ &= \frac{\binom{n-1}{t-1}\theta^{t-1}(1-\theta)^{n-t} \times \theta}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} \\ &= \frac{t}{n}, \end{aligned}$$

dove la penultima uguaglianza si ottiene ricordando che la somma di r v.a. bernoulliane indipendenti di parametro θ ha distribuzione Binom(r, θ). Abbiamo quindi ottenuto

$$d'(\mathbf{x}_n) = \frac{t}{n} = \frac{\sum_{i=1}^n x_i}{n} = \overline{x}_n.$$

Ciò vuol dire che, partendo dallo stimatore $d(\mathbf{X}_n) = X_1$, con una sola applicazione della procedura di R-B non solo abbiamo ottenuto uno stimatore migliore di d, ma addirittura lo stimatore UMVUE di θ. □

4.57 Esempio (Modello normale).[12] Per semplicità consideriamo un campione casuale di dimensione $n = 2$ da N$(\theta, 1)$. Sia $d(\mathbf{X}_n) = X_1$ lo stimatore non distorto di partenza e $T(\mathbf{X}_n) = X_1 + X_2$ la statistica sufficiente (minimale) con cui si costruisce lo stimatore migliorato $d'(\mathbf{X}_n)$. Per definizione $d'(\mathbf{X}_n)$ coincide con il valore atteso della v.a. $X_1|X_1+X_2$. Per determinare la funzione densità condizionata per questa v.a., osserviamo innanzitutto che l'evento $(X_1 = x_1, X_1 + X_2 = t)$ coincide con l'evento $(X_1 = x_1, X_2 = t - x_1)$ e quindi, per le funzioni di densità si ha

$$f_{X_1, X_1+X_2}(x_1, t; \theta) = f_{X_1,X_2}(x_1, t - x_1; \theta) = f_{X_1}(x_1; \theta) f_{X_2}(t - x_1; \theta),$$

dove l'ultima uguaglianza si ha per l'indipendenza di X_1 e X_2.

La densità condizionata che cerchiamo, sarà quindi:

$$\begin{aligned} f_{X_1|X_1+X_2}(x_1|t; \theta) &= \frac{f_{X_1,X_1+X_2}(x_1, t; \theta)}{f_{X_1+X_2}(t; \theta)} \\ &= \frac{f_{X_1,X_2}(x_1, t - x_1; \theta)}{f_{X_1+X_2}(t; \theta)} \\ &= \frac{f_{X_1}(x_1; \theta) f_{X_2}(t - x_1; \theta)}{f_{X_1+X_2}(t; \theta)} \\ &= \frac{\phi(x_1; \theta, 1) \cdot \phi(t - x_1; \theta, 1)}{\phi(t; 2\theta, 2)}, \end{aligned}$$

[12] Esempio tratto da L. Piccinato (2009). *Metodi per le decisioni statistiche*, Springer.

dove $\phi(\cdot;\mu,\sigma^2)$ indica la funzione di densità di una v.a. $\mathrm{N}(\mu;\sigma^2)$ e dove l'ultima uguaglianza si giustifica ricordando che, per l'indipendenza di X_1 e X_2, si ha $X_1+X_2 \sim \mathrm{N}(2\theta,2)$. Poiché

$$\begin{aligned}\phi(x_1;\theta,1) &= \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}(x_1-\theta)^2\right\}\\ \phi(t-x_1;\theta,1) &= \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}(t-x_1-\theta)^2\right\}\\ \phi(t;2\theta,2) &= \frac{1}{2\sqrt{\pi}}\exp\left\{-\frac{1}{4}(t-2\theta)^2\right\},\end{aligned}$$

con alcuni calcoli si trova che

$$f_{X_1|X_1+X_2}(x_1|t;\theta) = \phi\left(x_1;\frac{t}{2},\frac{1}{2}\right)$$

che è la funzione di densità di una v.a. $\mathrm{N}\left(\frac{t}{2},\frac{1}{2}\right)$. Si ha quindi che

$$d'(\mathbf{x}_n) = \mathbb{E}_\theta[X_1|T=t] = \frac{t}{2} = \frac{x_1+x_2}{2},$$

che coincide, anche questa volta, con lo stimatore UMVUE di θ. □

Il Teorema di Lehmann-Scheffé: unicità dello stimatore UMVUE

Abbiamo già osservato che è difficile in genere che esista uno stimatore ottimo e che quindi, nella teoria frequentista, si cerca lo stimatore migliore nella classe degli stimatori non distorti, ovvero lo stimatore UMVUE. Il Teorema di Lehmann-Scheffé fornisce le condizioni per l'unicità dello stimatore UMVUE. Per enunciare e dimostrare il teorema è necessario il concetto di *completezza* di una statistica.

4.58 Definizione (Completezza). Dato il modello statistico $\{\mathcal{X}^n, f_n(\mathbf{x}_n;\theta), \theta\in\Theta\}$, una statistica T si dice *completa* se

$$\mathbb{E}_\theta[q(T)] = 0 \quad \forall\theta\in\Theta \quad \Rightarrow \quad \mathbb{P}_\theta[q(T)=0] = 1 \quad \forall\theta\in\Theta,$$

(ovvero, $q(T)=0$ quasi certamente rispetto a tutte le leggi di probabilità del modello). □

4.59 Osservazione

1. La completezza è una proprietà riferita all'intera famiglia di distribuzioni di probabilità di T e non a una particolare distribuzione: la condizione che definisce la completezza deve infatti valere $\forall\theta\in\Theta$.

2. Nel caso di modelli che sono famiglie esponenziali con k parametri, le statistiche sufficienti (minimali)
$$\left(\sum_{i=1}^n T_1(x_i),\ldots,\sum_{i=1}^n T_k(x_i)\right)$$
che si ottengono con il criterio di fattorizzazione sono (in condizioni molto generali[13]) anche complete.

[13] Si richiede che lo spazio Θ contenga un insieme aperto di $\mathbb{R}^k$.

□

4.60 Esempio (Modello uniforme). Per mostrare la completezza della statistica $X_{(n)}$ in questo modello è necessario verificare le condizioni della Definizione 4.58. Sia g una funzione tale che $\mathbb{E}_\theta[g(T)] = 0$ per ogni θ. Mostriamo che $g(\cdot) = 0$ quasi certamente. Ricordando che $f_T(t;\theta) = nt^{n-1}\theta^{-n} I_{(0,\theta)}(t)$ e osservando che

$$\frac{d}{d\theta}\mathbb{E}_\theta[g(T)] = \frac{d}{d\theta}0 = 0,$$

si ha che

$$\begin{aligned} 0 = \frac{d}{d\theta}\mathbb{E}_\theta[g(T)] &= \frac{d}{d\theta}\int_0^\theta ng(t)t^{n-1}\theta^{-n}dt \\ &= \frac{d}{d\theta}\left\{\left(\theta^{-n}\right) \times \left(\int_0^\theta ng(t)t^{n-1}dt\right)\right\} \\ &= (\theta^{-n})\left[\frac{d}{d\theta}\int_0^\theta ng(t)t^{n-1}dt\right] + \left(\frac{d}{d\theta}\theta^{-n}\right)\int_0^\theta ng(t)t^{n-1}dt \\ &= \theta^{-n}ng(\theta)\theta^{n-1} + 0 \\ &= \theta^{-1}ng(\theta), \end{aligned}$$

dove la terzultima uguaglianza si ottiene applicando la regola di derivazione del prodotto di funzioni e la penultima uguaglianza discende dal fatto che, per il Teorema fondamentale del calcolo integrale[14] si ha che

$$\frac{d}{d\theta}\int_0^\theta ng(t)t^{n-1}dt = ng(\theta)\theta^{n-1}$$

e dove il secondo addendo è uguale a zero poiché $\int_0^\theta ng(t)t^{n-1}dt = \theta^n\mathbb{E}_\theta[g(T)] = \theta^n \times 0$ per ipotesi. Pertanto,

$$\theta^{-1}ng(\theta) = 0 \qquad \forall\theta > 0$$

e quindi, essendo $\theta^{-1}n \neq 0$, necessariamente $g(\cdot)$ è identicamente nulla. La completezza di $X_{(n)}$ è cosí dimostrata[15].

□

4.61 Teorema (Lehmann-Scheffé). Sia $\{\mathcal{X}^n, f_n(\mathbf{x}_n;\theta), \theta \in \Theta\}$ un modello statistico, T una statistica sufficiente e completa, D_S la classe degli stimatori basati su T e $d' \in D_S \cap D_{\mathcal{U}}$ uno stimatore non distorto di θ. Se la funzione di perdita $L(\theta, d)$ è convessa in d per ogni $\theta \in \Theta$, ogni altro stimatore in $D_S \cap D_{\mathcal{U}}$ coincide quasi certamente con d', ovvero:

$$\mathbb{P}_\theta[d(\mathbf{X}_n) = d'(\mathbf{X}_n)] = 1, \qquad \forall d, d' \in D_S \cap D_{\mathcal{U}}, \quad \forall\theta \in \Theta.$$

Dimostrazione. Sia d un qualsiasi stimatore in $D_S \cap D_{\mathcal{U}}$ tale che $d \neq d'$. Il fatto che $d, d' \in D_S$ implica che esistono due funzioni f e f' tali che $d(\mathbf{X}_n) = f(T)$ e $d'(\mathbf{X}_n) = f'(T)$, dove $T = T(\mathbf{X}_n)$. Il fatto che $d, d' \in D_{\mathcal{U}}$ implica invece che $\mathbb{E}_\theta[d] = \mathbb{E}_\theta[d'] = \theta,\ \forall\theta$. Sia

$$q(T) = f(T) - f'(T).$$

[14] Il teorema afferma che, per ogni funzione integrabile h, si ha che $\frac{d}{d\theta}\int_0^\theta h(t)dt = h(\theta)$.

[15] Nel libro Casella-Berger (2009), da cui l'esempio è tratto (pp. 286-287), si osserva che, a rigore, la proprietà mostrata è valida solo per le funzioni $q(\cdot)$ per cui è valido il Teorema fondamentale del calcolo integrale, ovvero per le funzioni integrabili nel senso di Riemann. Questa classe è tuttavia sufficientemente ampia.

Si ha che

$$\mathbb{E}_\theta[q(T)] = \mathbb{E}_\theta[f(T) - f'(T)] = \mathbb{E}_\theta[d] - \mathbb{E}_\theta[d'] = 0, \qquad \forall \theta \in \Theta.$$

Per la completezza di T si ha quindi che

$$\mathbb{P}_\theta[q(T) = 0] = \mathbb{P}_\theta[d(\mathbf{X}_n) = d'(\mathbf{X}_n)] = 1, \qquad \forall \theta \in \Theta.$$

Il teorema è quindi dimostrato. □

4.62 Osservazione

1. La principale conseguenza del teorema dimostrato è che se $d' \in D_S \cap D_{\mathcal{U}}$ e se T (sufficiente) è anche completa, allora l'UMVUE è unico (a meno di insiemi di misura nulla, ovvero irrilevanti).
2. Il Teorema di L-S assicura che il procedimento del Teorema di R-B si arresta dopo un solo passo, se la statistica sufficiente T su cui si basa il miglioramento dello stimatore di partenza è una statistica completa.
3. Se, indipendentemente dalla procedura di R-B, abbiamo a disposizione uno stimatore non distorto di θ, che è anche funzione di una statistica sufficiente e completa, questo stimatore è necessariamente l'UMVUE.
4. Ogni statistica, il cui valore atteso è esprimibile come funzione $g(\theta)$ del parametro del modello, che è funzione di una statistica sufficiente e completa, è UMVUE del proprio valore atteso $g(\theta)$.
5. Il fatto che le statistiche sufficienti (minimali) per i modelli che sono famiglie esponenziali sono anche complete, rende molto agevole l'individuazione dello stimatore UMVUE per il parametro di molti modelli comuni.

□

4.63 Esempio (Media campionaria). I modelli Ber(θ), Pois(θ), Esp(θ), N(θ, σ^2) (var. nota) sono tutte famiglie esponenziali. Per tutti questo modelli, lo stimatore $\overline{X}_n$ è non distorto, ed è pure statistica sufficiente minimale e completa. La media campionaria è quindi, UMVUE per il parametro θ di tutti questi modelli. Si tratta di una conferma (che non necessita di alcun calcolo) di quanto verificato con la disuguaglianza di Cramer-Rao. □

4.64 Esempio (UMVUE per la varianza, modello normale con valore atteso noto). Per il modello N(μ_0, θ) (μ_0 noto) lo stimatore $S_0^2 = \sum_{i=1}^n (X_i - \mu_0)^2/n$ è non distorto ed è inoltre una statistica sufficiente e completa (fam. esponenziale). Si tratta quindi dello stimatore UMVUE. Anche in questo caso si tratta di una conferma di quanto già mostrato con la procedura di Cramer-Rao. □

La procedura di R-B e L-S ci consente di risolvere problemi che non possono essere affrontati con la Disuguaglianza di Cramer-Rao.

4.65 Esempio (UMVUE per varianza, modello normale). Nell'Es. 4.52 abbiamo mostrato che lo stimatore non distorto $S_n^2 = \sum_{i=1}^n (X_i - \overline{X}_n)/(n-1)$ della varianza (θ_2) del modello N(θ_1, θ_2) (valore atteso incognito) ha varianza superiore a cr(θ_2) e che quindi non siamo in grado di dire se si tratta o meno di stimatore UMVUE. Infatti, il raggiungimento del `licr` non è assicurato per modelli con piú di un parametro incognito. Tuttavia, essendo S_n^2 funzione di $(\sum_{i=1}^n X, \sum_{i=1}^n X_i^2)$, statistica sufficiente e completa per il modello (si tratta

di famiglia esponenziale), possiamo ora affermare che, in virtú dei Teoremi 4.54 e 4.61 è l'unico UMVUE di θ_2. □

4.66 Esempio (Modello uniforme). Possiamo ora anche risolvere il problema della ricerca dello stimatore UMVUE per il modello uniforme in $[0, \theta]$, $\theta > 0$. Consideriamo lo stimatore $d(\mathbf{X}_n) = (n+1)X_{(n)}/n$. Abbiamo mostrato in precedenza che d è non distorto per θ. Inoltre, è funzione della statistica sufficiente (e minimale) $T(\mathbf{X}_n) = X_{(n)}$. Poiché il modello uniforme non è una famiglia esponenziale, non è scontato che $T(\mathbf{X}_n) = X_{(n)}$ sia anche completa. La completezza di questa statistica è stata però mostrata nell'Es. 4.60. Per i Teo. 4.54 e 4.61 possiamo quindi concludere che $d(\mathbf{X}_n) = (n+1)X_{(n)}/n$ è l'unico UMVUE di θ nel modello in esame. □

Possibili problemi dell'UMVUE: inammissibilità

Abbiamo piú volte sottolineato che l'esigenza di restringere la ricerca di stimatori ottimi nella classe $D_{\mathcal{U}}$ degli stimatori non distorti di un parametro è dettata dalla frequente inesistenza di uno stimatore migliore di ogni altro nella classe D di tutti gli stimatori. Tuttavia, è possibile che tra gli stimatori distorti vi siano stimatori migliori di un UMVUE. Con linguaggio della teoria delle decisioni, è cioè possibile che lo stimatore UMVUE sia *inammissibile*.

4.67 Definizione (Stimatore ammissibile). Dato il modello statistico $\{\mathcal{X}^n, f_n(\mathbf{x}_n; \theta), \theta \in \Theta\}$, uno stimatore d' di θ si dice *ammissibile* se non esistono stimatori d tali che

$$R(\theta, d) \leq R(\theta, d'), \qquad \forall \theta \in \Theta$$

(con disuguaglianza stretta per almeno un valore di $\theta \in \Theta$). □

Mostriamo due esempi di stimatori UMVUE inammissibili.

4.68 Esempio (Inammissibilità UMVUE nel modello normale, I) [16]**.** Si consideri un campione casuale di dimensione n dal modello $\mathrm{N}(0, \theta)$. Una statistica sufficiente e completa è $T(\mathbf{X}_n) = \sum_{i=1}^n X_i^2$. Lo stimatore

$$d'(\mathbf{X}_n) = \frac{\sum_{i=1}^n X_i^2}{n}$$

è lo stimatore UMVUE di θ. Infatti, per il teorema 3.25

$$\frac{T}{\theta} = \frac{nS_0^2}{\theta} \sim \chi_n^2 \quad \Rightarrow \quad \mathbb{E}_\theta\left[\frac{T}{n}\right] = \theta,$$

ovvero $d' \in D_{\mathcal{U}}$ e

$$\mathrm{MSE}_\theta[d'] = \mathbb{V}_\theta[d'] = \frac{2}{n}\theta^2.$$

Consideriamo ora lo stimatore

$$d(\mathbf{X}_n) = \frac{\sum_{i=1}^n X_i^2}{n+2},$$

per il quale si ha

$$\mathbb{V}_\theta[d] = \frac{2n}{(n+2)^2}\theta^2, \qquad \mathbb{E}_\theta[d] = \frac{n}{n+2}\theta.$$

Si ha quindi che, per ogni $\theta > 0$,

$$\mathrm{MSE}_\theta[d] = \mathbb{V}_\theta[d] + B_\theta^2[d] = \frac{2}{n+2}\theta^2 < \frac{2}{n}\theta^2 = \mathrm{MSE}_\theta[d'].$$

[16] Esempio tratto da Piccinato (2009), p. 291.

Lo stimatore distorto d è quindi migliore di quello ottimo non distorto d' per ogni θ e per ogni n. □

4.69 Esempio (Inammissibilità di UMVUE nel modello normale, II). Nel caso di un campione casuale di dimensione n da $\mathrm{N}(\theta_1, \theta_2)$, consideriamo per θ_2 gli stimatori S_n^2 (UMVUE) e $\widehat{\sigma}_n^2$ (stimatore distorto). Abbiamo mostrato in precedenza che

$$\mathbb{E}_\theta[S_n^2] = \theta_2, \quad \mathbb{V}_\theta[S_n^2] = \frac{2}{n-1}\theta_2^2, \quad \mathbb{E}_\theta[\widehat{\sigma}_n^2] = \frac{n-1}{n}\theta_2, \quad \mathbb{V}_\theta[\widehat{\sigma}_n^2] = \frac{2(n-1)}{n^2}\theta_2^2.$$

Con semplici calcoli si mostra quindi che per ogni $\theta_2 > 0$ e per ogni $n > 1$ si ha

$$\mathrm{MSE}_\theta[\widehat{\sigma}_n^2] = \frac{2n-1}{n^2}\theta_2^2 < \frac{2}{n-1}\theta_2^2 = \mathrm{MSE}_\theta[S_n^2].$$

La precedente relazione dimostra l'inammissibilità di S_n^2, UMVUE di θ_2. □

4.5 Proprietà asintotiche degli stimatori

Tutte le proprietà degli stimatori fin qui considerate (ottimalità, efficienza, non distorsione) si riferiscono a dimensioni campionarie arbitrarie ma finite. Si parla di *proprietà asintotiche* di una procedura inferenziale (ad esempio di uno stimatore puntuale, di uno stimatore per insiemi o di un test) quando si guarda al comportamento della procedura al crescere della dimensione campionaria. Nel caso degli stimatori, le proprietà asintotiche di uno stimatore $d \in D$ si riferiscono pertanto alla successione di v.a. definita da

$$(d_n,\ n \in \mathbb{N}),$$

dove, per ogni $n \in \mathbb{N}$, $d_n = d(\mathbf{X}_n)$. Le principali proprietà asintotiche di uno stimatore sono le seguenti

a) consistenza;
b) normalità asintotica;
c) efficienza (ottimalità) asintotica.

4.5.1 Consistenza

Si tratta di una proprietà praticamente imprescindibile per uno stimatore, o meglio, per la successione $(d_n,\ n \in \mathbb{N})$ degli stimatori di un parametro. Informalmente possiamo dire che uno stimatore è consistente se, al crescere di n, $d(\mathbf{X}_n)$ tende ad avvicinarsi sempre piú (fino a coincidere) al valore vero del parametro, qualunque esso sia. In pratica, la consistenza assicura che, al crescere della dimensione campionaria, l'accuratezza dello stimatore tende a migliorare.

4.70 Definizione (Consistenza). Dato il modello statistico $\{\mathcal{X}^n, f_n(\mathbf{x}_n; \theta), \theta \in \Theta\}$, la successione $(d_n,\ n \in \mathbb{N})$ di stimatori del parametro θ è *consistente* (o *consistente in senso debole*) se questa, $\forall \theta \in \Theta$, converge in probabilità a θ, ovvero se, per ogni $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}_\theta\left(|d_n - \theta| < \epsilon\right) = 1, \qquad \forall \theta \in \Theta.$$

□

Si osservi che la definizione di consistenza richiede la convergenza in probabilità della successione degli stimatori *per ogni* possibile valore di θ previsto dal modello statistico in esame.

4.71 Esempio (Consistenza della media campionaria, modello normale). Consideriamo un campione casuale di dimensione n da $\mathrm{N}(\theta, 1)$ e la successione delle medie campionarie:

$$\left(\overline{X}_n,\ n \in \mathbb{N}\right).$$

Ricordando che $\overline{X}_n \sim \mathrm{N}(\theta, 1/n)$ e indicando con Z la v.a. normale standardizzata, si ha che, per ogni $\epsilon > 0$,

$$\begin{aligned} \mathbb{P}_\theta(|\overline{X}_n - \theta| < \epsilon) &= \mathbb{P}(-\epsilon\sqrt{n} < Z < \epsilon\sqrt{n}) \\ &= \Phi(\epsilon\sqrt{n}) - \Phi(-\epsilon\sqrt{n}) \\ &\to 1 - 0 = 0 \quad \text{per} \quad n \to \infty, \end{aligned}$$

ovvero la consistenza di $\overline{X}_n$. Il risultato era scontato dal momento che, per la Legge dei Grandi Numeri (le cui ipotesi sono soddisfatte dal modello normale), $\overline{X}_n \to \mathbb{E}_\theta[X] = \theta$, per ogni $\theta \in \Theta$. □

4.72 Definizione (Consistenza in errore quadratico medio). Dato il modello statistico $\{\mathcal{X}^n, f_n(\mathbf{x}_n; \theta), \theta \in \Theta\}$, la successione $(d_n,\ n \in \mathbb{N})$ di stimatori del parametro θ è *consistente in errore quadratico medio* (oppure *consistente in MSE*) se

$$\lim_{n\to\infty} \mathrm{MSE}_\theta(d_n) = 0, \qquad \forall \theta \in \Theta.$$

□

La consistenza in MSE è una condizione piú forte della consistenza semplice. Vale infatti il seguente risultato.

4.73 Teorema Dato il modello statistico $\{\mathcal{X}^n, f_n(\mathbf{x}_n; \theta), \theta \in \Theta\}$ e la successione $(d_n, n \in \mathbb{N})$ di stimatori di θ, se

$$\lim_{n\to\infty} \mathrm{MSE}_\theta(d_n) = 0, \qquad \forall \theta \in \Theta,$$

allora la successione degli stimatori è consistente (in senso debole). □

Dimostrazione. In base alla Disuguaglianza di Cebicev [17] si ha infatti che

$$\mathbb{P}_\theta(|d_n - \theta| \geq \epsilon) \leq \frac{\mathbb{E}_\theta[(d_n - \theta)^2]}{\epsilon^2} = \frac{\mathrm{MSE}_\theta(d_n)}{\epsilon^2}, \qquad \forall \theta \in \Theta. \tag{4.19}$$

Per $n \to \infty$ $\mathrm{MSE}_\theta(d_n)$ tende a zero e quindi, essendo $\mathbb{P}_\theta(|d_n - \theta| \geq \epsilon) \geq 0$ per ogni n e per ogni θ, necessariamente $\mathbb{P}_\theta(|d_n - \theta| \geq \epsilon) \to 0$.

4.74 Osservazione

1. Ricordando che
$$\mathrm{MSE}_\theta(d_n) = \mathbb{V}_\theta(d_n) + [B_\theta(d_n)]^2,$$
la consistenza in MSE di d_n equivale, per ogni $\theta \in \theta$ alle seguenti due condizioni
$$\lim_{n\to\infty} \mathbb{V}_\theta(d_n) = 0, \qquad \text{e} \qquad \lim_{n\to\infty} [B_\theta(d_n)]^2 = 0.$$
La consistenza in MSE della successione d_n garantisce quindi che, al crescere del numero di osservazioni, la distribuzione campionaria dello stimatore si concentra sempre piú intorno al vero valore del parametro, qualunque esso sia.

[17] Data una v.a. X, per $r > 0$ e $a > 0$ si ha
$$\mathrm{P}(|X| \geq a) \leq \frac{E|X|^r}{a^r}.$$

2. Per quanto visto, a volte la proprietà 4.70 viene denominata *consistenza debole* (o semplice), perché meno forte della consistenza in errore quadratico medio.

□

4.75 Definizione (Non distorsione asintotica). Uno stimatore d_n di θ si dice *asintoticamente non distorto* se per la successione $(d_n,\ n \in \mathbb{N})$ si ha che

$$\lim_{n\to\infty} \mathbb{E}(d_n) = \theta, \qquad \forall \theta \in \Theta$$

ovvero

$$\lim_{n\to\infty} B_\theta(d_n) = 0, \qquad \forall \theta \in \Theta.$$

□

4.76 Esempio (Modello normale). Nel caso del modello $\mathrm{N}(\theta_1, \theta_2)$ abbiamo che $\mathbb{E}_\theta[\widehat{\sigma}^2_n] = \mathbb{E}_\theta[\sum_{i=1}^n (X_i - \overline{X}_n)^2/n] = \frac{n-1}{n}\theta_2$. Lo stimatore $\widehat{\sigma}^2_n$ è quindi distorto per θ_2, ma è asintoticamente non distorto in quanto, per ogni $\theta_2 > 0$, $B_{\theta_2}(\widehat{\sigma}^2_n) = -\theta_2/n \to 0$ per $n \to \infty$. □

4.77 Esempio (Modello uniforme). Abbiamo visto in precedenza che lo stimatore di massima verosimiglianza di θ nel modello uniforme in $[0, \theta]$ è $\widehat{\theta}_{mv} = X_{(n)}$ e che, per ogni $\theta > 0$, $\mathbb{E}_\theta[X_{(n)}] = \frac{n}{n+1}\theta$. In questo caso, per ogni $\theta > 0$, si ha che $B_\theta(X_{(n)}) = \theta/(n+1) \to 0$, per $n \to \infty$ e lo smv è quindi asintoticamente non distorto. □

4.5.2 Normalità asintotica

Per valutare e utilizzare lo stimatore di un parametro è indispensabile conoscere la sua distribuzione di probabilità che però, a volte è difficile da determinare. Risultano quindi particolarmente utili eventuali *approssimazioni asintotiche* di queste distribuzioni. Abbiamo visto in precedenza (Capitolo 2) che, sfruttando il teorema del limite centrale, è possibile ottenere distribuzioni approssimate per le medie campionarie di modelli non normali. Estendiamo ora questo ragionamento al caso di successioni di stimatori che non sono necessariamente medie campionarie.

4.78 Definizione (Normalità asintotica). Dato il modello statistico $\{\mathcal{X}^n, f_n(\mathbf{x}_n; \theta), \theta \in \Theta\}$, una successione di stimatori $(d_n,\ n \in \mathbb{N})$ di $\tau(\theta)$ è *asintoticamente normale* se esiste una successione di costanti $\kappa_n > 0$ tale che, per $n \to \infty$,

$$\kappa_n(d_n - \tau(\theta)) \xrightarrow{d} \mathrm{N}(0, v(\theta)), \qquad \forall \theta \in \Theta.$$

La quantità $v(\theta)$ si chiama *varianza asintotica*. □

Si noti che, per n sufficientemente elevato, si può assumere che

$$d_n \dot{\sim} \mathrm{N}\left(\tau(\theta), \frac{v(\theta)}{\kappa_n^2}\right) \qquad \text{e} \qquad \mathbb{V}_\theta[d_n] \simeq v_a(\theta) = \frac{v(\theta)}{k_n^2},$$

dove $v_a(\theta)$ indica la varianza asintotica dello stimatore d_n di θ.

4.79 Esempio (Modello uniforme). Nel caso di un campione casuale di dimensione n da un modello uniforme in $[0, \theta]$, lo stimatore dei momenti è $\widehat{\theta}_m = 2\overline{X}_n$ è non distorto per θ e la sua varianza è pari a $\theta^2/3n$. Per il Teorema del limite centrale abbiamo quindi che

$$\kappa_n[\widehat{\theta}_m(\mathbf{X}_n) - \theta] \xrightarrow{d} \mathrm{N}(0, v(\theta))$$

per $\kappa_n = \sqrt{n}$ e $v(\theta) = \theta^2/3$. Per n sufficientemente elevato si ha quindi che

$$2\overline{X}_n \dot{\sim} \mathrm{N}\left(\theta, \frac{\theta^2}{3n}\right).$$

Si noti che, in questo caso, $\mathbb{V}_\theta(d_n) = \frac{v(\theta)}{k_n^2} = \frac{\theta^2}{3n}$ (la varianza dell'approssimazione normale coincide con la varianza esatta dello stimatore). □

Come vedremo meglio nei prossimi paragrafi, sia gli stimatori dei momenti che gli stimatori di massima verosimiglianza, sotto opportune condizioni, godono della proprietà di normalità asintotica.

4.5.3 Efficienza asintotica

Ricordiamo che, in problemi regolari di stima, uno stimatore d non distorto di $\tau(\theta)$ si dice efficiente se la sua varianza uguaglia il limite inferiore di Cramer-Rao:

$$\mathbb{V}_\theta[d] = \mathrm{cr}[\tau(\theta)] = \frac{[\tau'(\theta)]^2}{I_n(\theta)}.$$

Diamo ora la versione asintotica di questa proprietà. Supponiamo che si possa utilizzare l'espressione $I_1(\theta) = -\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2}\ln f_X(X;\theta)\right]$ (vedi Teorema 4.35). Per campioni casuali abbiamo inoltre che $I_n(\theta) = nI_1(\theta)$.

4.80 Definizione. **(Efficienza asintotica).** Dato il modello statistico $\{\mathcal{X}^n, f_n(\mathbf{x}_n;\theta), \theta \in \Theta\}$, una successione di stimatori $(d_n,\ n \in \mathbb{N})$ di $\tau(\theta)$ è *asintoticamente efficiente* se

$$\sqrt{n}[d_n - \tau(\theta)] \xrightarrow{d} \mathrm{N}[0, v(\theta)]$$

e

$$v(\theta) = \frac{[\tau'(\theta)]^2}{I_1(\theta)} = \frac{[\tau'(\theta)]^2}{\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f_X(X;\theta)\right)^2\right]}.$$

□

Dal precedente risultato si ha quindi che, per n sufficientemente elevato,

$$\mathbb{V}_\theta[d_n] \simeq \frac{v(\theta)}{n} = \frac{[\tau'(\theta)]^2}{nI_1(\theta)},$$

dove $I_1(\theta) = \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f_X(X;\theta)\right)^2\right]$. La successione degli stimatori d_n di $\tau(\theta)$ è quindi asintoticamente efficiente se

$$\lim_{n\to\infty}\frac{\mathbb{V}_\theta[d_n]}{\mathrm{cr}[\tau(\theta)]} = 1.$$

4.81 Esempio. Nel caso del modello $\mathrm{N}(\theta_1, \theta_2)$ abbiamo visto che $\mathbb{V}_\theta[S_n^2] = 2\theta_2^2/(n-1) > 2\theta_2^2/n = \mathrm{cr}(\theta_2)$ e quindi che lo stimatore S_n^2 non è efficiente. Tuttavia

$$\lim_{n\to\infty}\frac{\mathbb{V}_\theta[S_n^2]}{\mathrm{cr}(\theta_2)} = 1,$$

e quindi, poichè mostreremo che la distribuzione di S_n^2 è asintoticamente normale, possiamo anche dire che S_n^2 è asintoticamente efficiente (ottimo) per θ_2. □

4.82 Osservazione.

1. Uno stimatore asintoticamente effieciente è asintoticamente non distorto e con varianza asintoticamente equivalente al `licr`.

2. La definizione di efficienza asintotica che abbiamo dato in questo testo fa riferimento al limite inferiore di Cramer-Rao, che però è valida nei problemi regolari di stima. Tuttavia è ragionevole definire asintoticamente efficiente uno stimatore con varianza asintoticamente equivalente (rapporto tendente a 1 al crescere di n) a quella dello stimatore UMVUE.

□

4.6 Proprietà degli stimatori dei momenti

Ricordiamo che abbiamo indicato con

$$\mu_r(\theta) = \mathbb{E}_\theta[X_i^r] \qquad \text{e} \qquad m_r(\mathbf{X}_n) = \frac{1}{n}\sum_{i=1}^n X_i^r$$

rispettivamente il momento di ordine r di X_i (che assumiamo esistere) e il momento campionario di ordine $r \geq 1$. Lo stimatore dei momenti di un parametro θ è una funzione dei momenti campionari dai quali, sotto opportune condizioni, eredita le proprietà. Il momento campionario $m_r(\mathbf{X}_n)$ è infatti la media campionaria delle v.a. $X_1^r, \ldots, X_n^r$, ovvero uno stimatore di $\mu_r(\theta)$ dotato di molte proprietà frequentiste, riassunte nel teorema che segue.

4.83 Teorema (Proprietà dei momenti campionari). Dato il modello statistico $\{\mathcal{X}^n, f_n(\mathbf{x}_n; \theta), \theta \in \Theta\}$ e un campione casuale $X_1, \ldots, X_n$, siano $\mu_r(\theta)$ e $m_r(\mathbf{X}_n)$ il momento di ordine r di X_i (che assumiamo esistere) e il momento campionario di ordine $r \geq 1$. Supponendo che esista il momento di ordine $2r$ di X_i, $\mu_{2r}(\theta)$, si ha allora che:

a) lo stimatore $m_r(\mathbf{X}_n)$ è uno stimatore non distorto di $\mu_r(\theta)$;

b) $\text{MSE}_\theta(m_r) = \mathbb{V}_\theta[m_r(\mathbf{X}_n)] = \frac{1}{n}[\mu_{2r}(\theta) - \mu_r(\theta)^2]$;

c) lo stimatore m_r è consistente per $\mu_r(\theta)$;

d) al crescere di n, la successione degli stimatori $m_r(\mathbf{X}_n)$ standardizzati converge in distribuzione alla v.a. $\text{N}(0,1)$:

$$\frac{m_r(\mathbf{X}_n) - \mathbb{E}_\theta[m_r(\mathbf{X}_n)]}{\sqrt{\mathbb{V}_\theta(m_r(\mathbf{X}_n))}} = \frac{\sqrt{n}[m_r(\mathbf{X}_n) - \mu_r(\theta)]}{\sqrt{\mu_{2r}(\theta) - \mu_r^2(\theta)}} \xrightarrow{d} \text{N}(0,1), \quad \forall \theta \in \Theta.$$

□

Dimostrazione. Per la parte a), si osservi che, per le proprietà del valore atteso si ha che

$$\mathbb{E}_\theta[m_r(\mathbf{X}_n)] = \mathbb{E}_\theta\left(\frac{1}{n}\sum_{i=1}^n X_i^r\right) = \mathbb{E}_\theta(X_i^r) = \mu_r(\theta), \qquad \forall \theta \in \Theta.$$

Per b), si ha che

$$\text{MSE}_\theta(m_r) = \mathbb{V}_\theta\left(\frac{1}{n}\sum_{i=1}^n X_i^r\right) = \frac{1}{n}\mathbb{V}_\theta(X_i^r) = \frac{1}{n}[\mathbb{E}_\theta(X^{2r}) - [\mathbb{E}_\theta(X^r)]^2] = \frac{1}{n}[\mu_{2r}(\theta) - \mu_r(\theta)^2].$$

Per quanto riguarda c), l'ipotesi di esistenza dei momenti di ordine r e $2r$ garantisce che, per $n \to \infty$, la successione degli errori quadratici medi $\text{MSE}_\theta[m_r(\mathbf{X}_n)]$ tende a zero per ogni

θ, ovvero che m_r è consistente in errore quadratico medio (e quindi anche in senso semplice) per $\mu_r(\theta)$. La normalità asintotica di m_r discende dal teorema del limite centrale. Per n sufficientemente elevato si ha quindi che

$$m_r(\mathbf{X}_n) \dot{\sim} \mathrm{N}\left(\mu_r(\theta), \frac{1}{n}[\mu_{2r}(\theta) - \mu_r(\theta)^2]\right).$$

Ricordiamo che, se il parametro è un vettore $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$, $k \geq 1$, lo stimatore dei momenti è soluzione del sistema

$$\mu_r(\boldsymbol{\theta}) = m_r(\mathbf{X}_n), \qquad r = 1, \ldots, k.$$

La soluzione di questo sistema è una funzione dei k momenti campionari coinvolti, ovvero

$$\widehat{\boldsymbol{\theta}}_m = (\widehat{\theta}_{m,1}, \ldots, \widehat{\theta}_{m,k}) = \mathbf{g}[m_1(\mathbf{X}_n), \ldots, m_k(\mathbf{X}_n)].$$

Mostriamo ora che, sotto ipotesi abbastanza generali, lo stimatore dei momenti è consistente.

4.84 Teorema (Consistenza dello stimatore dei momenti). Dato un modello statistico dipendente dal parametro $\boldsymbol{\theta}$ e una successione di osservazioni i.i.d. $(X_n,\ n \in \mathbb{N})$, se esistoni finiti i momenti delle v.a. X_i fino all'ordine k e se la soluzione del sistema dei momenti è $\widehat{\boldsymbol{\theta}}_m = (\widehat{\theta}_{m,1}, \ldots, \widehat{\theta}_{m,k}) = \mathbf{g}[m_1(\mathbf{X}_n), \ldots, m_k(\mathbf{X}_n)]$, con $\mathbf{g}$ funzione continua in $\mathbb{R}^k$, allora la successione degli stimatori $\widehat{\boldsymbol{\theta}}_m(\mathbf{X}_n)$ converge in probabilità a $\boldsymbol{\theta}$, per ogni $\boldsymbol{\theta} \in \Theta$. □

Dimostrazione. Per le ipotesi fatte, il precedente teorema (ovvero la legge dei grandi numeri) assicura che, per ogni $\boldsymbol{\theta}$,

$$m_r(\mathbf{X}_n) \xrightarrow{p} \mu_r(\boldsymbol{\theta}), \qquad r = 1, \ldots, k. \tag{4.20}$$

Per il teorema che garantisce la convergenza in probabilità di funzioni continue di vettori aleatori (si veda Casella-Berger (2002), Teorema 5.5.4, p.233), si ha allora che

$$\widehat{\boldsymbol{\theta}}_m(\mathbf{X}_n) = \mathbf{g}[m_1(\mathbf{X}_n), \ldots, m_k(\mathbf{X}_n)] \xrightarrow{p} \mathbf{g}[\mu_1(\boldsymbol{\theta}), \ldots, \mu_k(\boldsymbol{\theta})] = \boldsymbol{\theta}, \tag{4.21}$$

ovvero che $\widehat{\boldsymbol{\theta}}_m$ è consistente per $\boldsymbol{\theta}$. L'ultima uguaglianza della (4.21) si giustifica osservando, che per ogni n, $\mathbf{g}(m_1, \ldots, m_k)$ è soluzione del sistema $\mu_r(\boldsymbol{\theta}) = m_r(\mathbf{X}_n)$, $r = 1, \ldots, k$. e quindi, per $n \to \infty$,

$$\mu_r(\mathbf{g}[\mu_1(\boldsymbol{\theta}), \ldots, \mu_k(\boldsymbol{\theta})]) = \mu_r(\boldsymbol{\theta}), \qquad r = 1, \ldots, k,$$

ovvero che $\mathbf{g}[\mu_1(\boldsymbol{\theta}), \ldots, \mu_k(\boldsymbol{\theta})] = \boldsymbol{\theta}$.

Sotto opportune condizioni è possibile anche mostrare che gli stimatori dei momenti sono asintoticamente normali. Nel caso di un parametro scalare ($k = 1$), si ha che, per $n \to \infty$

$$\frac{\widehat{\theta}_m(\mathbf{X}_n) - \theta}{\sqrt{\mathbb{V}_\theta[\widehat{\theta}_m(\mathbf{X}_n)]}} \xrightarrow{d} \mathrm{N}(0, 1), \qquad \forall \theta \in \Theta.$$

Si noti che $\mathbb{V}_\theta[\widehat{\theta}_m(\mathbf{X}_n)]$ è una quantità che dipende da θ. Spesso non è agevole determinare la varianza di $\widehat{\theta}_m$ ma si può ricorrere al metodo delta (vedi Capitolo 2) per ottenerne un'approssimazione.

4.85 Esempio Nel caso del modello Beta$(\theta, 1)$, abbiamo visto in precedenza che $\widehat{\theta}_m(\mathbf{X}_n) = \overline{X}_n/(1-\overline{X}_n)$. Ponendo

$$\mathbb{E}_\theta[X] = \frac{\theta}{\theta+1} = \psi \qquad \text{e} \qquad \mathbb{V}_\theta[X] = \frac{\theta}{(\theta+2)(\theta+1)^2} = \sigma^2,$$

abbiamo che

$$\sqrt{n}(\overline{X}_n - \psi) \xrightarrow{d} \mathrm{N}(0, \sigma^2)$$

e quindi, per il metodo delta (Capitolo 2),

$$\sqrt{n}\left[g(\overline{X}_n) - g(\psi)\right] \xrightarrow{p} \mathrm{N}(0, g'(\psi)^2\sigma^2)$$

per ogni funzione g tale che $g'(\psi) = \frac{\partial}{\partial\psi}g(\psi) \neq 0$. In questo caso, poniamo

$$g(\overline{X}_n) = \frac{\overline{X}_n}{1-\overline{X}_n}$$

e quindi $g(\psi) = \psi/(1-\psi) = \theta$. Si ha allora che

$$g'(\psi) = \frac{1}{(1-\psi)^2} = (\theta+1)^2 \qquad \Rightarrow \qquad g'(\psi)^2\sigma^2 = \frac{\theta(\theta+1)^2}{\theta+2}.$$

Si ottiene quindi che

$$\sqrt{n}\left(\frac{\overline{X}_n}{1-\overline{X}_n} - \theta\right) \xrightarrow{d} \mathrm{N}\left(0, \frac{\theta(\theta+1)^2}{\theta+2}\right)$$

ovvero che, per n sufficientemente elevato,

$$\frac{\overline{X}_n}{1-\overline{X}_n} \dot{\sim} \mathrm{N}\left(\theta, \frac{\theta(\theta+1)^2}{n(\theta+2)}\right)$$

da cui si evince la non distorsione asintotica e la consistenza di $\widehat{\theta}_m(\mathbf{X}_n)$.

□

4.7 Proprietà degli stimatori di massima verosimiglianza

Gli stimatori di massima verosimiglianza, sotto opportune condizioni soddisfatte dai modelli parametrici piú comuni, godono di numerose proprietà, sia asintotiche che per n finito. Nel seguito, per rendere esplicita la dipendenza dalla dimensione campionaria n, indicheremo con $\widehat{\theta}_n$ lo stimatore di mv $\widehat{\theta}_{mv}(\mathbf{X}_n)$ e con $(\widehat{\theta}_n,\ n \in \mathbb{N})$ la successione di v.a. definita dagli stimatori di mv. Le principali proprietà degli stimatori di mv sono:

1. **Equivarianza**: se $\widehat{\theta}_{mv}$ è lo stimatore di mv di θ, allora $g(\widehat{\theta}_{mv})$ è lo stimatore di mv di $g(\theta)$.

2. **Consistenza**: la successione degli stimatori di mv $(\widehat{\theta}_n,\ n \in \mathbb{N})$ converge in probabilità al valore vero del parametro, θ^*.

3. **Normalità asintotica**: la successione $\sqrt{n}(\widehat{\theta}_n - \theta)$ converge in distribuzione a $\mathrm{N}(0, I_1(\theta)^{-1})$. Ciò vuol dire che, per n sufficientemente elevato, la distribuzione campionaria di $\widehat{\theta}_{mv}(\mathbf{X}_n)$ può essere approssimata con una densità normale.

4. **Normalità asintotica per funzioni dello stimatore di mv**: se g ha derivata non nulla, allora $\sqrt{n}[g(\widehat{\theta}_n) - g(\theta)]$ converge in distribuzione a $\mathrm{N}(0, g'(\theta)^2 I_1(\theta)^{-1})$. Ciò vuol dire che, per n sufficientemente elevato, la distribuzione campionaria di $g(\widehat{\theta}_{mv}(\mathbf{X}_n))$ può essere approssimata con una densità normale.

5. **Ottimalità (o efficienza) asintotica**: per i due punti precedenti, gli stimatori di massima verosimiglianza di θ e di $g(\theta)$ sono asintoticamente non distorti e con varianza asintoticamente equivalente al limite inferiore di Cramer-Rao.

4.7.1 Consistenza

In questo paragrafo enunciamo il teorema di consistenza per stimatori di massima verosimiglianza. Seguiamo qui la presentazione del problema proposta in Casella-Berger (2002, Cap. 10). Per enunciare il teorema di consistenza è necessario premettere alcune definizioni e delle condizioni di regolarità. Una dimostrazione del teorema in condizioni abbastanza generali è riportata in Appendice[18].

4.86 Definizione (Divergenza di Kullback-Leibler). Date due funzioni di densità f e g (con supporto $\mathcal{X}$), la *divergenza di Kullback-Leibler* tra f e g è data da

$$D(f, g) = \int_{\mathcal{X}} f(x) \log \frac{f(x)}{g(x)} dx.$$

□

Si può mostrare che, per ogni coppia f, g, si ha $D(f, g) \geq 0$ e che $D(f, g) = 0$ se e solo se $f = g$ quasi ovunque. Si noti che D non è una distanza, in quanto $D(f, g) \neq D(g, f)$ se $f \neq g$.

4.87 Definizione (Modello identificabile). Un modello statistico parametrico $\{\mathcal{X}, f_X(\cdot;\theta), \theta \in \Theta\}$ è *identificabile* se, per ogni coppia $\theta_1 \neq \theta_2$ si ha che $D[f_X(\cdot;\theta_1), f_X(x;\theta_2)] > 0$, ovvero che $f_X(\cdot;\theta_1) \neq f_X(\cdot;\theta_2)$. □

Per semplicità, indichiamo $D[f_X(\cdot;\theta_1), f_X(x;\theta_2)] > 0$ con $D(\theta_1, \theta_2)$.

Formuliamo ora le condizioni per garantire la consistenza delle successioni di stimatori di massima verosimiglianza[19].

4.88 Definizione (Condizioni di regolarità per consistenza smv). Dato il modello statistico parametrico $\{\mathcal{X}, f_X(\cdot;\theta), \theta \in \Theta\}$, le seguenti assunzioni sono sufficienti per dimostrare il successivo Teorema (4.89):
a) le v.a. $X_1, X_2, \ldots$ sono i.i.d. con distribuzione $f_X(\cdot;\theta)$;
b) il modello statistico è identificabile;
c) le distribuzioni $f_X(\cdot;\theta)$ hanno stesso supporto per ogni θ e $f_X(\cdot;\theta)$ è differenziabile rispetto a θ;
d) lo spazio parametrico Θ contiene un insieme aperto al quale appartiene il valore vero del parametro, θ^*. □

4.89 Teorema (Consistenza degli smv). Dato il modello statistico $\{\mathcal{X}, f_X(\cdot;\theta), \theta \in \Theta\}$ e una successione di v.a. $(X_n,\ n \in \mathbb{N})$ i.i.d. ciascuna con distribuzione $f_X(\cdot;\theta)$, sia $L(\theta; \mathbf{x}_n) = \prod_{i=1}^n f_X(x_i;\theta)$ la funzione di verosimiglianza di θ per un campione di dimensione

[18] Per la dimostrazione generale si veda, ad esempio, Stuart, Odd and Arnold (1999, Capitolo 18) o Lehmann e Casella (1998, Paragrafo 6.3)

[19] Casella-Berger (2002) osservano che è possibile dare condizioni ancora piú generali, ma tralasciamo questo aspetto troppo tecnico per la presente trattazione.

n e $\left(\widehat{\theta}_n,\ n \in \mathbb{N}\right)$ la successione degli stimatori di massima verosimiglianza di θ. Indicato con θ^* il vero valore del parametro, sotto le condizioni espresse nella Definizione 4.88 si ha che

$$\widehat{\theta}_n \xrightarrow{p} \theta^*.$$

Inoltre, per ogni funzione continua g di θ,

$$g(\widehat{\theta}_n) \xrightarrow{p} g(\theta^*).$$

□

La dimostrazione della prima parte del teorema è riportata in appendice.

4.90 Esempio Il modello normale soddisfa le ipotesi del teorema precedente. Nel caso del modello $\mathrm{N}(\mu_0, \theta)$ (valore atteso noto) $\widehat{\theta}_n = S_0^2 \xrightarrow{p} \theta$. Nel modello $\mathrm{N}(\theta_1, \theta_2)$ si ha $\widehat{\theta}_n = \widehat{\sigma}_n^2 \xrightarrow{p} \theta_2$. Si noti che sia S_0^2 che S_n^2 sono stimatori consistenti in MSE e quindi anche consistenti in senso debole. □

4.91 Esempio Il modello esponenziale negativo soddisfa le ipotesi del teorema precedente. In questo caso $\widehat{\theta}_n = \frac{1}{\overline{X}_n} \xrightarrow{p} \theta$. □

4.7.2 Normalità asintotica

La dimostrazione della normalità asintotica dello stimatore di mv richiede ulteriori condizioni di regolarità (oltre alle 4.88) per $f_X(\cdot;\theta)$, specificate, ad esempio, in Casella-Berger (2002, p. 516). Queste ipotesi sono soddisfatte nei modelli piú comuni, utilizzati in questo testo. La dimostrazione del risultato si trova nell Appendice 4.9.2.

4.92 Teorema (Normalità asintotica degli smv). Dato il modello statistico $\{\mathcal{X}, f_X(\cdot;\theta), \theta \in \Theta\}$ e una successione di v.a. $(X_n,\ n \in \mathbb{N})$ i.i.d. ciascuna con distribuzione $f_X(\cdot;\theta)$, sia $L(\theta; \mathbf{x}_n) = \prod_{i=1}^n f_X(x_i;\theta)$ la funzione di verosimiglianza di θ per un campione di dimensione n e $\left(\widehat{\theta}_n,\ n \in \mathbb{N}\right)$ la successione degli stimatori di massima verosimiglianza di θ. Sotto le condizioni di regolarità 4.88 e quelle aggiuntive in Casella-Berger (2002, p.516), si ha che

$$\frac{\widehat{\theta}_n - \theta}{\sqrt{I_n(\theta)^{-1}}} \xrightarrow{d} \mathrm{N}(0,1), \tag{4.22}$$

Inoltre

$$\frac{\widehat{\theta}_n - \theta}{\sqrt{I_n(\widehat{\theta}_n)^{-1}}} \xrightarrow{d} \mathrm{N}(0,1). \tag{4.23}$$

□

4.93 Osservazione
a) Si noti che la 4.22 equivale a

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} \mathrm{N}\left(0, [I_1(\theta)]^{-1}\right).$$

che ci permette di affermare che lo smv è asintoticamente efficiente (o asintoticamente ottimo).

b) Dal punto di vista pratico il teorema precedente ci consente di dire che, per n sufficientemente elevato

$$\widehat{\theta}_n \dot{\sim} \mathrm{N}\widehat{\theta}_n \dot{\sim} \mathrm{N}\left(\theta, \frac{1}{nI_1(\theta)}\right).$$

c) Dal teorema discende anche che

$$\mathbb{V}_\theta(\widehat{\theta}_n) \simeq [nI_1(\theta)]^{-1} = v_a(\theta),$$

dove v_a indica la varianza asintotica di $\widehat{\theta}_n$. Esistono due possibili stimatori $\widehat{I_n(\theta)}$ per l'informazione attesa di Fisher:

$$nI_1(\widehat{\theta}_n) \qquad \text{e} \qquad \mathcal{I}_n(\mathbf{X}_n).$$

Infatti, se $\mathcal{I}_n = -\frac{\partial^2}{\partial\theta^2} \ln L(\theta; \mathbf{X}_n)$ è una funzione continua di θ, $\mathcal{I}_n(\mathbf{X}_n) = -\frac{\partial^2}{\partial\theta^2} \ln L(\theta; \mathbf{X}_n)|_{\theta=\widehat{\theta}_n} \xrightarrow{p} I_n(\theta)$. Per la varianza asintotica di $\widehat{\theta}_n$ abbiamo quindi due stimatori:

$$\widehat{v_a(\widehat{\theta}_n)} = I_n(\widehat{\theta}_n)^{-1} \qquad \text{e} \qquad \widehat{v_a(\widehat{\theta}_n)} = \mathcal{I}_n(\mathbf{X}_n)^{-1}.$$

Abbiamo quindi il seguente ulteriore risultato asintotico:

$$\frac{\widehat{\theta}_n - \theta}{\sqrt{[\mathcal{I}_n(\mathbf{X}_n)]^{-1}}} \xrightarrow{d} \mathrm{N}(0,1). \tag{4.24}$$

Per i casi in cui i due stimatori non coincidono, Efron e Hinkley (1978)[20] dimostrano la superiorità dello stimatore basato sull'informazione osservata rispetto a $I_n(\widehat{\theta}_n)^{-1}$. □

4.94 Esempio (Modello esponenziale negativo). Nel caso di un campione casuale da un modello esponenziale negativo, lo smv di θ è $\widehat{\theta}_n = 1/\overline{X}_n$ e $I_n(\theta) = n/\theta^2$. Abbiamo quindi che

$$\frac{\sqrt{n}({\overline{X}_n}^{-1} - \theta)}{\theta} \xrightarrow{d} \mathrm{N}(0,1).$$

Per n elevato si ha

$$\frac{1}{\overline{X}_n} \dot{\sim} \mathrm{N}\left(\theta, \frac{\theta^2}{n}\right).$$

Lo stimatore della varianza asintotica di ${\overline{X}_n}^{-1}$ è quindi $\widehat{v_a({\overline{X}_n}^{-1})} = 1/(n{\overline{X}_n}^2)$. □

4.95 Esempio (Modello normale). Nel caso del modello normale $\mathrm{N}(\theta_1, \theta_2)$, si ha che $I_n(\theta_2) = n/(2\theta_2^2)$ e quindi

$$\sqrt{n}(\widehat{\sigma}_n^2 - \theta_2) \xrightarrow{d} \mathrm{N}(0, 2\theta_2^2)$$

Per n elevato possiamo quindi usare la seguente approssimazione

$$\widehat{\sigma}_n^2 \dot{\sim} \mathrm{N}\left(\theta_2, \frac{2\theta_2^2}{n}\right).$$

Lo stimatore della varianza asintotica di $\widehat{\sigma}_n^2$ è quindi $\widehat{v_a(\widehat{\sigma}_n^2)} = 2\widehat{\sigma}_n^4/n$. In questo esempio è possibile confrontare la distribuzione asintotica di $\widehat{\sigma}^2$ con quella esatta: $\widehat{\sigma}_n^2 \sim \mathrm{Gamma}[(n -$

[20] Efron B.F. e Hinkley D.V. (1978). Assessing the accuracy of the maximum likelihood estimator: observed versus expected Fisher information. *Biometrika*, 65, 457-487.

$1)/2, n/(2\sigma^2)]$. Possiamo anche confrontare la varianza esatta con varianza asintotica. Abbiamo che $\mathbb{V}_\theta(\widehat{\sigma}^2_n) = \frac{2(n-1)}{n^2}\theta_2^2$ e quindi

$$\frac{\mathbb{V}_\theta(\widehat{\sigma}^2_n)}{v_a(\widehat{\sigma}^2_n)} = \frac{\frac{2(n-1)}{n^2}\theta_2^2}{\frac{2\theta_2^2}{n}} = \frac{n-1}{n} \to 1.$$

□

4.96 Esempio (Modello beta). Consideriamo il modello Beta$(\theta, 1)$ in cui $f_X(x;\theta) = \theta x^{\theta-1}, \quad \theta > 0,\ x \in [0,1]$. Nel caso di campioni casuali, si verifica facilmente che $\widehat{\theta}_n = -n/\sum_{i=1}^n \ln X_i$ e che $I_n(\theta) = n/\theta^2$. Per n elevato possiamo quindi usare la seguente approssimazione

$$-\frac{n}{\sum_{i=1}^n \ln X_i} \dot{\sim} \mathrm{N}\left(\theta, \frac{\theta^2}{n}\right).$$

Lo stimatore della varianza asintotica è quindi $n/(\sum_{i=1}^n \ln X_i)^2$. □

4.7.3 Normalità asintotica per funzioni dello stimatore di mv

Si tratta di una applicazione del metodo delta al caso in cui si considerano funzioni degli stimatori di massima verosimiglianza.

4.97 Teorema (Metodo delta per funzioni di smv). Nelle ipotesi del Teorema 4.92, se g è una funzione di θ con derivata prima $g'(\theta) \neq 0$, allora

$$\frac{g(\widehat{\theta}_n) - g(\theta)}{|g'(\theta)|\sqrt{I_n(\theta)^{-1}}} \xrightarrow{d} \mathrm{N}(0,1).$$

Inoltre,

$$\frac{g(\widehat{\theta}_n) - g(\theta)}{|g'(\widehat{\theta}_n)|\sqrt{\widehat{I_n(\theta)}^{-1}}} \xrightarrow{d} \mathrm{N}(0,1).$$

dove $\widehat{I_n(\theta)}$ può essere sia $I_n(\widehat{\theta}_n)$ che $\mathcal{I}_n(\mathbf{X}_n)$. □

Dimostrazione. Si applica il metodo delta alla successione $\sqrt{n}(\widehat{\theta}_n - \theta)$ che, per il Teorema 4.92, converge in distribuzione a $\mathrm{N}(0,1)$. La dimostrazione della seconda parte del teorema si ottiene osservando che $I_n(\widehat{\theta}_n)$ e $\mathcal{I}_n(\mathbf{X}_n)$ convergono entrambe in probabilità a $I_n(\theta)$ e applicando il Teorema di Slutsky in modo analogo a quanto fatto nel Teorema 4.92.

4.98 Osservazione. Dal teorema discende che, per n grande

$$\frac{g(\widehat{\theta}_n) - g(\theta)}{|g'(\theta)|\sqrt{I_n(\theta)^{-1}}} \dot{\sim} \mathrm{N}(0,1)$$

ovvero che

$$g(\widehat{\theta}_{mv}) \dot{\sim} \mathrm{N}\left(g(\theta), \frac{g'(\theta)^2}{I_n(\theta)}\right).$$

La varianza asintotica di $g(\widehat{\theta}_n)$ e il suo stimatore sono rispettivamente

$$v_a[g(\widehat{\theta}_n)] = g'(\theta)^2 I_n(\theta)^{-1} \qquad \text{e} \qquad \widehat{v_a[g(\widehat{\theta}_n)]} = g'(\widehat{\theta}_n)^2 \widehat{I_n(\theta)}^{-1}$$

dove $\widehat{I_n(\theta)}$ può essere, al solito, l'informazione osservata o l'informazione attesa calcolata in $\widehat{\theta}_n$. □

4.99 Esempio (**Stima del log-odds.**) Consideriamo il modello Ber(θ). Si definiscono ripettivamente *odds* e *log-odds* le quantità

$$\frac{\theta}{1-\theta}, \qquad \text{e} \qquad \log\frac{\theta}{1-\theta}.$$

Consideriamo la seconda quantità. Per la proprietà di equivarianza, lo smv di $g(\theta) = \log[\theta/(1-\theta)]$ è $g(\widehat{\theta}_n) = \log \overline{X}_n/(1-\overline{X}_n)$. Con semplici calcoli si verifica che $g'(\theta) = [\theta(1-\theta)]^{-1}$. Nel modello bernoulliano $I_n(\theta) = n/[\theta(1-\theta)]$. Si ha quindi che $g'(\theta)^2 I_n(\theta)^{-1} = v_a[g(\widehat{\theta}_n)] = [n\theta(1-\theta)]^{-1}$ e che

$$\log\frac{\overline{X}_n}{1-\overline{X}_n} \mathrel{\dot\sim} \mathrm{N}\left(\log\frac{\theta}{1-\theta}, \frac{1}{n\theta(1-\theta)}\right).$$

Lo stimatore della varianza asintotica di $g(\widehat{\theta}_n)$ è

$$\widehat{v_a[g(\widehat{\theta}_n)]} = g'(\widehat{\theta}_n)^2 \mathcal{I}_n(\mathbf{X}_n)^{-1} = \frac{1}{n\overline{X}_n(1-\overline{X}_n)}.$$

□

4.100 Esempio (**Stima di una probabilità, modello di Poisson**.) Consideriamo il modello Pois(θ) e supponiamo di voler stimare $g(\theta) = \mathbb{P}_\theta(X=0) = e^{-\theta}$. La stima di mv di $g(\theta)$ è quindi (invarianza) $g(\widehat{\theta}_n) = e^{-\overline{X}_n}$. In questo caso abbiamo $I_n(\theta) = n/\theta$ e $g'(\theta) = -e^{-\theta}$. Pertanto la varianza asintotica di $g(\widehat{\theta}_n)$ è $v_a[g(\widehat{\theta}_n)] = g'(\theta)^2 I_n(\theta)^{-1} = e^{-2\theta}\theta/n$ e inoltre

$$g(\widehat{\theta}_n) = e^{-\overline{X}_n} \mathrel{\dot\sim} \mathrm{N}\left(e^{-\theta}, e^{-2\theta}\frac{\theta}{n}\right).$$

La stima della varianza asintotica di $g(\widehat{\theta}_n)$ è $\widehat{v_a[g(\widehat{\theta}_n)]} = g'(\widehat{\theta}_n)^2 \mathcal{I}_n(\mathbf{X}_n)^{-1} = e^{-2\overline{X}_n}\overline{X}_n/n$. □

4.8 Il caso multiparametrico

Molti dei risultati riportati nei paragrafi precedenti valgono, nel modo in cui sono stati enunciati, per una dimensione arbitraria (finita) del parametro incognito del modello. Ad esempio, i teoremi di Rao-Blackwell e Lehmann-Scheffè non specificano la dimensione del parametro e sono quindi validi anche nel caso vettoriale, in cui $\dim(\boldsymbol{\theta}) = k > 1$. Necessitano invece di una generalizzazione (in forma matriciale) la definizione di informazione di Fisher e i risultati collegati (teorema di Cramer-Rao, definizione di normalità ed efficienza asintotica, metodo delta...). Ci poniamo in un caso del tutto regolare, in cui la funzione di log-verosimiglianza è differenziabile due volte e in cui sono leciti gli scambi tra derivata e integrale.

4.101 Definizione (Matrice di informazione). Sia $\{\mathcal{X}^n, f_n(\cdot;\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$ un modello statistico (per campioni i.i.d.) in cui $\boldsymbol{\theta}^T = (\theta_1, \ldots, \theta_k) \in \Theta \subseteq \mathbb{R}^k$, $\ell_n(\boldsymbol{\theta}) = \ell_n(\boldsymbol{\theta}; \mathbf{X}_n) = \sum_{i=1}^n \ln f_X(X_i;\boldsymbol{\theta})$ indica la funzione di log-verosimiglianza e sia H la matrice hessiana di dimensione (k,k) di ℓ_n, i cui elementi generici sono:

$$H_{ii} = \frac{\partial^2}{\partial\theta_i^2}\ell_n(\boldsymbol{\theta}), \qquad H_{ij} = \frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_n(\boldsymbol{\theta}), \qquad i,j = 1,\ldots,k, \quad i \neq j.$$

Si definisce *matrice di informazione attesa di Fisher* la matrice semidefinita positiva $I_n(\theta)$ il cui elemento generico è

$$I_n(\boldsymbol{\theta})_{ij} = -\mathbb{E}_\theta(H_{ij}), \qquad i = 1, \ldots, k, \quad j = 1, \ldots, k.$$

□

4.102 Esempio (Modello normale). Consideriamo il modello $\mathrm{N}(\theta_1, \theta_2)$, in cui $\boldsymbol{\theta} = (\theta_1, \theta_2)$. Abbiamo mostrato in precedenza che, nel caso di un campione casuale di dimensione n, la matrice hessiana è

$$H = \begin{pmatrix} -\frac{n}{\theta_2} & \frac{\sum(X_i-\theta_1)}{\theta_2^2} \\ -\frac{\sum(X_i-\theta_1)}{\theta_2^2} & \frac{n}{2\theta_2^2} - \frac{\sum(X_i-\theta_1)^2}{\theta_2^3} \end{pmatrix}$$

da cui si ottiene che

$$I_n(\boldsymbol{\theta}) = -\mathbb{E}_\theta(H) = -\begin{pmatrix} -\frac{n}{\theta_2} & 0 \\ 0 & \frac{n}{2\theta_2^2} - \frac{n\theta_2}{\theta_2^3} \end{pmatrix} = \begin{pmatrix} \frac{n}{\theta_2} & 0 \\ 0 & \frac{n}{2\theta_2^2} \end{pmatrix}$$

e quindi che

$$J_n(\boldsymbol{\theta}) = I_n(\boldsymbol{\theta})^{-1} = \begin{pmatrix} \frac{\theta_2}{n} & 0 \\ 0 & \frac{2\theta_2^2}{n} \end{pmatrix}.$$

□

Disuguaglianza di Cramer-Rao ($\mathbf{k > 1}$)

Consideriamo ora uno $\widehat{\boldsymbol{\theta}}$ stimatore di $\boldsymbol{\theta}$: $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \ldots, \widehat{\theta}_k)$. Indichiamo con $\mathbb{V}_\theta(\widehat{\boldsymbol{\theta}}) = \mathbb{E}_\theta[(\widehat{\boldsymbol{\theta}} - \mathbb{E}_\theta(\boldsymbol{\theta}))(\widehat{\boldsymbol{\theta}} - \mathbb{E}_\theta(\widehat{\boldsymbol{\theta}}))^T]$ la matrice delle varianze e covarianze di $\widehat{\boldsymbol{\theta}}$. il cui generico elemento (i, j), $i \neq j$, è la covarianza di $(\widehat{\theta}_i, \widehat{\theta}_j)$ mentre il generico elemento (i, i) è la varianza di $\widehat{\theta}_i$, $i = 1, \ldots, k$. Indichiamo con $J_n(\boldsymbol{\theta}) = I_n(\boldsymbol{\theta})^{-1}$ la matrice inversa di I_n. Sotto le ipotesi di regolarità usuali, si ha allora che

$$\mathbb{V}_\theta(\widehat{\boldsymbol{\theta}}) \geq J_n(\boldsymbol{\theta})$$

nel senso che la matrice $\mathbb{V}_\theta(\widehat{\boldsymbol{\theta}}) - J_n(\boldsymbol{\theta})$ è semidefinita positiva.

Normalità asintotica

Consideriamo la successione degli smv (vettori) $\left(\widehat{\boldsymbol{\theta}}_n^T = (\widehat{\theta}_{1,n}, \ldots, \widehat{\theta}_{k,n}),\ n \in \mathbb{N}\right)$. Si ha allora che[21]

$$I_n(\boldsymbol{\theta})^{-1/2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{d} \mathrm{N}(0_k, \mathbf{I}_k),$$

dove 0_k e $\mathbf{I}_k$ indicano rispettivamente il vettore nullo di dimensione k e la matrice identità di ordine k. Per n sufficientemente grande, si ha quindi che

$$\widehat{\boldsymbol{\theta}}_n \dot{\sim} \mathrm{N}[\boldsymbol{\theta}, J_n(\boldsymbol{\theta})].$$

Dalla normalità asintotica del vettore $\widehat{\boldsymbol{\theta}}_n$ seguono le proprietà di consistenza e ottimalità (efficienza) asintotica dello stimatore di mv anche per $k > 1$.

[21] Si ricordi che, data una matrice quadrata A, la matrice $A^{1/2}$ è la matrice di stesse dimensioni tale che $A^{1/2}A^{1/2} = A$.

Metodo delta multiparametrico

Consideriamo ora una funzione $g(\boldsymbol{\theta})$ del vettore $k-$dimensionale $\boldsymbol{\theta}$ e assumiamo in particolare che $g : \Theta \to \mathbb{R}^1$. Sia, inoltre, $\nabla g(\boldsymbol{\theta})$ il gradiente di $g(\boldsymbol{\theta})$:

$$\nabla g(\boldsymbol{\theta})^T = \left(\frac{\partial}{\partial \theta_1} g(\boldsymbol{\theta}), \ldots, \frac{\partial}{\partial \theta_k} g(\boldsymbol{\theta}) \right).$$

Vale allora il seguente teorema, che estende al caso multiparametrico il metodo delta.

4.103 Teorema Dato il modello statistico $\{\mathcal{X}^n, f_n(\cdot; \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$, assumendo valide le condizioni di regolarità che assicurano la normalità asintotica della successione degli stimatori di massima verosimiglianza di $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^k$, e assumendo che $\nabla g(\widehat{\boldsymbol{\theta}}_n) \neq 0$, si ha che, per $n \to \infty$,

$$\frac{g(\widehat{\boldsymbol{\theta}}_n) - g(\boldsymbol{\theta})}{\sqrt{v_a[g(\widehat{\boldsymbol{\theta}}_n)]}} \xrightarrow{d} \mathrm{N}(0,1),$$

dove la varianza asintotica $v_a[g(\widehat{\boldsymbol{\theta}}_n)]$ è stimata da

$$\widehat{v_a[g(\widehat{\boldsymbol{\theta}}_n)]} = [\nabla g(\widehat{\boldsymbol{\theta}}_n)]^T J_n(\widehat{\boldsymbol{\theta}}_n)[\nabla g(\widehat{\boldsymbol{\theta}}_n)].$$

□

4.104 Esempio (**Coefficiente variazione, modello normale.**) Consideriamo nuovamente il modello $\mathrm{N}(\theta_1, \theta_2)$ e supponiamo di essere interessati a stimare il coefficiente di variazione: $g(\theta_1, \theta_2) = \frac{\sqrt{\theta_2}}{\theta_1}$. Lo stimatore di massima verosimiglianza è allora $g(\widehat{\boldsymbol{\theta}}_n) = \frac{\widehat{\sigma}_n}{\overline{X}_n}$. Con semplici calcoli si mostra che

$$\nabla g(\boldsymbol{\theta})^T = \left(-\frac{\sqrt{\theta}_2}{\theta_1^2}, \frac{1}{2\theta_1\sqrt{\theta}_2} \right)$$

e quindi che

$$\nabla g(\boldsymbol{\theta})^T J_n(\boldsymbol{\theta}) \nabla g(\boldsymbol{\theta}) = \left(-\frac{\theta_2}{\theta_1^2}, \frac{1}{2\theta_1\sqrt{\theta}_2} \right) \begin{pmatrix} \frac{\theta_2}{n} & 0 \\ 0 & \frac{2\theta_2^2}{n} \end{pmatrix} \begin{pmatrix} -\frac{\theta_2}{\theta_1^2} \\ \frac{1}{2\theta_1\sqrt{\theta}_2} \end{pmatrix} = \frac{1}{n} \left(\frac{\theta_2^2}{\theta_1^4} + \frac{\theta_2}{2\theta_1^2} \right)$$

Pertanto,

$$\frac{\widehat{\sigma}_n}{\overline{X}_n} \dot{\sim} \mathrm{N}\left[\frac{\sqrt{\theta}_2}{\theta_1}, \frac{1}{n} \left(\frac{\theta_2^2}{\theta_1^4} + \frac{\theta_2}{2\theta_1^2} \right) \right].$$

In questo caso

$$\widehat{v_a(g(\widehat{\boldsymbol{\theta}}))} = [\nabla g(\widehat{\boldsymbol{\theta}}_n)]^T J_n(\widehat{\boldsymbol{\theta}}_n)[\nabla g(\widehat{\boldsymbol{\theta}}_n)] = \frac{1}{n} \left(\frac{\widehat{\sigma}_n^4}{\overline{X}_n{}^4} + \frac{\widehat{\sigma}_n^2}{2\overline{X}_n{}^2} \right).$$

□

4.9 Appendice

4.9.1 Consistenza dello stimatore di massima verosimiglianza

Diamo qui una dimostrazione abbastanza formale della consistenza dello stimatore di massima verosimiglianza[22]

4.105 Teorema (Consistenza dello stimatore di mv). Dato il modello statistico $\{\mathcal{X}, f_X(\cdot;\theta), \theta \in \Theta\}$ identificabile e una successione $X_1, X_2, \ldots$ di v.a. con distribuzione $f_X(\cdot;\theta)$, sia $L(\theta;\mathbf{x}_n) = \prod_{i=1}^n f_X(x_i;\theta)$ la funzione di verosimiglianza di θ per un campione di dimensione n e $(\widehat{\theta}_n,\ n \in \mathbb{N})$ la successione degli stimatori di massima verosimiglianza di θ. Siano inoltre

$$M_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \log \frac{f_X(X_i;\theta)}{f_X(X_i;\theta^*)} = \frac{1}{n}[\ell_n(\theta) - \ell_n(\theta^*)] \tag{4.25}$$

dove $\ell_n(\theta) = \ell(\theta;\mathbf{X}_n) = \ln L(\theta;\mathbf{X}_n)$ è la funzione di log-verosimiglianza di θ associata al campione aleatorio $\mathbf{X}_n$ e

$$M(\theta) = -D(\theta^*, \theta), \tag{4.26}$$

dove θ^* indica il valore vero del parametro θ. Se

$$\sup_{\theta\in\Theta} |M_n(\theta) - M(\theta)| \xrightarrow{p} 0 \tag{4.27}$$

e se, per ogni $\epsilon > 0$,

$$\sup_{\theta:|\theta-\widehat{\theta}_n|\geq\epsilon} M(\theta) < M(\theta^*), \tag{4.28}$$

allora

$$\widehat{\theta}_n \xrightarrow{p} \theta^*.$$

□

Dimostrazione. Si noti innanzitutto che $\widehat{\theta}_n$ massimizza sia $\ell_n(\theta)$ che $M_n(\theta)$ (poiché $\ell_n(\theta^*)$ è costante rispetto a θ). Per la legge dei grandi numeri,

$$M_n(\theta) \xrightarrow{d} M(\theta) \tag{4.29}$$

dove

$$\begin{aligned} M(\theta) = \mathbb{E}_{\theta^*}\left[\log \frac{f_X(X_i;\theta)}{f_X(X_i;\theta^*)}\right] &= \int \log \frac{f_X(X_i;\theta)}{f_X(X_i;\theta^*)} f(x;\theta^*)dx \\ &= -\int \log \frac{f_X(X_i;\theta^*)}{f_X(X_i;\theta)} f(x;\theta^*)dx \\ &= -D(\theta^*,\theta). \end{aligned}$$

Per n sufficientemente elevato abbiamo quindi che $M_n(\theta) \approx -D(\theta^*,\theta)$ che, essendo $D \geq 0$, assume valore massimo in $\theta = \theta^*$, mentre $-D(\theta^*,\theta) < 0$ per ogni $\theta \neq \theta^*$. Abbiamo quindi stabilito che $M_n(\theta)$, massimizzato da $\widehat{\theta}_n$, converge in probabilità a $-D(\theta^*,\theta)$, massimizzato da θ^*. Pertanto ci aspettiamo (ma va dimostrato) che $\widehat{\theta}_n$ converga in probabilità a θ^*. Poiché $\widehat{\theta}_n$ massimizza $M_n(\theta)$, si ha che

$$M_n(\widehat{\theta}_n) \geq M_n(\theta^*).$$

[22] La trattazione della consistenza dello stimatore di mv è tratta da Wasserman (2004), Cap. 9.

Pertanto

$$\begin{aligned} M(\theta^*) - M(\widehat{\theta}_n) &= M_n(\theta^*) - M(\widehat{\theta}_n) + M(\theta^*) - M_n(\theta^*) \\ &\leq M_n(\widehat{\theta}_n) - M(\widehat{\theta}_n) + M(\theta^*) - M_n(\theta^*) \\ &\leq \sup_\theta |M_n(\theta) - M(\theta)| + M(\theta^*) - M_n(\theta^*) \\ &\xrightarrow{p} 0 \end{aligned}$$

poiché $\sup_\theta |M_n(\theta) - M(\theta)| \xrightarrow{p} 0$ per la (4.27) e $M_n(\theta^*) \xrightarrow{p} M(\theta^*)$ per la (4.29). Poiché, quindi, $M(\widehat{\theta}_n) \xrightarrow{p} M(\theta^*)$, abbiamo che, per ogni $\delta > 0$,

$$\mathbb{P}_{\theta^*}\left(M(\widehat{\theta}_n) < M(\theta^*) - \delta\right) \to 0.$$

Consideriamo ora un arbitrario $\epsilon > 0$, per la (4.28) e le proprietà dell'estremo superiore di un insieme, esiste un $\delta > 0$ tale che

$$|\theta - \theta^*| \geq \epsilon \quad \Rightarrow \quad M(\theta) < M(\theta^*) - \delta.$$

Pertanto

$$\{\mathbf{x}_n : \ |\widehat{\theta}_n - \theta^*| > \epsilon\} \subseteq \{\mathbf{x}_n; \ M(\widehat{\theta}_n) < M(\theta^*) - \delta\}$$

e quindi, per ogni $\epsilon > 0$,

$$\mathbb{P}_{\theta^*}(|\widehat{\theta}_n - \theta^*| > \epsilon) \leq \mathbb{P}_{\theta^*}\left(M(\widehat{\theta}_n) < M(\theta^*) - \delta\right) \to 0$$

che prova la consistenza di $\widehat{\theta}_n$.

4.9.2 Dimostrazione della normalità asintotica dello smv

Dimostriamo qui il Teorema 4.92

Dimostrazione. Sia $\ell_n(\theta) = \ell_n(\theta; \mathbf{x}_n) = \ln L(\theta; \mathbf{x}_n)$. Poiché $\widehat{\theta}_n$ è smv, si ha $\ell'(\widehat{\theta}_n) = 0$. Consideriamo lo sviluppo in serie di Taylor di ℓ' in un intorno di θ e calcolato in $\widehat{\theta}_n$. Si ha

$$0 = \ell'(\widehat{\theta}_n) = \ell'(\theta) + (\widehat{\theta}_n - \theta)\ell''(\theta) + r_n$$

dove r_n è una quantità che tende a zero al crescere di n. Arrestando lo sviluppo al secondo ordine, dalla precedente relazione otteniamo

$$\sqrt{n}(\widehat{\theta}_n - \theta) \simeq \sqrt{n}\frac{\ell'(\theta)}{-\ell''(\theta)} = \frac{\frac{1}{\sqrt{n}}\ell'(\theta)}{-\frac{1}{n}\ell''(\theta)} = \frac{N_n}{D_n}.$$

Possiamo ora mostrare che:

$$N_n = \frac{1}{\sqrt{n}}\ell'(\theta) \xrightarrow{d} W \sim \mathrm{N}[0, I_1(\theta)]$$

e che

$$D_n = -\frac{1}{n}\ell''(\theta) \xrightarrow{p} I_1(\theta),$$

ovvero che, per il teorema di Slutsky

$$\frac{N_n}{D_n} \xrightarrow{d} \frac{W}{I_1(\theta)} \sim \mathrm{N}[0, I_1(\theta)^{-1}]$$

da cui si ottiene la (4.22). Per mostrare la prima delle due precedenti convergenze poniamo $Y_i = \partial \ln f_X(X_i;\theta)/\partial\theta$ e $\overline{Y}_n = \sum_{i=1}^n Y_i/n$. Sappiamo che (vedi Teorema 4.43) $\mathbb{E}_\theta(Y_i) = 0$ e $\mathbb{V}_\theta(Y_i) = I_1(\theta)$. Per il teorema del limite centrale si ha quindi che

$$\frac{1}{\sqrt{n}}\ell'(\theta) = \frac{1}{\sqrt{n}} n\overline{Y}_n = \sqrt{n}(\overline{Y}_n - 0) \xrightarrow{d} W \sim \mathrm{N}[0, I_1(\theta)].$$

Per verificare la convergenza in probabilità di D_n a $I_1(\theta)$, poniamo $A_i = -\partial^2 \ln f_X(X_i;\theta)/\partial\theta^2$ e $\overline{A}_n = \sum_{i=1}^n A_i/n = D_n$. Poiché $\mathbb{E}_\theta(\overline{A}_n) = \mathbb{E}_\theta(A_i) = I_1(\theta)$. Per la legge dei grandi numeri abbiamo che

$$D_n = \overline{A}_n \xrightarrow{p} I_1(\theta).$$

La (4.23) si ottiene in virtú del teorema di Slutsky osservando che

$$I_n(\widehat{\theta}_n)^{1/2}(\widehat{\theta}_n - \theta) = \frac{I_n(\widehat{\theta}_n)^{1/2}}{I_n(\theta)^{1/2}} \times I_n(\theta)^{1/2}(\widehat{\theta}_n - \theta)$$

dove il primo fattore converge in probabilità a 1 e il secondo converge in distribuzione a $\mathrm{N}(0,1)$.

Capitolo 5

Inferenza frequentista: stima mediante insiemi

Dopo aver esaminato il tema della stima puntuale di un parametro nell'ottica dell'inferenza statistica frequentista (Cap. 4), in questo capitolo si considera il problema di stima intervallare. Si introducono alcune definizioni generali (stimatore intervallare, probabilità di copertura, livello di confidenza) e poi, come nel caso della stima puntuale, si esaminano: (a) un metodo per ottenere stimatori intervallari (metodo delle quantità pivotali); (b) un metodo per la valutazione. Successivamente vengono considerati gli intervalli di stima ottenuti con approssimazioni asintotiche. Il capitolo si chiude con l'illustrazione di un metodo per affrontare il fondamentale problema pre-sperimentale della scelta della dimensione campionaria, basato sul controllo probabilistico della lunghezza aleatoria di un intervallo stima.

5.1 Definizioni preliminari

5.1 Definizione (Stima e stimatore intervallare). Dato il modello statistico $\{\mathcal{X}^n, f_n(\mathbf{x}_n;\theta), \theta \in \Theta\}$ in cui $\Theta \subseteq \mathbb{R}^1$ e un campione osservato $\mathbf{x}_n \in \mathcal{X}^n$, una *stima intervallare* di θ è definita attraverso l'insieme

$$C(\mathbf{x}_n) = [L(\mathbf{x}_n), U(\mathbf{x}_n)]$$

dove

a) $L(\cdot)$ e $U(\cdot)$ sono due funzioni definite in $\mathcal{X}^n$ e a valori in Θ, ovvero: $L : \mathcal{X}^n \to \Theta$ e $U : \mathcal{X}^n \to \Theta$;

b) $L(\mathbf{x}_n) \leq U(\mathbf{x}_n)$, $\forall \mathbf{x}_n \in \mathcal{X}^n$;

c) $C(\mathbf{x}_n) \subseteq \Theta$, $\forall \mathbf{x}_n \in \mathcal{X}^n$.

L'oggetto aleatorio

$$C(\mathbf{X}_n) = [L(\mathbf{X}_n), U(\mathbf{X}_n)]$$

si chiama *stimatore intervallare* di θ. □

Si noti che nella precedente definizione abbiamo utilizzato un intervallo chiuso, ma in generale possiamo anche considerare intervalli aperti o semiaperti di $\mathbb{R}^1$.

5.2 Esempio Nel caso del modello $\mathrm{N}(\theta, 1)$, sono stimatori intervallari di θ gli insiemi $C = [\overline{X}_n - 1/\sqrt{n}, \overline{X}_n + 1/\sqrt{n}]$, $C'(\mathbf{X}_n)[\overline{X}_n - 1, \overline{X}_n + 1]$, $C''(\mathbf{X}_n) = [X_1 - 1, X_1 + 1]$. □

Così come per la stima puntuale, si pone il problema di stabilire quali sono le buone proprietà di uno stimatore intervallare e come scegliere tra più alternative. Il problema della valutazione degli stimatori intervallari è trattato nel Paragrafo 5.3. Basandoci semplicemente sull'intuizione possiamo anticipare che una stima intervallare $C(\mathbf{x}_n)$ è buona se: (a) contiene il vero valore del parametro; (b) la sua lunghezza non è eccessiva. Poiché però θ è incognito (altrimenti non si porrebbe il problema di stimarlo!), non possiamo mai stabilire se $\theta \in C(\mathbf{x}_n)$. Possiamo invece calcolare la probabilità che $C(\mathbf{X}_n)$ contenga θ. Inoltre, la lunghezza di $C(\mathbf{x}_n)$ varia al variare di $\mathbf{x}_n \in \mathcal{X}^n$, e quindi il valore di tale lunghezza ci dice poco del comportamento complessivo dell'oggetto aleatorio $C(\mathbf{X}_n)$. In altri termini, dal punto di vista frequentista, come sempre, ciò che si può fare è valutare uno *stimatore* intervallare $C(\mathbf{X}_n)$ e non una *stima* intervallare $C(\mathbf{x}_n)$. In generale possiamo quindi dire che uno stimatore intervallare $C(\mathbf{X}_n)$ è buono se, con elevata probabilità, calcolata prima di osservare i dati, ovvero con riferimento allo spazio dei campioni, si ha che: (a) la sua lunghezza aleatoria non è elevata; (b) contiene il vero valore di θ, qualunque esso sia. Per precisare queste due idee abbiamo bisogno delle seguenti due definizioni.

5.3 Definizione (Lunghezza di uno stimatore intervallare). Dato $C(\mathbf{X}_n) = [L(\mathbf{X}_n), U(\mathbf{X}_n)]$, stimatore intervallare di θ, la variabile aleatoria

$$\mathcal{L}(\mathbf{X}_n) = U(\mathbf{X}_n) - L(\mathbf{X}_n),$$

è la *lunghezza aleatoria* di C; il suo valore atteso

$$e_n = \mathbb{E}_\theta[\mathcal{L}(\mathbf{X}_n)] = \mathbb{E}_\theta[U(\mathbf{X}_n)] - \mathbb{E}_\theta[L(\mathbf{X}_n)]$$

si chiama *lunghezza attesa* di $C(\mathbf{X}_n)$. Il numero $\mathcal{L}(\mathbf{x}_n)$, realizzazione della v.a. $\mathcal{L}(\mathbf{X}_n)$, rappresenta invece la lunghezza dell'intervallo $C(\mathbf{x}_n)$, corrispondente a un determinato intervallo $\mathbf{x}_n \in \mathcal{X}^n$. □

5.4 Definizione (Probabilità di copertura). La *probabilità di copertura* di $C(\mathbf{X}_n)$, stimatore intervallare di un parametro θ, è la probabilità che l'intervallo aleatorio $[L(\mathbf{X}_n), U(\mathbf{X}_n)]$ contenga il vero valore di θ:

$$\mathbb{C}\mathrm{op}_\theta(C) = \mathbb{P}_\theta[\theta \in C(\mathbf{X}_n)] = \mathbb{P}_\theta[L(\mathbf{X}_n) \leq \theta \leq U(\mathbf{X}_n)]. \tag{5.1}$$

□

5.5 Osservazione

1. Nella definizione (5.1), il parametro θ è una quantità fissa incognita. La probabilità $\mathbb{P}_\theta$ si riferisce alle v.a. aleatorie $L(\mathbf{X}_n)$ e $U(\mathbf{X}_n)$, ovvero all'oggetto aleatorio $C(\mathbf{X}_n)$.

2. In generale, la probabilità di copertura di un intervallo, $\mathbb{C}\mathrm{op}_\theta(C)$, è una funzione non costante di θ. A seconda di qual è il vero valore di θ, che non conosciamo, la garanzia probabilistica di un intervallo può infatti variare. In alcuni casi particolari e fortunati, la probabilità di copertura è invece costante al variare dei possibili valori di θ.

□

5.6 Definizione (Coefficiente di confidenza). Il *coefficiente (o livello) di confidenza* di $C(\mathbf{X}_n)$, stimatore intervallare di un parametro θ, è il minimo valore della probabilità che

l'intervallo aleatorio $[L(\mathbf{X}_n), U(\mathbf{X}_n)]$ contenga il vero valore di θ, ottenuto al variare di θ in Θ:

$$\inf_{\theta\in\Theta} \mathbb{C}\text{op}_\theta(C) = \inf_{\theta\in\Theta} \mathbb{P}_\theta[\theta \in C(\mathbf{X}_n)] = \inf_{\theta\in\Theta} \mathbb{P}_\theta[L(\mathbf{X}_n) \leq \theta \leq U(\mathbf{X}_n)]$$

□

Il coefficiente di confidenza di $C(\mathbf{X}_n)$ è quindi la garanzia minima che lo stimatore intervallare dà di contenere θ.

5.7 Esempio (Modello normale). Nel caso del modello $\mathrm{N}(\theta, 1)$, lo stimatore $C(\mathbf{X}_n) = [\overline{X}_n - 1/\sqrt{n}, \overline{X}_n + 1/\sqrt{n}]$ ha probabilità di copertura che non varia con θ:

$$\begin{aligned}
\mathbb{C}\text{op}_\theta(C) = \mathbb{P}_\theta[\theta \in C(\mathbf{X}_n)] &= \mathbb{P}_\theta(\overline{X}_n - 1/\sqrt{n} \leq \theta \leq \overline{X}_n + 1/\sqrt{n}) \\
&= \mathbb{P}_\theta(-1/\sqrt{n} \leq \overline{X}_n - \theta \leq 1/\sqrt{n}) \\
&= \mathbb{P}_\theta\left(-1 \leq \sqrt{n}(\overline{X}_n - \theta) \leq 1\right) \\
&= \mathbb{P}_\theta(-1 \leq Z \leq 1) \\
&= \Phi(1) - \Phi(-1) \\
&= 0.68.
\end{aligned}$$

In questo caso, il valore 0.68 è quindi sia probabilità di copertura che livello di confidenza. □

5.8 Esempio[1] (Modello uniforme). Consideriamo per il modello uniforme in $[0, \theta]$ lo stimatore intervallare $C(\mathbf{X}_n) = [Y + a, Y + b]$, dove $Y = X_{(n)}$ e a, b sono costanti positive tali che $a < b$. Ricordando che la distribuzione campionaria di Y è $f_Y(y; \theta) = ny^{n-1}/\theta^n$, $y \in [0, \theta]$, possiamo calcolare facilmente la probabilità di copertura di $C(\mathbf{X}_n)$. Infatti

$$\begin{aligned}
\mathbb{C}\text{op}_\theta(C) = \mathbb{P}_\theta[\theta \in C(\mathbf{X}_n)] &= \mathbb{P}_\theta(Y + a \leq \theta \leq Y + b) \\
&= \mathbb{P}_\theta(\theta - b \leq Y \leq \theta - a) \\
&= \theta^{-n} \int_{\theta-b}^{\theta-a} n y^{n-1} dy \\
&= \theta^{-n}[(\theta - a)^n - (\theta - b)^n] \\
&= \left(1 - \frac{a}{\theta}\right)^n - \left(1 - \frac{b}{\theta}\right)^n.
\end{aligned}$$

La probabilità di copertura di C, in questo caso, dipende da θ. Poichè

$$\lim_{\theta\to\infty} \left[\left(1 - \frac{a}{\theta}\right)^n - \left(1 - \frac{b}{\theta}\right)^n\right] = 0$$

possiamo concludere che il coefficiente di confidenza di C è uguale a zero. □

5.9 Definizione (Intervallo di confidenza). Dato un numero reale $\alpha \in (0, 1)$, uno stimatore intervallare $C(\mathbf{X}_n)$ di un parametro θ, con coefficiente di confidenza pari a $1 - \alpha$, ovvero tale che

$$\inf_{\theta\in\Theta} \mathbb{P}_\theta[\theta \in C(\mathbf{X}_n)] = \inf_{\theta\in\Theta} \mathbb{P}_\theta[L(\mathbf{X}_n) \leq \theta \leq U(\mathbf{X}_n)] = 1 - \alpha$$

si chiama *intervallo di confidenza di livello* $1 - \alpha$ e si indica $C_{1-\alpha}(\mathbf{X}_n)$. □

Così come fatto per la teoria della stima puntuale nel precedente capitolo, possiamo distinguere tra:

a) metodi per la stima intervallare (prg. 5.2);

b) valutazione degli stimatori intervallari (prg. 5.3).

[1] Esempio tratto da Casella-Berger (2002), p. 419-420.

5.2 Metodi per la stima intervallare

Così come esistono molteplici modi per costruire uno stimatore puntuale di un parametro (ad esempio: stimatori dei momenti, di massima verosimiglianza, UMVUE...), allo stesso modo non esiste un unico metodo per ottenere stimatori intervallari. I metodi più noti sono due:

a) metodo delle quantità pivotali (prg. 5.2.1);

b) metodo dell'inversione della regione di accettazione di un test statistico.

In questo capitolo ci soffermiamo sul primo dei due metodi.

5.2.1 Metodo delle quantità pivotali

Il metodo consente di costruire un intervallo di confidenza di livello prefissato $1-\alpha$ per il parametro di un modello, quando per tale modello si è in grado di individuare una *quantità pivotale*, qui di seguito definita. La caratteristica del metodo è quella di produrre intervalli con probabilità di copertura costante, coincidente quindi con il livello di confidenza desiderato.

5.10 Definizione (Quantità pivotale). Dato il modello statistico $\{\mathcal{X}^n, f_n(\mathbf{x}_n;\theta), \theta \in \Theta\}$, si definisce *quantità pivotale* l'oggetto aleatorio

$$Q:\ \mathcal{X}^n \times \Theta \to \mathbb{R}$$

con le seguenti proprietà:

a) $Q(\mathbf{X}_n, \theta)$ ha distribuzione campionaria $f_Q(\cdot)$ che non dipende da θ;

b) $Q(\mathbf{x}_n, \theta)$ è una funzione monotona di θ per ogni campione $\mathbf{x}_n$ di $\mathcal{X}^n$.

□

Una quantità pivotale è quindi una funzione di dati e parametro, la cui distribuzione di probabilità non dipende dal parametro. Grazie alla proprietà (a), data una quantità pivotale Q, è in linea di principio semplice individuare due valori q_1 e q_2, indipendenti da θ, tali che, per un qualsiasi valore $\alpha \in (0,1)$ si abbia (esattamente nel caso di v.a. $\mathbf{X}_n$ assolutamente contiunue, approssimativamente nel caso discreto)

$$\mathbb{P}(q_1 \leq Q(\mathbf{X}_n, \theta) \leq q_2) = 1-\alpha, \qquad \forall \theta \in \Theta. \tag{5.2}$$

L'intervallo $[q_1, q_2]$ si chiama *intervallo pivotale*. La proprietà (b) garantisce l'invertibilità di Q rispetto a θ e quindi la possibilità di determinare due funzioni $L(\mathbf{X}_n)$ ed $U(\mathbf{X}_n)$, che dipendono da Q^{-1}, $\mathbf{X}_n$, q_1 e q_2 tali che, comunque si fissa un numero $\alpha \in (0,1)$, si abbia che

$$\mathbb{P}(q_1 \leq Q(\mathbf{X}_n, \theta) \leq q_2) = \mathbb{P}[L(\mathbf{X}_n) \leq \theta \leq U(\mathbf{X}_n)] = \mathbb{P}[\theta \in C(\mathbf{X}_n)] = 1-\alpha, \qquad \forall \theta \in \Theta.$$

La regione aleatoria $C(\mathbf{X}_n) = [L(\mathbf{X}_n), U(\mathbf{X}_n)]$ è quindi, per definizione, un insieme di confidenza di livello $1-\alpha$.

5.11 Esempio. Sia X_1 una v.a. con distribuzione $\mathrm{N}(\theta, \sigma^2)$. In questo caso, per ogni $\theta \in \mathbb{R}$, la funzione

$$Q(X_1, \theta) = Z = \frac{X_1 - \theta}{\sigma} \sim \mathrm{N}(0,1),$$

ha cioè distribuzione $\mathrm{N}(0,1)$ che non dipende dal parametri incognito ed è inoltre una funzione invertibile di θ. Anche se di scarsa utilità, in quanto basata su una singola osservazione, $Q(X_1, \theta)$ è quindi una quantità pivotale per θ. Esistono infatti infinite coppie di valori reali (q_1, q_2) che garantiscono che

$$\mathbb{P}(q_1 \leq Z \leq q_2) = \int_{q_1}^{q_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = 1 - \alpha$$

ovvero che, $\forall \theta \in \mathbb{R}$, si abbia

$$\begin{aligned} 1 - \alpha &= \mathbb{P}\left(q_1 \leq \frac{X_1 - \theta}{\sigma} \leq q_2\right) \\ &= \mathbb{P}(-X_1 + q_1\sigma \leq -\theta \leq -X_1 + q_2\sigma) \\ &= \mathbb{P}(X_1 - q_2\sigma \leq \theta \leq X_1 - q_1\sigma). \end{aligned}$$

In questo caso possiamo porre $L(X_1) = X_1 - q_2\sigma$ e $U(X_1) = X_1 - q_1\sigma$ ed ottenere un intervallo aleatorio

$$C_{1-\alpha}(X_1) = [X_1 - q_2\sigma, X_1 - q_1\sigma]$$

che, qualunque sia il vero valore di θ, con probabilità $1 - \alpha$ contiene tale valore.
Si noti che se consideriamo $Q(X_1, \cdot)$ come funzione di θ e se poniamo

$$Q(X_1, \theta) = \frac{X_1 - \theta}{\sigma} = q,$$

allora la funzione inversa di $Q(X_1, \cdot)$ risulta definita da

$$Q^{-1}(X_1, q) = X_1 - q\sigma$$

e quindi

$$C_{1-\alpha}(X_1) = [Q^{-1}(X_1, q_2), Q^{-1}(X_1, q_1)].$$

Si noti inoltre che, ponendo $q_1 = -z_{1-\frac{\alpha}{2}}$ e $q_2 = z_{1-\frac{\alpha}{2}}$, l'intervallo diventa

$$C_{1-\alpha}(X_1) = [X_1 - z_{1-\frac{\alpha}{2}}\sigma, X_1 + z_{1-\frac{\alpha}{2}}\sigma].$$

□

5.12 Osservazione.

1. Per la stima intervallare del parametro di un determinato modello statistico esistono, in genere, più possibili quantità pivotali. In pratica, come vedremo negli esempi, si parte in genere da stimatori puntuali (o da statistica ad essi collegate) di θ e si passa quindi a una sua trasformata che soddisfi le proprietà richieste da una quantià pivotale (vedi Def. 5.10). Naturalmente, se possibile, le quantità pivotali vengono elaborate sulla base di stimatori/statistiche sufficienti per il modello.

2. Scelta una quantità pivotale e fissato il valore $1 - \alpha$, esistono molteplici coppie di valori q_1 e q_2 in (5.2) che definiscono un intervallo pivotale con probabilità pari a $1 - \alpha$ (le scelte sono infinite nel caso di quantità pivotali dotate di funzione di densità di probabilità).

3. Il metodo può essere usato solo se il risultante insieme $C(\mathbf{X}_n)$ non dipende dal parametro incognito θ (se l'intervallo dipendesse da θ non sarebbe uno stimatore).

4. La definizione della quantità pivotale Q è, in un certo senso, speculare a quella di stimatore puntuale), $(\mathbf{X}_n)$. Infatti, mentre $d(\mathbf{X}_n)$ deve essere funzione di $\mathbf{X}_n$ ma non di θ e la sua distribuzione campionaria dipenda da θ, per $Q(\mathbf{X}_n, \theta)$ vale esattamente l'opposto: deve essere anche funzione di θ ma la sua legge di probabilità, $f_Q(\cdot)$ non può dipendere da θ.

□

Per quanto riguarda la prima osservazione, senza entrare in dettagli formali, è ragionevole scegliere, tra possibili quantità pivotali a disposizione, una costruita partendo da statistiche sufficienti per il modello in esame, ovvero da stimatori del parametro del modello con buone proprietà frequentiste. Una scelta comune consiste nel definire $Q(\mathbf{X}_n, \theta)$ a partire da stimatori di massima verosimiglianza o da stimatori UMVUE. Il paragrafo che segue è dedicato alla scelta dei quantili q_1 e q_2 che definiscono l'intervallo pivotale, da cui discende l'intervallo di confidenza,

5.2.2 Tipologie di insiemi pivotali

Consideriamo il caso di quantità pivotali che siano funzioni di variabili aleatorie assolutamente continue, dotate quindi di funzione di densità di probabilità. In questo caso, come si è detto, esistono infinite scelte di q_1 e q_2 che garantiscono che l'intervallo pivotale $[q_1, q_2]$ abbia probabilità $1 - \alpha$. Esaminiamo ora le due principali procedure di scelta dei suddetti valori, che defiscono, rispettivamente (a) intervalli pivolati a code uguali; e (b) intervalli pivotali di densità massima.

Intervalli pivotali a code uguali (ET)

Tali intervalli (in inglese *Equal tails intervals*, da cui la sigla intervalli ET) si ottengono ponendo in (5.2):

$$q_1 = q_{\alpha/2} \qquad \text{e} \qquad q_2 = q_{1-\frac{\alpha}{2}},$$

dove q_ϵ indica il quantile di livello ϵ di $Q(\mathbf{X}_n, \theta)$. Con questa scelta, per costruzione, l'intervallo $[q_{\alpha/2}, q_{1-\frac{\alpha}{2}}]$ ha probabilità pari a $1 - \alpha$ qualunque sia il vero valore di θ e lascia alla propria destra e alla propria sinistra intervalli o semirette equiprobabili di valori che può assumere l'oggetto aleatorio $Q(\mathbf{X}_n, \theta)$. Il metodo appena descritto, che è quello usato più frequentemente nella pratica e negli esempi che seguono, ha il vantaggio delle semplicità. Infatti, in molti importanti, la distribuzione di $Q(\mathbf{X}_n, \theta)$ è nota ed è quindi agevole determinare i valori numerici dei quantili necessari. Si osservi però che l'intervallo pivotale ET non è, in genere, quello di lunghezza minima tra gli intervalli pivotali per un determinato parametro.
La Figura 5.1 riporta il grafico delle funzioni di due quantità pivotali e i corrispondenti insiemi a code uguali (ET). Nel primo caso (più comune) la funzione $f_Q(y)$ è asimmetrica rispetto al suo punto di massimo interno al dominio. Nel secondo caso (per il quale si veda l'Esempio 5.18), la funzione $f_Q(y)$ è crescente nel suo dominio.

Intervalli pivotali di massima densità (HDI)

Questi intervalli (in inglese *Highest density intervals*, da cui l'acronimo intervalli HDI) si ottengono ponendo in (5.2):

$$q_1 = q_1^* \qquad \text{e} \qquad q_2 = q_2^*,$$

e ottenendo l'intervallo $H_{1-\alpha}^* = [q_1^*, q_2^*]$ tale che:

a) $\int_{q_1^*}^{q_2^*} f_Q(y)dy = 1 - \alpha$;

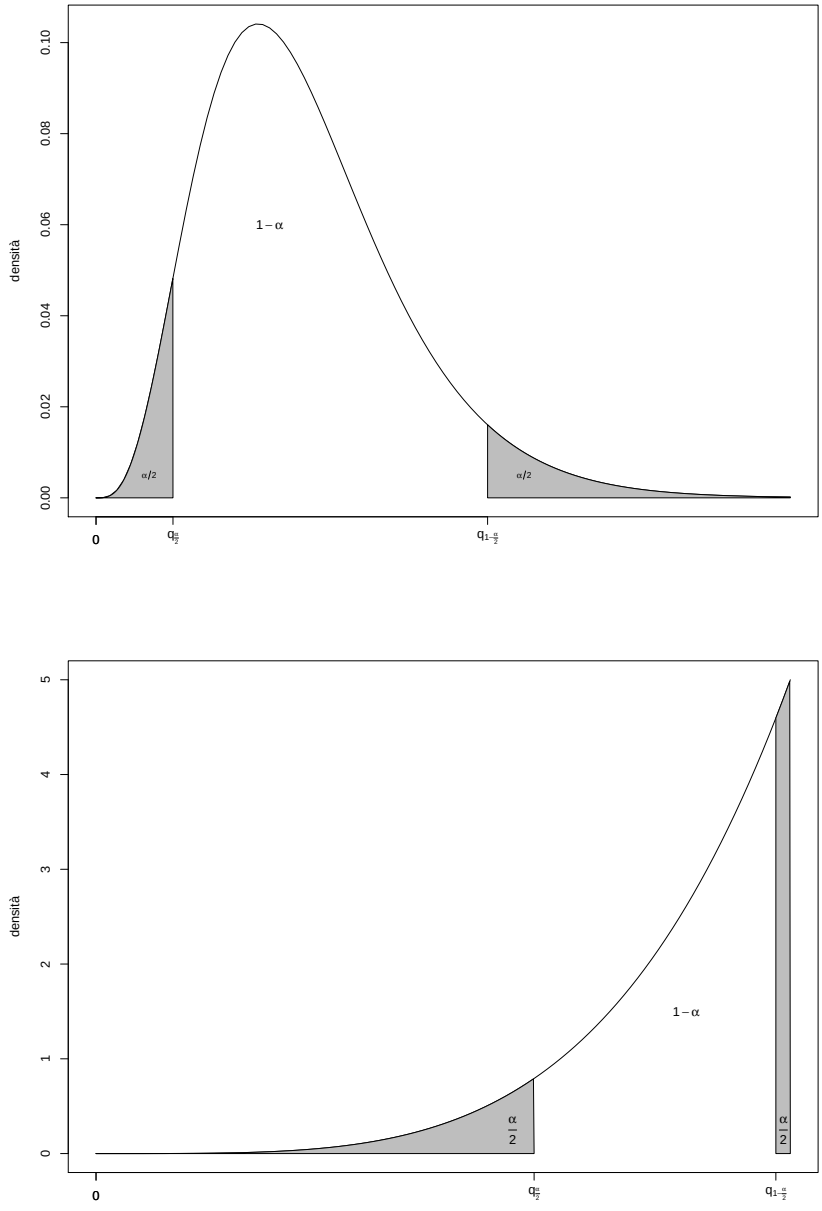

Figura 5.1: *Grafico di funzioni di densità di quantità pivotali e corrispondenti insiemi a code uguali (ET).*

b) $\forall y \in H^*_{1-\alpha}$ e $\forall y' \notin H^*_{1-\alpha}$, si ha che $f_Q(y) \geq f_Q(y')$.

Dalla condizione (b) discende che un insieme HDI di livello $1-\alpha$ è un *insieme di livello* della funzione $f_Q(y)$: esiste cioè un valore $h_{1-\alpha}$

$$H^*_{1-\alpha} = \{y \in \mathbb{R} : \ f_Q(y) \geq h_{1-\alpha}\}.$$

La determinazione numerica dei valori q_1^* e q_2^* risulta in genere più onerosa di quella necessaria per determinare i valori $q_{\alpha/2}$ e $q_{1-\frac{\alpha}{2}}$. Come vedremo nel Paragrafo 5.3, gli insiemi pivotali HDI sono, in molti casi rilevanti, quelli di lunghezza minima tra gli intervalli pivotali di livello $1-\alpha$ fissato.
La Figura 5.2 riporta il grafico delle funzioni delle stesse quantità pivotali della Figura 5.1 e i corrispondenti insiemi di massima densità (HDI). Si osservi che nel primo caso le probabilità corrispondenti alle code di f_Q non sono uguali. Nel secondo caso, l'estremo inferiore dell'intervallo di massima densità finale coincide con il percentile q_α della distribuzione di $Q(\mathbf{X}_n, \theta)$ (si veda l'esempio 5.18).

5.13 Osservazione. Dagli intervalli pivotali si passa agli intervalli di confidenza. È possibile dimostrare che, se si scelgono q_1 e q_2 in modo che l'intervallo pivotale sia ET, il risultante intervallo di confidenza C per θ è esso stesso un intervallo ET. Al contrario, l'intervallo C che si ottiene utilizzando i quantili $q_1^\star$ e $q_2^\star$ dell'insieme pivotale HDI è quello di lunghezza minima solo se $Q(\cdot, \theta)$ è funzione monotona *crescente* di θ. □

5.2.3 Intervalli di confidenza per i parametri del modello normale

In questo paragrafo deriviamo gli intervalli di confidenza per i parametri del modello normale, partendo, in tutti i casi, da insiemi pivotali a code uguali.

5.14 Esempio (Intervalli di confidenza per valore atteso modello normale). Consideriamo il modello $\mathrm{N}(\theta, \sigma^2)$

1. **Varianza nota**.

   In questo caso, come noto, la funzione

   $$Q(\mathbf{X}_n, \theta) = \frac{\sqrt{n}(\overline{X}_n - \theta)}{\sigma}$$

   ha distribuzione $\mathrm{N}(0,1)$ per ogni $\theta \in \mathbb{R}$ ed è inoltre una funzione invertibile di θ. Preso il percentile $z_{1-\frac{\alpha}{2}}$ della v.a. $\mathrm{N}(0,1)$, possiamo affermare che, $\forall \theta \in \mathbb{R}$,

   $$\begin{aligned} 1-\alpha &= \mathbb{P}\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\sqrt{n}(\overline{X}_n - \theta)}{\sigma} \leq z_{1-\frac{\alpha}{2}}\right) \\ &= \mathbb{P}\left(-z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \overline{X}_n - \theta \leq z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) \\ &= \mathbb{P}\left(\overline{X}_n - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \theta \leq \overline{X}_n + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) \end{aligned}$$

   da cui si evince che

   $$C_{1-\alpha}(\mathbf{X}_n) = \left[\overline{X}_n - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{X}_n + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$$

   è un intervallo di confidenza di livello $1-\alpha$ per θ.

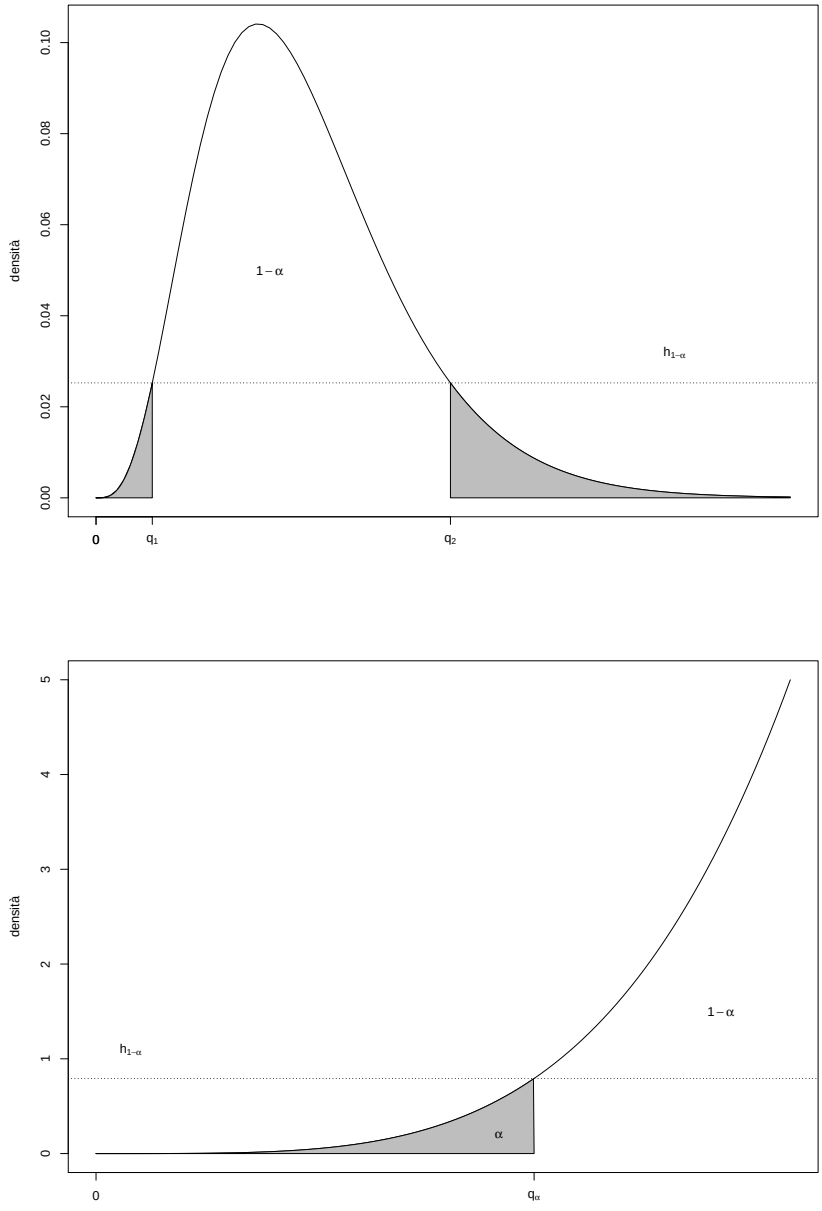

Figura 5.2: *Grafico di funzioni di densità di quantità pivotali e corrispondenti insiemi di massima densità (HDI).*

Nel caso di varianza incognita il precedente intervallo non può essere utilizzato, poichè dipende da σ (e non è quindi uno stimatore).

In questo caso (particolare) la lunghezza di $C_{1-\alpha}$ non è una v. aleatoria e quindi

$$e_n = \mathbb{E}_\theta[\mathcal{L}(\mathbf{X}_n)] = \mathcal{L}(\mathbf{X}_n) = 2z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}.$$

5.15 Osservazione

(a) Per valori fissati di α e σ, e_n tende a zero al crescere di n.

(b) Per la simmetria della funzione di densità di Q rispetto al suo punto di massimo in zero, la scelta dei percentili $\pm z_{1-\frac{\alpha}{2}}$ conduce all'intervallo pivotale che è sia di tipo a code uguali che di massima densità. Tale scelta non è obbligata. Ogni altra coppia di percentili z_1, z_2 della v.a. $\mathrm{N}(0,1)$ tali che $\mathbb{P}[z_1 \leq Z \leq z_2] = 1-\alpha$ porterebbe a un distinto intervallo di confidenza $C'_{1-\alpha}(\mathbf{X}_n) = [\overline{X}_n - z_2\frac{\sigma}{\sqrt{n}}, \overline{X}_n - z_1\frac{\sigma}{\sqrt{n}}]$ di medesimo livello di quello trovato. Tuttavia, come vedremo, la scelta fatta produce l'intervallo di confidenza di lunghezza minima tra tutti quelli di livello $1-\alpha$.

(c) Se poniamo $z_1 = z_\alpha$ e $z_2 = \infty$ si ottiene l'insieme (illimitato)

$$\left(-\infty, \overline{X}_n - z_\alpha\frac{\sigma}{\sqrt{n}}\right]$$

detto insieme di confidenza *unilaterale* (o *one-sided*). Analogamente, se si pone $z_1 = -\infty$ e $z_2 = z_{1-\alpha}$, si ottiene l'altro insieme unilaterale di confidenza di livello $1-\alpha$

$$\left[\overline{X}_n - z_{1-\alpha}\frac{\sigma}{\sqrt{n}}, \infty\right).$$

□

La Figura 5.3 riporta il grafico della funzione di densità della quantità pivotale $Q(\mathbf{X}_n, \theta) = \frac{\sqrt{n}(\overline{X}_n - \theta)}{\sigma} \sim \mathrm{N}(0,1)$ per intervallo di confidenza per θ nel modello $\mathrm{N}(\theta, \sigma^2)$ con $1-\alpha = 0.90$, $\sigma^2 = 1$, $n = 10$. Sull'asse delle ascisse sono indicati anche i percentili di livello $\alpha/2$ e $1-\alpha/2$ della v.a. $\mathrm{N}(0,1)$, che sono simmetrici rispetto all'origine degli assi e che consentono di individuare l'intervallo di lunghezza minima tra gli infiniti intervalli di confidenza di livello $1-\alpha$ determinabili usando la medesima quantità pivotale.

2. **Varianza incognita**.

In questo caso non possiamo utilizzare la quantità pivotale dell'esempio precedente, dal momento che l'intervallo risultante dipende dal parametro incognito di disturbo σ^2. Consideriamo allora la funzione

$$Q(\mathbf{X}_n, \theta) = \frac{\sqrt{n}(\overline{X}_n - \theta)}{S_n},$$

dove $S_n^2 = \sum_{i=1}^n (X_i - \overline{X}_n)^2/(n-1)$, ha distribuzione t di Student con $n-1$ gradi di libertà ed è una funzione invertibile di θ. Si tratta quindi di una quantità pivotale. Utilizzando il percentile $t_{n-1;1-\frac{\alpha}{2}}$ della t_{n-1}, si ha che, $\forall\theta \in \mathbb{R}$,

$$\begin{aligned}
1-\alpha &= \mathbb{P}\left(-t_{n-1;1-\frac{\alpha}{2}} \leq \frac{\sqrt{n}(\overline{X}_n - \theta)}{S_n} \leq t_{n-1;1-\frac{\alpha}{2}}\right) \\
&= \mathbb{P}\left(-t_{n-1;1-\frac{\alpha}{2}}\frac{S_n}{\sqrt{n}} \leq \overline{X}_n - \theta \leq t_{n-1;1-\frac{\alpha}{2}}\frac{S_n}{\sqrt{n}}\right) \\
&= \mathbb{P}\left(\overline{X}_n - t_{n-1;1-\frac{\alpha}{2}}\frac{S_n}{\sqrt{n}} \leq \theta \leq \overline{X}_n + t_{n-1;1-\frac{\alpha}{2}}\frac{S_n}{\sqrt{n}}\right)
\end{aligned}$$

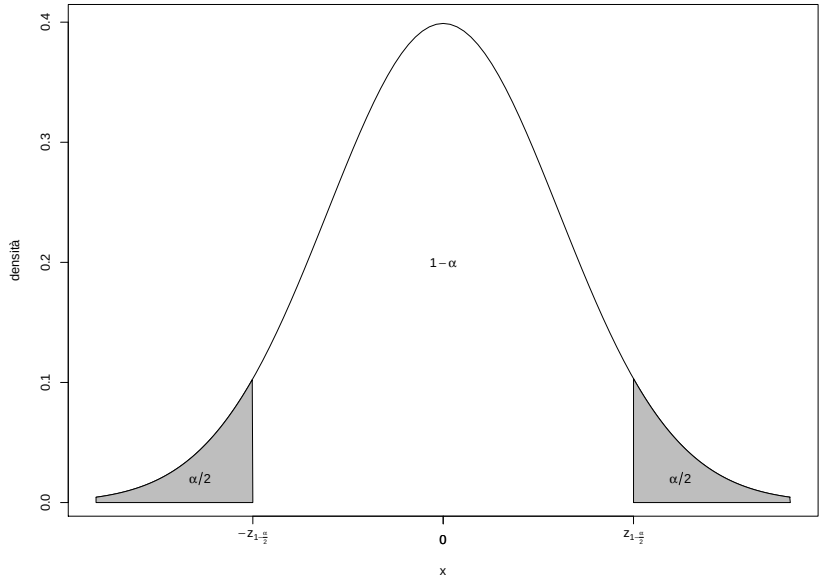

Figura 5.3: *Grafico della funzione di densità della quantità pivotale* $Q(\mathbf{X}_n, \theta) = \frac{\sqrt{n}(\overline{X}_n - \theta)}{\sigma} \sim N(0,1)$ *per intervallo di confidenza per* θ *nel modello* $N(\theta, \sigma^2)$ *con* $1-\alpha = 0.90$, $\sigma^2 = 1$.

L'insieme

$$C_{1-\alpha}(\mathbf{X}_n) = \left[\overline{X}_n - t_{n-1;1-\frac{\alpha}{2}} \frac{S_n}{\sqrt{n}}, \overline{X}_n + t_{n-1;1-\frac{\alpha}{2}} \frac{S_n}{\sqrt{n}}\right]$$

è quindi un intervallo di confidenza per θ di livello $1-\alpha$. Anche in questo caso, la simmetria di f_Q rispetto al suo massimo fa sì che l'intervallo pivotale da cui si ottiene $C_{1-\alpha}(\mathbf{X}_n)$ sia sia di tipo ET che HDI.

La lunghezza di $C_{1-\alpha}$ è

$$\mathcal{L}(\mathbf{X}_n) = 2t_{n-1;1-\frac{\alpha}{2}} \frac{S_n}{\sqrt{n}}$$

e

$$e_n = \mathbb{E}_\theta[\mathcal{L}(\mathbf{X}_n)] = 2t_{n-1;1-\frac{\alpha}{2}} \frac{\mathbb{E}_\theta[S_n]}{\sqrt{n}} = t_{n-1;1-\frac{\alpha}{2}} \kappa(n,\alpha)\sigma,$$

con

$$\kappa(n,\alpha) = \frac{\sqrt{2}}{\sqrt{n-1}} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}.$$

Si può verificare che, per un valore fissato di α, e_n tende a zero al crescere di n. Inoltre, come nel caso precedente, la scelta dei percentili $\pm t_{n-1;1-\frac{\alpha}{2}}$ tra le infinite coppie di percentili della v.a. t_{n-1} che garantiscono un intervallo di confidenza di livello $1-\alpha$ è quella che porta all'intervallo di confidenza di lunghezza minima (vedi teorema 5.22).

La Figura 5.4 riporta il grafico della funzione di densità della quantità pivotale $Q(\mathbf{X}_n, \theta) = \frac{\sqrt{n}(\overline{X}_n - \theta)}{S_n} \sim \mathrm{t}_{n-1}$ per intervallo di confidenza per θ_1 nel modello $\mathrm{N}(\theta_1, \theta_2)$ con $1-\alpha = 0.90$, θ_2 incognita, $n = 10$. Sull'asse delle ascisse sono indicati anche i percentili di livello $\alpha/2$ e $1-\alpha/2$ della v.a. T_{n-1}, che sono simmetrici rispetto all'origine degli assi e che consentono di

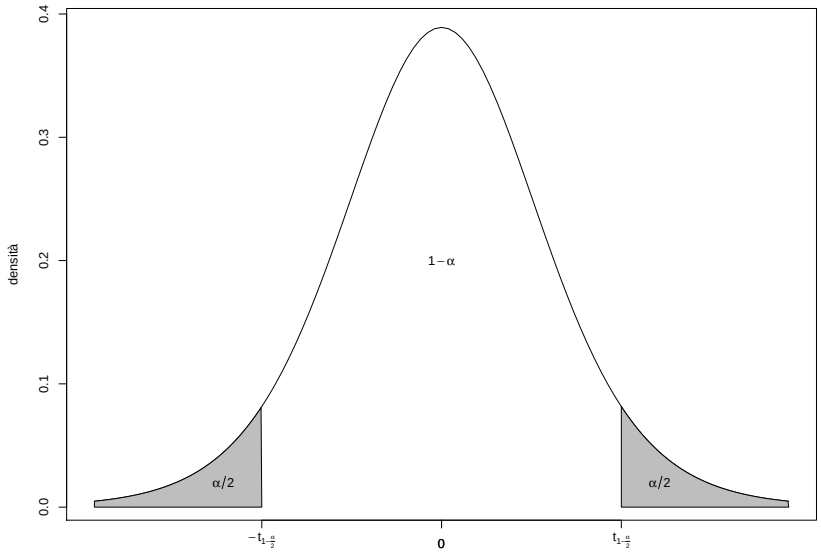

Figura 5.4: *Grafico della funzione di densità della quantità pivotale* $Q(\mathbf{X}_n, \theta) = \frac{\sqrt{n}(\overline{X}_n - \theta_1)}{S_n} \sim t_{n-1}$ *per intervallo di confidenza per* θ_1 *(valore atteso) nel modello* $N(\theta_1, \theta_2)$ *con* $1-\alpha = 0.90$, $n = 10$.

individuare l'intervallo di lunghezza minima tra gli infiniti intervalli di confidenza di livello $1 - \alpha$ determinabili usando la medesima quantità pivotale.

□

5.16 Esempio (Intervalli di confidenza per la varianza del modello normale). Determiniamo ora quantità pivotali e intervalli di confidenza per la varianza del modello normale. In questo caso, a differenza dei due precedenti, le funzioni di densità delle quantità pivotali utilizzate sia nel caso di valore atteso noto che incognito non sono simmetriche rispetto al punto di massimo e quindi insiemi a code uguali e insiemi di massima densità non coincidono. Per gli insiemi ET sono disponibili espressioni esplicite (fornite negli esempi che seguono) che non sono invece disponibili per gli insiemi di densità massima.

1. **Valore atteso noto**

   Consideriamo il modello $\mathrm{N}(\mu_0, \theta)$, in cui il valore atteso è noto e la varianza θ incognita. La funzione
   $$Q(\mathbf{X}_n, \theta) = \frac{nS_0^2}{\theta} = \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\theta}$$
   ha distribuzione χ^2_n (indipendente da θ) ed è funzione invertibile di θ. Si tratta quindi di una quantità pivotale per θ. Siano a e b due valori tali che $\mathbb{P}(a \leq \chi^2_n \leq b) = 1 - \alpha$, si ha che, $\forall \theta \in \mathbb{R}^+$,
   $$1 - \alpha = \mathbb{P}\left(a \leq \frac{nS_0^2}{\theta} \leq b\right) \tag{5.3}$$
   $$= \mathbb{P}\left(\frac{nS_0^2}{b} \leq \theta \leq \frac{nS_0^2}{a}\right) \tag{5.4}$$

Esistono infinite scelte per a e b. La più semplice e utilizzata è considerare i percentili di livello $\alpha/2$ e $1-\alpha/2$ della v.a. χ^2_n:

$$a = \chi^2_{n;\frac{\alpha}{2}}, \qquad b = \chi^2_{n;1-\frac{\alpha}{2}}.$$

Si ottiene così l'intervallo

$$C_{1-\alpha}(\mathbf{X}_n) = \left[\frac{nS_0^2}{\chi^2_{n;1-\frac{\alpha}{2}}}, \frac{nS_0^2}{\chi^2_{n;\frac{\alpha}{2}}}\right].$$

La lunghezza di $C_{1-\alpha}(\mathbf{X}_n)$ è

$$\mathcal{L}(\mathbf{X}_n) = \kappa_0(n,\alpha) n S_0^2, \qquad \text{con} \qquad \kappa(n,\alpha) = \frac{1}{\chi^2_{n;\frac{\alpha}{2}}} - \frac{1}{\chi^2_{n;1-\frac{\alpha}{2}}}.$$

Pertanto

$$e_n = \mathbb{E}_\theta[\mathcal{L}(\mathbf{X}_n)] = n\kappa_0(n,\alpha)\theta$$

Si noti che l'intervallo considerato non è quello di lunghezza minima tra quelli di coefficiente di confidenza $1-\alpha$.

La Figura 5.5 riporta il grafico della funzione di densità della quantità pivotale $Q(\mathbf{X}_n, \theta_2) = \frac{(n-1)S_n^2}{\theta_2} \sim \chi^2_{n-1}$ per intervallo di confidenza per θ_2 (varianza) nel modello $\mathrm{N}(\theta_1, \theta_2)$ con $1-\alpha = 0.90$, $n = 10$. Sull'asse delle ascisse sono indicati anche i percentili di livello $\alpha/2$ e $1-\alpha/2$ della v.a. χ^2_{n-1}. Tale scelta consente di derminare l'intervallo di confidenza "a code uguali", che non risulta essere quello di lunghezza minima tra gli infiniti intervalli di confidenza di livello $1-\alpha$ determinabili usando la medesima quantità pivotale.

2. **Valore atteso incognito**

Consideriamo ora il modello $\mathrm{N}(\theta_1, \theta_2)$, in cui il valore atteso e la varianza sono entrambi incogniti. Vogliamo determinare un intervallo di confidenza per θ_2. La funzione

$$Q(\mathbf{X}_n, \theta_2) = \frac{(n-1)S_n^2}{\theta_2} = \frac{\sum_{i=1}^n (X_i - \overline{X}_n)^2}{\theta_2}$$

ha distribuzione χ^2_{n-1} (indipendente da θ_2) ed è funzione invertibile di θ_2. Si tratta quindi una quantità pivotale per θ_2. In modo del tutto analogo al caso precedente si ottiene quindi il seguente intervallo di confidenza

$$C_{1-\alpha}(\mathbf{X}_n) = \left[\frac{(n-1)S_n^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}}, \frac{(n-1)S_n^2}{\chi^2_{n-1;\frac{\alpha}{2}}}\right].$$

La lunghezza di $C_{1-\alpha}(\mathbf{X}_n)$ è

$$\mathcal{L}(\mathbf{X}_n) = (n-1)\kappa(n,\alpha) S_n^2, \qquad \text{con} \qquad \kappa(n,\alpha) = \frac{1}{\chi^2_{n-1;\frac{\alpha}{2}}} - \frac{1}{\chi^2_{n-1;1-\frac{\alpha}{2}}}.$$

Pertanto

$$e_n = \mathbb{E}_\theta[\mathcal{L}(\mathbf{X}_n)] = (n-1)\kappa(n,\alpha)\theta_2.$$

Si noti che, anche in questo caso, l'intervallo considerato non è quello di lunghezza (attesa) minima tra quelli di coefficiente di confidenza $1-\alpha$.

□

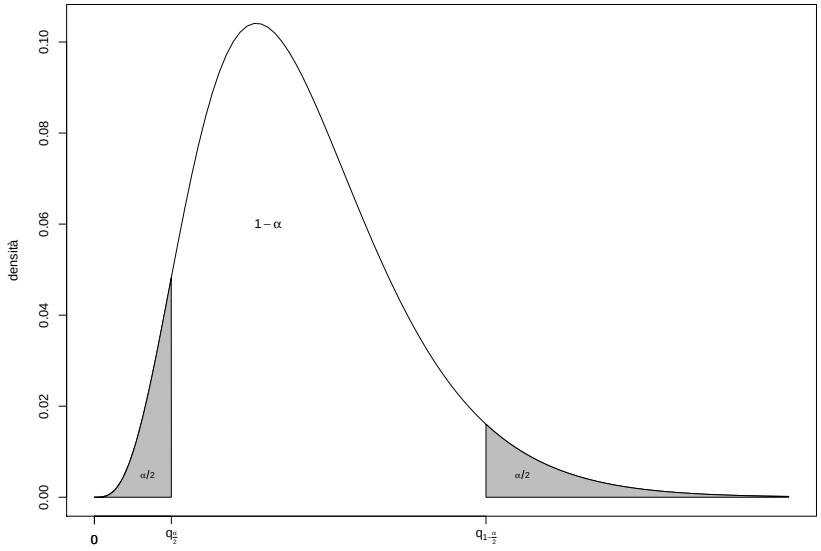

Figura 5.5: *Grafico della funzione di densità della quantità pivotale* $Q(\mathbf{X}_n, \theta_2) = \frac{(n-1)S_n^2}{\theta_2} \sim \chi^2_{n-1}$ *per intervallo di confidenza per* θ_2 *(varianza) nel modello* $N(\theta_1, \theta_2)$ *con* $1 - \alpha = 0.90$*,* $n = 10$*.*

5.2.4 Altri esempi

5.17 Esempio (Modello esponenziale). Dato un campione casuale da un modello $\text{Esp}(\theta)$, sappiamo che[2]

$$\sum_{i=1}^{n} X_i | \theta \sim \text{Ga}(n, \texttt{scale} = \theta).$$

Pertanto

$$Q(\mathbf{X}_n, \theta) = \frac{\sum_{i=1}^n X_i}{\theta} \sim \text{Ga}(n, 1)$$

è una quantità pivotale per θ. Presi, ad esempio, i percentili $q_{n;\frac{\alpha}{2}}$ e $q_{n;1-\frac{\alpha}{2}}$ della v.a. $\text{Ga}(n, 1)$, si ha che, $\forall \theta > 0$,

$$\begin{aligned} 1 - \alpha &= \mathbb{P}\left(q_{n;\frac{\alpha}{2}} \leq \frac{\sum_{i=1}^n X_i}{\theta} \leq q_{n;1-\frac{\alpha}{2}} \right) && (5.5) \\ &= \mathbb{P}\left(\frac{\sum_{i=1}^n X_i}{q_{n;1-\frac{\alpha}{2}}} \leq \theta \leq \frac{\sum_{i=1}^n X_i}{q_{n;\frac{\alpha}{2}}} \right) && (5.6) \end{aligned}$$

Un intervallo di confidenza di livello $1 - \alpha$ è quindi

$$C_{1-\alpha}(\mathbf{X}_n) = \left[\frac{\sum_{i=1}^n X_i}{q_{n;1-\frac{\alpha}{2}}}, \frac{\sum_{i=1}^n X_i}{q_{n;\frac{\alpha}{2}}} \right]$$

[2]Utilizzando per la v.a. Gamma la parametrizzazione `scala`, $f_X(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$, presa una costante $c > 0$ si ha che $cX \sim \text{Ga}(\alpha, c\beta)$.

di lunghezza e lunghezza attesa rispettivamente uguali a

$$\mathcal{L}(\mathbf{X}_n) = \left(\frac{1}{q_{n;\frac{\alpha}{2}}} - \frac{1}{q_{n;1-\frac{\alpha}{2}}} \right) \sum_{i=1}^{n} X_i \quad \text{e} \quad e_n = \left(\frac{1}{q_{n;\frac{\alpha}{2}}} - \frac{1}{q_{n;1-\frac{\alpha}{2}}} \right) n\theta.$$

Si può verificare che, al crescere di n, e_n tende a zero. □

5.18 Esempio (Modello uniforme). Riprendiamo in considerazione il modello uniforme in $[0, \theta]$, $\theta > 0$. La statistica sufficiente per il modello è $X_{(n)}$, che ha distribuzione campionaria $f_{X_{(n)}}(y;\theta) = ny^{n-1}/\theta^n$, $y \in [0, \theta]$. È semplice verificare che $Q(\mathbf{X}_n, \theta) = X_{(n)}/\theta$ ha distribuzione $f_Q(t) = nt^{n-1}$, $t \in [0, 1]$ e che quindi è una quantità pivotale per θ. Più precisamente,

$$Q(\mathbf{X}_n, \theta) = \frac{X_{(n)}}{\theta} \sim \text{Beta}(n, 1).$$

La Figura 5.6 riporta il grafico della funzione di densità di $Q(\mathbf{X}_n, \theta)$ con $n = 5$ e del suo percentile di livello α. In questo caso è semplice determinare sia l'insieme ET che quello HDI. Quest'ultimo si ottiene osservando che, $\forall \theta > 0$, si ha

$$\begin{aligned} 1 - \alpha &= \mathbb{P}\left(q_{n,\alpha} \leq \frac{X_{(n)}}{\theta} \leq 1 \right) \\ &= \mathbb{P}\left(X_{(n)} \leq \theta \leq \frac{X_{(n)}}{q_{n,\alpha}} \right), \end{aligned}$$

dove $q_{n,\alpha}$ è il quantile di livello α della v.a. Beta$(n, 1)$. In questo esempio è semplice ottenere un'espressione esplicita di $q_{n,\alpha}$ in funzione di α. Infatti:

$$1 - \alpha = \int_{q_{n,\alpha}}^{1} nt^{n-1} dt = t^n|_{q_{n,\alpha}}^{1} = 1^n - (q_{n,\alpha})^n = 1 - (q_{n,\alpha})^n,$$

da cui si ottiene che

$$q_{n,\alpha} = \alpha^{1/n}.$$

L'intervallo di confidenza di livello $1 - \alpha$ che si ottiene è quindi

$$C_{1-\alpha}(\mathbf{X}_n) = \left[X_{(n)} \leq \theta \leq \alpha^{-1/n} X_{(n)} \right].$$

Si osservi che l'intervallo ottenuto ha la stessa struttura dell'insieme di verosimiglianza di livello q ottenuto in 2.2.2: ponendo $q = q_{n,\alpha}$ i due insiemi coincidono.
La lunghezza e la lunghezza attesa dell'intervallo determinato sono

$$\mathcal{L}(\mathbf{X}_n) = (\alpha^{-1/n} - 1) X_{(n)} \quad \text{e} \quad e_n = (\alpha^{-1/n} - 1) \frac{n}{n+1} \theta.$$

In modo analogo è possibile determinare l'insieme ET, che sarà quindi:

$$C_{1-\alpha}(\mathbf{X}_n) = \left[\frac{X_{(n)}}{q_{n,1-\frac{\alpha}{2}}}, \frac{X_{(n)}}{q_{n,\frac{\alpha}{2}}} \right].$$

□

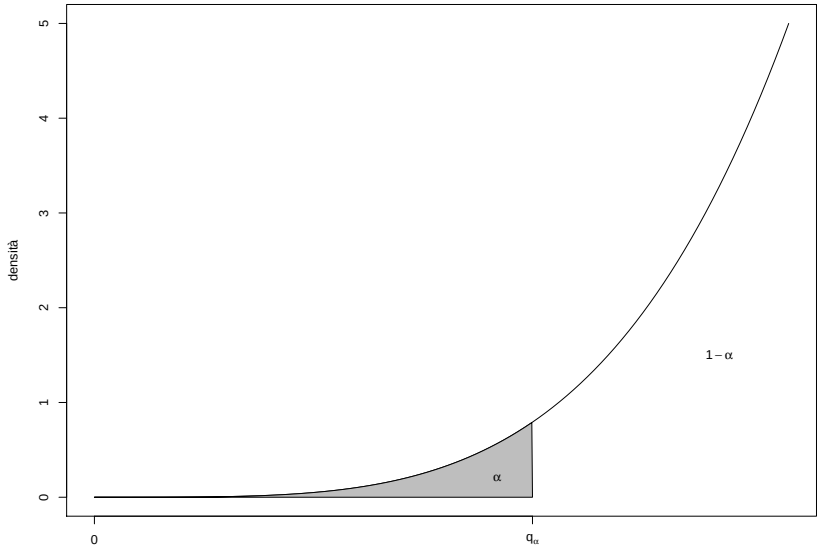

Figura 5.6: *Grafico della funzione di densità della quantità pivotale* $Q(\mathbf{X}_n, \theta) = \frac{X_{(n)}}{\theta} \sim$ *Beta*$(n, 1)$ *per intervallo di confidenza per* θ *nel modello Uniforme nell'intervallo* $[0, \theta]$ *con* $1 - \alpha = 0.90$*,* $n = 5$*.*

5.2.5 Confronto tra parametri di due popolazioni normali

5.19 Esempio (Confronto tra valori attesi di due distribuzioni normali). Consideriamo due campioni casuali indipendenti tra loro, costituiti rispettivamente da n_a e n_b osservazioni, provenienti da due popolazioni normali di parametri (μ_a, σ_a^2) e (μ_b, σ_b^2):

$$X_i^a \sim \mathrm{N}(\mu_a, \sigma_a^2), \quad i = 1, \ldots, n_a \qquad \text{e} \qquad X_i^b \sim \mathrm{N}(\mu_b, \sigma_b^2), \quad i = 1, \ldots, n_b.$$

Dal momento che

$$\overline{X}_{n_a} \sim \mathrm{N}\left(\mu_a, \frac{\sigma_a^2}{n_a}\right) \qquad \text{e} \qquad \overline{X}_{n_b} \sim \mathrm{N}\left(\mu_b, \frac{\sigma_b^2}{n_b}\right),$$

per le proprietà delle v.a. normali e per l'indipendenza di $\overline{X}_{n_a}$ e $\overline{X}_{n_b}$ si ha che

$$\overline{X}_{n_a} - \overline{X}_{n_b} \sim \mathrm{N}\left(\mu_a - \mu_b, \frac{\sigma_a^2}{n_a} + \frac{\sigma_b^2}{n_b}\right).$$

Per confrontare tra loro i valori attesi μ_a e μ_b possiamo determinare l'intervallo di confidenza centrale del parametro

$$\mu = \mu_a - \mu_b.$$

1. **Varianze note.** Nel caso di varianze note, una quantità pivotale per μ basata sul campione $\mathbf{X}_n = (\mathbf{X}_{n_a}, \mathbf{X}_{n_b})$ è quindi data dalla funzione

$$Q(\mathbf{X}_n, \mu) = \frac{\overline{X}_{n_a} - \overline{X}_{n_b} - \mu}{\sqrt{\frac{\sigma_a^2}{n_a} + \frac{\sigma_b^2}{n_b}}} \sim \mathrm{N}(0, 1)$$

Scelto un livello di confidenza pari a $1-\alpha$, si ha che, $\forall \mu \in \mathbb{R}$,

$$1-\alpha = \mathbb{P}_\theta\left(-z_{1-\frac{\alpha}{2}} \leq Q(\mathbf{X}_n, \mu) \leq z_{1-\frac{\alpha}{2}}\right)$$

da cui, con gli stessi passaggi dell'esempio 5.14, si ottiene il seguente intervallo di confidenza centrale per μ:

$$\overline{X}_{n_a} - \overline{X}_{n_b} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_a^2}{n_a} + \frac{\sigma_b^2}{n_b}}.$$

2. **Varianze incognite.** Consideriamo, per semplicità, il caso di varianze incognite ma uguali: $\sigma^2_{n_a} = \sigma^2_{n_b} = \sigma^2$. In questo caso

$$\frac{\overline{X}_{n_a} - \overline{X}_{n_b} - \mu}{\sigma\sqrt{\frac{1}{n_a} + \frac{1}{n_b}}} \sim \mathrm{N}(0,1) \tag{5.7}$$

e per ottenere una q. pivotale per μ dobbiamo sostituire σ con un suo stimatore. Siano $S^2_{n_a}$ e $S^2_{n_b}$ le varianze campionarie corrette che, come noto, sono stimatori non distorti di σ^2. Per sfruttare l'informazione che entrambi i campioni forniscono su σ^2, consideriamo lo stimatore

$$S_p^2 = \frac{(n_a-1)S^2_{n_a} + (n_b-1)S^2_{n_b}}{n_a + n_b - 2}$$

(detto stimatore *pooled* di σ^2) ottenuto dalla media ponderata di $S^2_{n_a}$ e $S^2_{n_b}$, con pesi proporzionali a $n_a - 1$ e $n_b - 1$. Ricordando che

$$\frac{(n_a-1)S^2_{n_a}}{\sigma^2} \sim \chi^2_{n_a-1} \qquad \text{e} \qquad \frac{(n_b-1)S^2_{n_b}}{\sigma^2} \sim \chi^2_{n_b-1}$$

e che le due v.a. sono indipendenti tra loro (grazie all'indipendenza delle due varianze campionarie), abbiamo che, per la proprietà di additività della v.a. χ^2,

$$\frac{(n_a+n_b-2)S_p^2}{\sigma^2} = \frac{(n_a-1)S^2_{n_a}}{\sigma^2} + \frac{(n_b-1)S^2_{n_b}}{\sigma^2} \sim \chi^2_{n_a+n_b-2}. \tag{5.8}$$

Quest'ultima v.a. è indipendente da $\overline{X}_{n_a} - \overline{X}_{n_b}$ e quindi possiamo costruire una v.a. t dal rapporto tra la v.a. $\mathrm{N}(0,1)$ in (5.7) e la radice del rapporto tra il χ^2 in (5.8) e i corrispondenti gradi di libertà:

$$\frac{\frac{\overline{X}_{n_a} - \overline{X}_{n_b} - \mu}{\sigma\sqrt{\frac{1}{n_a}+\frac{1}{n_b}}}}{\sqrt{\frac{(n_a+n_b-2)S_p^2}{\sigma^2}\frac{1}{(na+n_b-2)}}} = \frac{\overline{X}_{n_a} - \overline{X}_{n_b} - \mu}{S_p\sqrt{\frac{1}{n_a}+\frac{1}{n_b}}} \sim t_{n_a+n_b-2}.$$

Possiamo quindi utilizzare come q. pivotale per μ la funzione

$$Q(\mathbf{X}_n, \mu) = \frac{\overline{X}_{n_a} - \overline{X}_{n_b} - \mu}{S_p\sqrt{\frac{1}{n_a}+\frac{1}{n_b}}}$$

da cui, con gli usuali passaggi, otteniamo l'intervallo di confidenza centrale di livello $1-\alpha$ per μ:

$$\overline{X}_{n_a} - \overline{X}_{n_b} \pm t_{n_a+n_b-2;1-\frac{\alpha}{2}} S_p \sqrt{\frac{1}{n_a} + \frac{1}{n_b}}.$$

□

5.20 Esempio(Confronto tra varianze di due distribuzioni normali). Consideriamo, come nel precedente esempio, due popolazioni normali, ma supponiamo di essere ora interessati a un intervallo di confidenza per il confronto tra le varianze incognite σ_a^2 e σ_b^2. Consideriamo quindi il parametro

$$\psi = \frac{\sigma_a^2}{\sigma_b^2}$$

per il quale cerchiamo una q. pivotale.

1. **Valori attesi noti.** Consideriamo i due stimatori UMVUE di σ_a^2 e σ_b^2:

$$S_{0_a}^2 = \frac{1}{n_a}\sum_{i=1}^{n_a}(X_i^a - \mu_a)^2 \qquad \text{e} \qquad S_{0_b}^2 = \frac{1}{n_b}\sum_{i=1}^{n_b}(X_i^b - \mu_b)^2.$$

Sappiamo che

$$\frac{n_a S_{0_a}^2}{\sigma_a^2} \sim \chi_{n_a}^2 \qquad \text{e} \qquad \frac{n_b S_{0_b}^2}{\sigma_b^2} \sim \chi_{n_b}^2.$$

Per l'indipendenza e per il teorema 3.29, riguardante la costruzione di una v.a. F di Fisher abbiamo che

$$\frac{\frac{n_a S_{0_a}^2}{\sigma_a^2}\frac{1}{n_a}}{\frac{n_b S_{0_b}^2}{\sigma_b^2}\frac{1}{n_b}} = \frac{S_{0_a}^2}{S_{0_b}^2}/\frac{\sigma_a^2}{\sigma_b^2} \sim F_{n_a,n_b}.$$

La funzione

$$Q(\mathbf{X}_n, \psi) = \frac{S_{0_a}^2}{S_{0_b}^2}/\frac{\sigma_a^2}{\sigma_b^2}$$

è una q. pivotale per ψ. Indicati con $F_{n_a,n_b;\frac{\alpha}{2}}$ e $F_{n_a,n_b;1-\frac{\alpha}{2}}$ i percentili di livello $\frac{\alpha}{2}$ e $1-\frac{\alpha}{2}$ della F di Fisher con (n_a, n_b) gradi di libertà, dalla relazione che segue, valida per ogni ψ e per ogni valore di confidenza prescelto, $1-\alpha$

$$1-\alpha = \mathbb{P}\left[F_{n_a,n_b;\frac{\alpha}{2}} \leq \frac{S_{0_a}^2}{S_{0_b}^2}/\frac{\sigma_a^2}{\sigma_b^2} \leq F_{n_a,n_b;1-\frac{\alpha}{2}}\right]$$

si ottiene l'intervallo di confidenza (a code uguali) di livello $1-\alpha$:

$$\left[\frac{S_{0_a}^2/S_{0_b}^2}{F_{n_a,n_b;1-\frac{\alpha}{2}}}, \frac{S_{0_a}^2/S_{0_b}^2}{F_{n_a,n_b;\frac{\alpha}{2}}},\right]$$

2. **Valori attesi incogniti.** Nel caso di valori attesi incogniti, la q. pivotale per ψ si costruisce a partire dalle varianze campionarie corrette:

$$S_{n_a}^2 = \frac{1}{n_a-1}\sum_{i=1}^{n_a}(X_i^a - \overline{X}_{n_a})^2 \qquad \text{e} \qquad S_{n_b}^2 = \frac{1}{n_b-1}\sum_{i=1}^{n_b}(X_i^b - \overline{X}_{n_b})^2.$$

La funzione

$$Q(\mathbf{X}_n, \psi) = \frac{S_{n_a}^2}{S_{n_b}^2}/\frac{\sigma_a^2}{\sigma_b^2} \sim F_{n_a-1,n_b-1}$$

è quindi la q. pivotale da utilizzare. In modo del tutto analogo al caso precedente si ottiene quindi l'intervallo di confidenza di livello $1-\alpha$

$$\left[\frac{S_{n_a}^2/S_{n_b}^2}{F_{n_a-1,n_b-1;1-\frac{\alpha}{2}}}, \frac{S_{n_a}^2/S_{n_b}^2}{F_{n_a-1,n_b-1;\frac{\alpha}{2}}},\right],$$

dove $F_{n_a-1,n_b-1;\frac{\alpha}{2}}$ e $F_{n_a-1,n_b-1;1-\frac{\alpha}{2}}$ sono i percentili di livello $\frac{\alpha}{2}$ e $1-\frac{\alpha}{2}$ della F d Fisher con (n_a-1, n_b-1) gradi di libertà.

□

5.3 Valutazione degli intervalli e ottimalità

Esistono molti diversi criteri per il confronto tra stimatori intervallari di un parametro. Come nel caso di stima puntuale, la teoria delle decisioni statistiche offre il contesto ideale per il confronto tra procedure alternative. Si tratta, anche in questo caso, di definire delle misure che quantificano le perdite associate all'uso di stimatori intervallari concorrenti e nel confrontare le perdite attese corrispondenti. Senza entrare nel dettaglio (si rimanda a Piccinato (2009) per una trattazione decisionale), è abbastanza intuitivo assumere che, per valutare uno stimatore intervallare $C(\mathbf{X}_n)$, si tenga conto della sua *lunghezza aleatoria* e del suo *livello di confidenza*. Uno stimatore intervallare risulta tanto più valido quanto maggiore è la probabilià di osservare un intervallo con lunghezza non elevata e quanto più elevato è il suo livello di confidenza, ovvero la minima probabilità di copertura. La massimizzazione contestuale di queste due probabilità non è però molto agevole. Un modo molto più semplice per confrontare stimatori intervallari è quello di fissare il livello di confidenza di stimatori concorrenti e di scegliere tra intervalli con uguale livello di confidenza, quello con lunghezza attesa e_n minima (ricordiamo che la lunghezza di uno stimatore C di θ è infatti, in generale, aleatoria). In altri termini: l'intervallo *ottimo* è quello di lunghezza attesa e_n minima tra quelli il cui livello di confidenza è pari a un valore fissato $1-\alpha$.

5.21 Esempio(Modello normale). Consideriamo il modello $\mathrm{N}(\theta, \sigma^2)$ (varianza nota). Nell'esempio 5.14 abbiamo ottenuto, come 'intervallo di confidenza di livello $1-\alpha$ per θ, l'intervallo

$$C_{1-\alpha}(\mathbf{X}_n) = [\overline{X}_n - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \overline{X}_n + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}]. \tag{5.9}$$

Si tratta di uno tra gli infiniti intervalli di confidenza di stesso livello che avremmo potuto ottenere con il metodo delle quantità pivotali. Infatti, per ogni altra coppia di percentili z_1 e z_2 della v.a. $\mathrm{N}(0,1)$, tali che $\mathbb{P}[z_1 \leq Z \leq z_z] = 1-\alpha$, l'intervallo

$$C'_{1-\alpha}(\mathbf{X}_n) = \left[\overline{X}_n - z_2\frac{\sigma}{\sqrt{n}}, \overline{X}_n + z_1\frac{\sigma}{\sqrt{n}}\right]$$

è un intervallo di confidenza per θ di medesimo livello di $C_{1-\alpha}$. Se però confrontiamo le lunghezze (attese) dei due intervalli, e_n e e'_n, abbiamo che (come conseguenza del teorema che enunceremo a seguire questo esempio), per ogni valore di $\alpha \in (0,1)$ e per ogni $n \in \mathbb{N}$,

$$e'_n = (z_2 - z_1)\frac{\sigma}{\sqrt{n}} > 2z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} = e_n.$$

Nel problema considerato, l'intervallo (5.9) è quindi quello di lunghezza minima tra quelli di livello $1-\alpha$ scelto. □

Per la determinazione dell'intervallo di confidenza di livello $1-\alpha$ di lunghezza minima risulta utile il seguente teorema (la cui dimostrazione si trova in Casella-Berger 2002, p. 441-442). L'idea è la seguente: si trovano gli insiemi pivotali di lunghezza minima e da questi si ottengono gli insiemi di confidenza per θ di lunghezza minima.

5.22 Teorema Sia $f(\cdot)$ una funzione di densità di probabilità unimodale con moda nel punto x^*. Sia $[a,b]$ un intervallo che soddisfa i seguenti 3 requisiti:

(a) $\int_a^b f(x)dx = 1-\alpha, \qquad \alpha \in (0,1);$

(b) $f(a) = f(b) > 0;$

(c) $x^* \in [a,b].$

L'intervallo $[a,b]$ è quello di lunghezza minima tra quelli che soddisfano la condizione (a). □

Quando una quantità pivotale Q ha densità $f_Q(\cdot)$ che soddisfa le ipotesi del teorema enunciato, questo può essere utilizzato per la determinazione dei intervalli pivotali $[q_1^\star, q_2^\star]$ di fissato livello $1-\alpha$ e di lunghezza minima pari a $(q_2^\star - q_1^\star)$. Se l'intervallo di confidenza per θ risultante ha lunghezza proporzionale a $(q_2^\star - q_1^\star)$, tale intervallo è quello di lunghezza minima tra quelli di livello $1-\alpha$ per θ. Ciò avviene se la quantità pivotale Q ha densità f_Q che, oltre a soddisfare le ipotesi del precedente teorema, è anche funzione *crescente* di θ: in tal modo tale la sua lunghezza risulta proporzionale alla lunghezza dell'intervallo pivotale. Se invece Q è funzione decrescente di θ, l'applicazione del teorema per la determinazione dell'intervallo di confidenza HD per θ è meno banale. Per illustrare quanto detto, consideriamo per semplicità i due casi seguenti.

1. $Q(\mathbf{X}_n, \theta) = \theta T(\mathbf{X}_n)$, dove $T(\mathbf{X}_n)$ è una opportuna funzione dei dati campionari. In questo caso Q è funzione *crescente* di θ ed è semplice verificare che, scelta una coppia di valori (q_1, q_2) che definisca un intervallo pivotale di livello $1-\alpha$, l'intervallo di confidenza risultante per θ è $C_{1-\alpha} = \left[\frac{q_1}{T(\mathbf{X}_n)}, \frac{q_2}{T(\mathbf{X}_n)}\right]$, la cui lunghezza è pari a $\mathcal{L}(\mathbf{X}_n) = (q_2 - q_1)T(\mathbf{X}_n)$, ovvero proporzionale alla lunghezza dell'intervallo pivotale $(q_2 - q_1)$. Se scegliamo i valori $[q_1^\star, q_2^\star$ che minimizzano la lunghezza della quantità pivotale, questi minimizzano anche la lunghezza di $C_{1-\alpha}$. L'intevallo di confidenza di livello $1-\alpha$ di lunghezza minima è quindi $C_{1-\alpha}^\star = \left[\frac{q_1^\star}{T(\mathbf{X}_n)}, \frac{q_2^\star}{T(\mathbf{X}_n)}\right]$.

2. $Q(\mathbf{X}_n, \theta) = \frac{T(\mathbf{X}_n)}{\theta}$, dove $T(\mathbf{X}_n)$ è, come sopra, una opportuna funzione dei dati campionari. In questo caso Q è funzione *decrescente* di θ ed è semplice verificare che, scelta una coppia di valori (q_1, q_2) che definisca un intervallo pivotale di livello $1-\alpha$, l'intervallo di confidenza risultante per θ è $C_{1-\alpha} = \left[\frac{T(\mathbf{X}_n)}{q_2}, \frac{T(\mathbf{X}_n)}{q_1}\right]$ la cui lunghezza è pari a $\mathcal{L}(\mathbf{X}_n) = \left(\frac{1}{q_1} - \frac{1}{q_2}\right) T(\mathbf{X}_n)$, ovvero proporzionale a $\left(\frac{1}{q_1} - \frac{1}{q_2}\right)$. Ora, i valori $q_1^\star, q_2^\star$ che minimizzano la lunghezza $(q_2 - q_1)$ dell'intervallo pivotale $[q_1, q_2]$ non minimizzano la lunghezza $\mathcal{L}(\mathbf{X}_n) = \left(\frac{1}{q_1} - \frac{1}{q_2}\right) T(\mathbf{X}_n)$ dell'intervallo di confidenza per θ. Questo problema si può risolvere usando la quantità pivotale $\overline{Q}(\mathbf{X}_n, \theta) = \frac{1}{Q(\mathbf{X}_n, \theta)} = \frac{\theta}{T(\mathbf{X}_n)}$, la cui distribuzione è quella dell'inversa della v.a. Q di partenza. Si applica quindi il teorema a $f_{\overline{Q}}$ e si ottiene l'intervallo pivotale HD $[\overline{q}_1^\star, \overline{q}_2^\star]$ e da questo l'intervallo di confidenza HD per θ $C_{1-\alpha}^\star = [\overline{q}_1^\star T(\mathbf{X}_n), \overline{q}_2^\star T(\mathbf{X}_n)]$ la cui lunghezza $(\overline{q}_2^\star - \overline{q}_1^\star)T(\mathbf{X}_n)$ è minima. Ovviamente $\overline{q}_i^\star$, $i = 1, 2$ sono ora opportuni quantili della distribuzione di $\overline{Q}$, inversa della quantità pivotale di partenza.

Si noti che, nell'esempio 5.14, per un fissato valore $1-\alpha$, i percentili $\pm z_{1-\frac{\alpha}{2}}$ sono gli unici valori reali che soddisfano le tre condizioni del teorema precedente. L'intervallo (5.9) è quindi quello di lunghezza minima tra quelli di livello $1-\alpha$. Lo stesso vale per i percentili $\pm t_{n-1;1-\frac{\alpha}{2}}$ dell'esempio con varianza incognita. Sempre grazie al teorema precedente possiamo invece escludere che, in generale, gli intervalli (a code uguali) degli esempi 5.16 e 5.17 siano quelli di lunghezza minima tra quelli di livello di confidenza fissato.

5.4 Intervalli di confidenza asintotici

Non per tutti i modelli statistici è agevole o possibile determinare quantità pivotali. In alcuni casi (ad esempio quando si considerano v.a. discrete), nonostante sia possibile determinare intervalli di confidenza per i parametri, sorgono difficiltà analitiche non banali. Nei problemi in cui sono disponibili stimatori asintoticamente normali per il parametro, possiamo tuttavia ottenere piuttosto agevolmente degli intervalli di confidenza approssimati per il parametro. L'idea è che gli intervalli che si ottengono hanno un livello di confidenza che, al crescere della numerosità campionaria tende al livello *nominale* (ovvero quello dichiarato) $1-\alpha$.

Abbiamo visto che (definizione 4.78) una successione $(d_n, n \in \mathbb{N})$ di stimatori di θ è asintoticamente normale se esiste una successione numerica $(\kappa_n, n \in \mathbb{N})$ tale che

$$k_n(d_n - \theta) \xrightarrow{d} \mathrm{N}(0, v(\theta)).$$

In questo caso, per n sufficientemente elevato, abbiamo che

$$d_n \dot{\sim} \mathrm{N}(\theta, v_a(d_n)),$$

dove $v_a(d_n) = v(\theta)/\kappa_n^2$ è la varianza asintotica di d_n, quantità che in genere dipende da θ. Se $\widehat{v_a(d_n)}$ è uno stimatore consistente di $v_a(d_n)$, allora per la funzione $\widetilde{Q}(\mathbf{X}_n, \theta)$si ha

$$\widetilde{Q}(\mathbf{X}_n, \theta) = \frac{d_n - \theta}{\sqrt{\widehat{v_a(d_n)}}} \dot{\sim} \mathrm{N}(0, 1),$$

che quindi definiamo *quantità pivotale asintotica* per θ. Abbiamo infatti che, per un generico livello di confidenza $1-\alpha$ desiderato, presi i percentili $\pm z_{1-\frac{\alpha}{2}}$ della v.a. normale standardizzata,

$$\mathbb{P}_\theta\left(-z_{1-\frac{\alpha}{2}} \leq \widetilde{Q}(\mathbf{X}_n, \theta) \leq z_{1-\frac{\alpha}{2}}\right) \simeq 1-\alpha.$$

Dalla precedente relazione si ottiene un intervallo di confidenza di livello approssimativamente uguale a $1-\alpha$ per θ:

$$\widetilde{C}_{1-\alpha}(\mathbf{X}_n) = \left[d_n - z_{1-\frac{\alpha}{2}}\sqrt{\widehat{v_a(d_n)}}, d_n + z_{1-\frac{\alpha}{2}}\sqrt{\widehat{v_a(d_n)}}\right].$$

5.23 Esempio (Modello uniforme). Nel caso del modello uniforme in $[0, \theta]$ abbiamo verificato che lo stimatore dei momenti è $\widehat{\theta}_m(\mathbf{X}_n) = 2\overline{X}_n$ e che

$$2\overline{X}_n \dot{\sim} \mathrm{N}\left(\theta, \frac{\theta^2}{3n}\right)$$

In questo caso uno stimatore della varianza asintotica è $\widehat{v_a(2\overline{X}_n)} = (2\overline{X}_n)^2/3n$. Si ottiene quindi il seguente intervallo di confidenza asintotico di livello $1-\alpha$ per θ:

$$2\overline{X}_n \pm z_{1-\frac{\alpha}{2}}\frac{2\overline{X}_n}{\sqrt{3n}}. \tag{5.10}$$

La Figura 5.7 mostra l'andamento della probabilità di copertura dell'intervallo asintotico (5.10) in funzione della numerosità campionaria. Per valori bassi di n la probabilià di copertura effettiva dell'intervallo si discosta in modo non irrilevante dal valore *nominale*$1-\alpha$, qui posto uguale a 0.95. Ad esempio, per $n = 10$, la probabili$\widehat{\text{a}}$ di copertura dell'intervallo è pari a 0.923. Tuttavia, al crescere di n, il valore effettivo della probabilità di copeetura tende ad essere sempre più vicino a quello nominale. Ad esempio, per $n = 50$, la probabili$\widehat{\text{a}}$ di copertura dell'intervallo è pari a 0.945.

Ch5-copertura-s.mom-unif

□

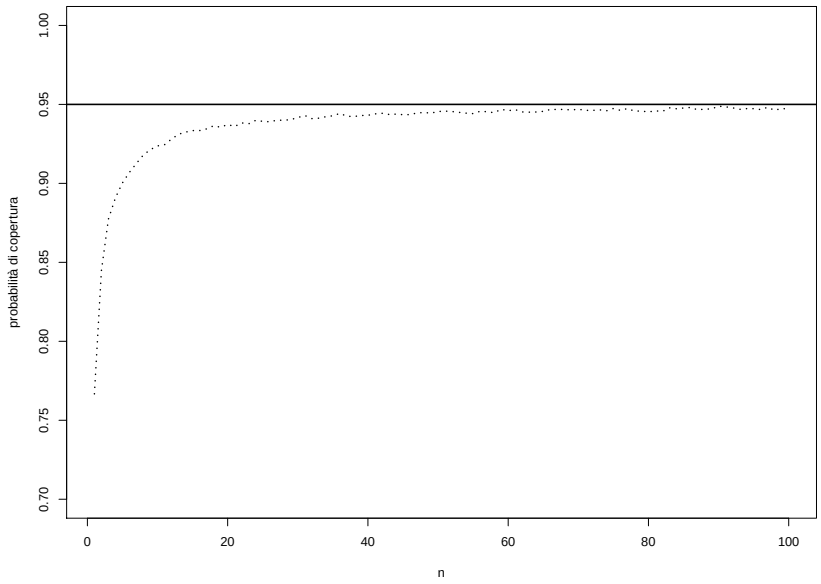

Figura 5.7: *Grafico della probabilità di copertura dell'intervallo asintotico (5.10) per il parametro incognito nel modello Uniforme nell'intervallo* $[0, \theta]$ *con* $1 - \alpha = 0.95$ *in funzione della numerosità campionaria.*

5.4.1 Intervalli asintotici ottenuti da stimatori di massima verosimiglianza

Dalla normalità asintotica degli stimatori di massima verosimiglianza (vedi 4.7.2) si ottiene, per n sufficientemente elevato, la seguente quantità pivotale approssimata per θ:

$$\widetilde{Q}_{mv}(\mathbf{X}_n, \theta) = \frac{\widehat{\theta}_{mv}(\mathbf{X}_n) - \theta}{\sqrt{\widehat{I_n(\theta)}^{-1}}},$$

dove $\widehat{I_n(\theta)}$ indica lo stimatore dell'informazione attesa di Fisher (ovvero $\mathcal{I}_n(\mathbf{X}_n)$ oppure $I_n(\widehat{\theta}_{mv}(\mathbf{X}_n))$). Da questa quantità pivotale si ottiene l'intervallo di confidenza approssimato di livello $(1 - \alpha)$ per θ:

$$\widehat{\theta}_{mv}(\mathbf{X}_n) \pm z_{1-\frac{\alpha}{2}} \sqrt{\widehat{I_n(\theta)}^{-1}} \tag{5.11}$$

5.24 Esempio (Modello bernoulliano). In questo caso, ricordando che, per un campione casuale di dimensione n si ha

- $\widehat{\theta}_{mv}(\mathbf{X}_n) = \overline{X}_n$
- $I_n(\theta) = \frac{n}{\theta(1-\theta)}$
- $\mathcal{I}_n(\mathbf{X}_n) = \frac{n}{\overline{X}_n(1-\overline{X}_n)}$

l'intervallo di confidenza approssimato di livello $1-\alpha$ cercato per è

$$\overline{X}_n \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}}.$$

□

5.25 Esempio (Modello Poisson). In questo caso, ricordando che, per un campione casuale di dimensione n si ha

- $\widehat{\theta}_{mv}(\mathbf{X}_n) = \overline{X}_n$
- $I_n(\theta) = \frac{n}{\theta}$
- $\mathcal{I}_n(\mathbf{X}_n) = \frac{n}{\overline{X}_n}$

l'intervallo di confidenza approssimato di livello $1-\alpha$ cercato per è

$$\overline{X}_n \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\overline{X}_n}{n}}.$$

□

5.26 Esempio (Modello beta). Consideriamo il modello Beta$(\theta, 1)$ (vedi esempio 4.96). In questo caso, ricordando che, per un campione casuale di dimensione n si ha

- $\widehat{\theta}_{mv}(\mathbf{X}_n) = -\frac{n}{\sum_{i=1}^n \ln X_i}$
- $I_n(\theta) = \frac{n}{\theta^2}$
- $\mathcal{I}_n(\mathbf{X}_n) = \frac{(\sum_{i=1}^n \ln X_i)^2}{n}$

l'intervallo di confidenza approssimato di livello $1-\alpha$ cercato per è

$$-\frac{n}{\sum_{i=1}^n \ln X_i} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{n}{(\sum_{i=1}^n \ln X_i)^2}}.$$

□

5.27 Esempio (Confronto tra parametri di due modelli bernoulliani). Supponiamo di disporre di due campioni indipendenti di dimensioni rispettivamente n_a e n_b, ciascuno proveniente da due popolazioni bernoulliane di parametro θ_a e θ_b. Per valori di n sufficientemente elevati abbiamo che

$$\overline{X}_{n_a} - \overline{X}_{n_b} \dot{\sim} \mathrm{N}\left(\theta_a - \theta_b, \frac{\theta_a(1-\theta_a)}{n_a} + \frac{\theta_b(1-\theta_b)}{n_b}\right).$$

Utilizzando per la varianza la stima $\overline{X}_{n_a}(1-\overline{X}_{n_a})/n_a + \overline{X}_{n_b}(1-\overline{X}_{n_b})/n_b$ si ottiene la seguente quantità pivotale approssimata per $\theta = \theta_a - \theta_b$

$$\widetilde{Q}(\mathbf{X}_n;\theta) = \frac{(\overline{X}_{n_a} - \overline{X}_{n_b}) - \theta}{\sqrt{\frac{\overline{X}_{n_a}(1-\overline{X}_{n_a})}{n_a} + \frac{\overline{X}_{n_b}(1-\overline{X}_{n_b})}{n_b}}}$$

da cui un intervallo di confidenza di livello approssimato di livello $1-\alpha$ per θ:

$$(\overline{X}_{n_a} - \overline{X}_{n_b}) \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\overline{X}_{n_a}(1-\overline{X}_{n_a})}{n_a} + \frac{\overline{X}_{n_b}(1-\overline{X}_{n_b})}{n_b}}$$

□

Intervalli di confidenza asintotici per funzioni del parametro del modello

Analogamente, se si considera una funzione $g(\theta)$ per θ, tale che $g'(\theta) \neq 0$, per quanto visto nel Paragrafo 4.7.3 si ottiene, per n sufficientemente elevato, la seguente quantità pivotale approssimata per $g(\theta)$:

$$\widetilde{Q}_{mv}(\mathbf{X}_n, g(\theta)) = \frac{g(\widehat{\theta}_{mv}) - g(\theta)}{|g'(\widehat{\theta}_{mv})| \sqrt{\widehat{I_n(\theta)}^{-1}}},$$

da cui l'intervallo di confidenza approssimato di livello $(1-\alpha)$ per $g(\theta)$:

$$g[\widehat{\theta}_{mv}(\mathbf{X}_n)] \pm z_{1-\frac{\alpha}{2}} |g'(\widehat{\theta}_{mv})| \sqrt{\widehat{I_n(\theta)}^{-1}}. \tag{5.12}$$

5.28 Esempio (IC per log-odds, modello bernoulliano). Dato un campione casuale di dimensione n sufficientemente elevato da una popolazione bernoulliana, possiamo ottenere un intervallo di confidenza approssimato di livello $1-\alpha$ per il parametro log-odds, $g(\theta) = \log \frac{\theta}{1-\theta}$. In questo caso (vedi esempio 4.99) si ha:

- $\widehat{\theta}_{mv}(\mathbf{X}_n) = \overline{X}_n$
- $g(\widehat{\theta}_{mv}) = \log \frac{\overline{X}_n}{1-\overline{X}_n}$
- $g'(\theta) = \frac{1}{\theta(1-\theta)}$
- $I_n(\theta) = \frac{n}{\theta(1-\theta)}$
- $g'(\widehat{\theta}_{mv})^2 \mathcal{I}_n(\mathbf{X}_n)^{-1} = \frac{1}{n\overline{X}_n(1-\overline{X}_n)}$

e quindi l'intervallo di confidenza approssimato di livello $1-\alpha$ cercato è

$$\log \frac{\overline{X}_n}{1-\overline{X}_n} \pm z_{1-\frac{\alpha}{2}} \frac{1}{[n\overline{X}_n(1-\overline{X}_n)]^{1/2}}.$$

□

5.29 Esempio (IC per una probabilità, modello di Poisson). Dato un campione casuale di dimensione n sufficientemente elevato da una popolazione di Poisson di parametro θ, possiamo ottenere un intervallo di confidenza approssimato di livello $1-\alpha$ per il parametro $g(\theta) = e^{-\theta} = \mathbb{P}_\theta(X = 0)$. In questo caso (vedi esempio 4.100) si ha:

- $\widehat{\theta}_{mv}(\mathbf{X}_n) = \overline{X}_n$
- $g(\widehat{\theta}_{mv}) = e^{-\overline{X}_n}$
- $g'(\theta) = -e^{-\theta}$
- $I_n(\theta) = \frac{n}{\theta}$
- $g'(\widehat{\theta}_{mv})^2 \mathcal{I}_n(\mathbf{X}_n)^{-1} = e^{-2\overline{X}_n} \times \frac{\overline{X}_n}{n}$

e quindi l'intervallo di confidenza approssimato di livello $1-\alpha$ cercato è

$$e^{-\overline{X}_n} \pm z_{1-\frac{\alpha}{2}} e^{-\overline{X}_n} \sqrt{\frac{\overline{X}_n}{n}}.$$

□

5.5 Determinazione della numerosità campionaria

La scelta della dimensione campionaria è una fase del processo inferenziale che si deve affrontare prima di avere osservato i dati. Il problema rientra nella più ampia tematica della *scelta del disegno* (o dell'*esperimento*) ed è di natura *pre-sperimentale*, a differenza della stima (puntuale e intervallare) e della verifica di ipotesi, che sono di tipo *post-sperimentale*. La scelta del numero di osservazioni da includere nel campione, pur essendo una delle domande più frequenti rivolte a uno statistico, è un problema non semplice da affrontare. In generale, nelle situazioni più comuni, la variabilità delle procedure di stima (ad esempio delle distribuzioni degli stimatori puntuali o delle caratteristiche degli stimatori intervallari) tende a dimunuire al crescere della dimensione campionaria. Questo vuol dire che, al crescere di n, l'accuratezza di procedure *consistenti* tende a crescere.

Su questa ipotesi si basano i più semplici metodi di scelta della dimensione del campione. Uno dei più comuni tra questi metodi si basa sul controllo della lunghezza degli stimatori intervallari, $\mathcal{L}$. Dal momento che, prima di effettuare l'esperimento, $\mathcal{L}(\mathbf{X}_n)$ è, una variabile aleatoria, il criterio si basa sul controllo probabilistico della sua distribuzione di probabilità. In particolare, si può considerare la lunghezza attesa di un intervallo, e_n e, supponendo che la successione $(e_n, n \in \mathbb{N})$ sia almeno definitivamente (almeno da un certo valore di n in poi) monotona decrescente (come di fatto avviene nelle situazioni usuali), determinare il minimo valore di n tale che e_n sia al disotto di una soglia prefissata, ℓ_0. Esplicitando la dipendenza di $e_n = \mathbb{E}_\theta[\mathcal{L}_{1-\alpha}(\mathbf{X}_n)]$ da θ e da α, definiamo con

$$n^*(\theta, \alpha, \ell_0) = \min\{n \in \mathbb{N} : e_n(\theta, \alpha) \leq \ell_0\} \tag{5.13}$$

la numerosità campionaria minima che garantisce una lunghezza non superiore al valore ℓ_0 prescelto. È evidente quindi che n^* dipende non solo dal livello di confidenza e dal valore ℓ_0 prescelti, ma anche dal valore del parametro incognito θ per il quale intendiamo effettuare la stima. Questo apparente paradosso (dover conoscere il valore del parametro per scegliere la dimensione del campione con cui stimare il parametro stesso) è una caratteristica dei problemi di scelta dell'esperimento i cui criteri si basano su elaborazioni della distribuzione campionaria $f_n(\cdot;\theta)$ di $\mathbf{X}_n$, che dipende ovviamente da θ. Per aggirare il problema, si ipotizza un valore θ_D per il parametro, detto *valore di disegno* e, per una data coppia di valori (α, ℓ_0) si sceglie come numerosità

$$n^*(\theta_D, \alpha, \ell_0).$$

5.30 Esempio (Modello normale). Consideriamo il problema della scelta di n nei problemi di inferenza sul valore atteso del modello normale.

1. **Varianza nota**. Si consideri il modello $\mathrm{N}(\theta, \sigma^2)$ (con varianza nota). In questo caso abbiamo visto (Esempio 5.14) che, utilizzando l'intervallo di confidenza centrale di livello $1-\alpha$, si ha
$$\mathcal{L}(\mathbf{X}_n) = 2z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$
che non dipende da $\mathbf{X}_n$. Pertanto $e_n = \mathcal{L}$. Osservando che
$$e_n \leq \ell_0 \quad \Longleftrightarrow \quad n \geq \frac{4z_{1-\frac{\alpha}{2}}{}^2\sigma^2}{\ell_0^2},$$
si ottiene (omettendo in questo caso la dipendenza di n^* da θ)
$$n^*(\alpha, \ell_0) = \left\lceil \frac{4z_{1-\frac{\alpha}{2}}{}^2\sigma^2}{\ell_0^2} \right\rceil.$$

2. **Varianza incognita**. Abbiamo visto che, in questo caso, l'insieme

$$C_{1-\alpha}(\mathbf{X}_n) = \left[\overline{X}_n - t_{n-1;1-\frac{\alpha}{2}}\frac{S_n}{\sqrt{n}}, \overline{X}_n + t_{n-1;1-\frac{\alpha}{2}}\frac{S_n}{\sqrt{n}}\right]$$

è un intervallo di confidenza per θ di livello $1-\alpha$ e che

$$\mathcal{L}(\mathbf{X}_n) = 2t_{n-1;1-\frac{\alpha}{2}}\frac{S_n}{\sqrt{n}}.$$

Si ha quindi che

$$e_n = \mathbb{E}_\theta[\mathcal{L}(\mathbf{X}_n)] = t_{n-1;1-\frac{\alpha}{2}}\frac{\mathbb{E}_\theta[S_n]}{\sqrt{n}} = t_{n-1;1-\frac{\alpha}{2}}\kappa(n,\alpha)\sigma,$$

con

$$\kappa(n,\alpha) = \frac{\sqrt{2}}{\sqrt{n-1}}\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}.$$

La determinazione esplicita di n^* non è possibile e va eseguita numericamente e richiede un valore di disegno σ_D^2 per il parametro di disturbo.

□

In modo del tutto analogo a quanto visto nel precedente esempio possiamo calcolare (in genere numericamente) la dimensione campionaria ottima, utilizzando le espressioni di e_n determinate negli esempi di questo capitolo.

5.6 Intervalli di confidenza e insiemi di verosimiglianza

Intervalli di confidenza e insiemi di verosimiglianza hanno giustificazioni e interpretazioni diverse. In molti casi esiste tuttavia un legame strutturale e formale tra le due procedure. Se consideriamo il modello $\mathrm{N}(\theta,\sigma^2)$ (varianza nota) abbiamo che, in corrispondenza di un campione osservato $\mathbf{x}_n$, gli insiemi L_q e l'intervallo di confidenza osservato di livello $1-\alpha$ sono rispettivamente uguali a

$$L_q(\mathbf{x}_n) = \left[\overline{x}_n - k_q\frac{\sigma}{\sqrt{n}}, \overline{x}_n + k_q\frac{\sigma}{\sqrt{n}}\right], \qquad k_q = \sqrt{-2\ln q}, \qquad q \in (0,1)$$

e

$$C_{1-\alpha}(\mathbf{x}_n) = \left[\overline{x}_n - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \overline{x}_n + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right], \qquad \alpha \in (0,1).$$

I due insiemi coincidono (numericamente, ma non nella interpretazione) se assumiamo

$$q = e^{-\frac{1}{2}z^2_{1-\frac{\alpha}{2}}} \quad \text{ovvero} \qquad \alpha = 2[1-\Phi(\sqrt{-2\ln q})]. \tag{5.14}$$

In questo caso la corrispondenza si ha per ogni $n \in \mathbb{N}$. Possiamo quindi dire che l'insieme di verosimiglianza aleatorio $L_q(\mathbf{X}_n)$ è un intervallo di confidenza di livello $2[1-\Phi(\sqrt{-2\ln q})]$ e che l'intervallo di confidenza osservato è un insieme di verosimiglianza di livello $q = e^{-\frac{1}{2}z^2_{1-\frac{\alpha}{2}}}$.

La reciproca legittimazione di queste procedure si ritrova, un pò più in generale (ovvero anche al di fuori del modello normale), quando possiamo determinare intervalli di confidenza asintotici e insiemi di verosimiglianza approssimati. Abbiamo infatti che per $k_q = \sqrt{-2\ln q}$ e $q \in (0,1)$.

$$\widetilde{L}_q(\mathbf{x}_n) = \left[\widehat{\theta}_{mv}(\mathbf{x}_n) - k_q\sqrt{\mathcal{I}_n(\mathbf{x}_n)^{-1}}, \widehat{\theta}_{mv}(\mathbf{x}_n) + k_q\sqrt{\mathcal{I}_n(\mathbf{x}_n)^{-1}}\right],$$

e

$$\widetilde{C}_{1-\alpha}(\mathbf{x}_n) = \left[\widehat{\theta}_{mv}(\mathbf{x}_n) - z_{1-\frac{\alpha}{2}}\sqrt{\mathcal{I}_n(\mathbf{x}_n)^{-1}}, \widehat{\theta}_{mv}(\mathbf{x}_n) + z_{1-\frac{\alpha}{2}}\sqrt{\mathcal{I}_n(\mathbf{x}_n)^{-1}}\right], \qquad \alpha \in (0,1).$$

Quindi, anche in questo caso gli intervalli di stima coincidono numericamente se tra q e α valgono le relazioni espresse in (5.14).

Capitolo 6

Inferenza frequentista: test

Il capitolo illustra la trattazione frequestista del problema di verifica delle ipotesi. Dopo la formulazione generale del problema e l'introduzione dei concetti di base (ipotesi, statistica test, errori, probabilità di errore), si esaminano partitamente il caso del confronto tra ipotesi semplici e quello tra ipotesi di cui almeno una composta. Anche in questo caso, come nei capitoli precedenti, vengono illustrati: (a) metodi per ottenere i test; (b) metodi per la valutazione dei test. Il capitolo illustra quindi tradizionale teoria dei test ottimi: test più potenti, uniformemente più potenti, testi non distorti. Viene successivamente introdotto il concetto di p-*value*, sottolineandone .legami e differenze con la teoria classica della verifica di ipotesi. Si, considerano, infine, alcuni test asintotici.

6.1 Definizioni preliminari

Dato un modello statistico $\{\mathcal{X}^n, f_n(\mathbf{x}_n;\theta), \theta \in \Theta\}$, si consideri una partizione dello spazio parametrico Θ, costituita da due sottoinsiemi $\{\Theta_0, \Theta_1\}$, dove $\Theta_1 = \Theta_0^c$ e dove quindi

$$\Theta_0 \cup \Theta_1 = \Theta \qquad \text{e} \qquad \Theta_0 \cap \Theta_1 = \emptyset.$$

In un problema di verifica di ipotesi si vuole stabilire se è più plausibile che il valore vero del parametro incognito θ appartenga a Θ_0 oppure a Θ_1. L'ipotesi che il vero valore di θ appartenga a Θ_0 viene denominata *ipotesi nulla*, mentre l'ipotesi che il vero valore di θ sia un elemento di Θ_1 è chiamata *ipotesi alternativa*.
Si possono avere ipotesi di due tipi:

1. **Ipotesi semplici**: si ha nel caso in cui si suppone che $\Theta_i = \theta_i$ (un punto di Θ). Nel caso in cui entrambe le ipotesi siano semplici si ha quindi che $\Theta = \{\theta_0, \theta_1\}$ e si parla quindi di *sistema di ipotesi semplici*:

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta = \theta_1. \tag{6.1}$$

2. **Ipotesi composte**: si hanno quando Θ_i è un sottoinsieme non puntuale di Θ. Nel caso in cui $\Theta \subseteq \mathbb{R}^1$, i più comuni tipi di ipotesi composta sono:

   (a) **Ipotesi unilaterale (one-sided)**: fissato un valore θ_0, si ha quando si pone $H_i : \theta \leq \theta_0$ oppure $H_i : \theta \geq \theta_0$ (oppure nel caso in cui le disuguaglianze sono strette). Abbiamo un *sistema di ipotesi unilaterali* quando entrambe le ipotesi sono unilaterali:

$$H_0 : \theta \leq \theta_0 \quad \text{vs.} \quad H_1 : \theta > \theta_0. \tag{6.2}$$

oppure

$$H_0 : \theta \geq \theta_0 \quad \text{vs.} \quad H_1 : \theta < \theta_0. \tag{6.3}$$

Sono comuni anche i *sistemi di ipotesi misti*, con ipotesi nulla semplice e ipotesi alternativa unilaterale. Se, ad esempio, supponiamo che $\Theta_0 = \{\theta_0\}$ e $\Theta_1 = (\theta_0, +\infty)$, abbiamo

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta > \theta_0. \tag{6.4}$$

Se invece supponiamo che $\Theta_0 = \{\theta_0\}$ e $\Theta_1 = (-\infty, \theta_0)$, abbiamo

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta < \theta_0. \tag{6.5}$$

(b) **Ipotesi bilalerali (two-sided)**: $H_i : \theta \neq \theta_0$, oppure, presi θ_1, $\theta_2 \in \theta$ (con $\theta_1 < \theta_2$),

$$H_i : \theta \leq \theta_1 \quad \text{oppure} \quad \theta \geq \theta_2. \tag{6.6}$$

Il più comune sistema di ipotesi misto con ipotesi alternativa bilaterale è il seguente:

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta \neq \theta_0. \tag{6.7}$$

Effettuare un test vuol dire scegliere tra l'ipotesi H_0 e l'ipotesi H_1, utilizzando i dati campionari. Intendiamo qui che la scelta di H_0 comporta automaticamente la non accettazione di H_1 e, viceversa, il rifiuto di H_0 equivale all'accettazione di H_1: una soltanto tra le due ipotesi viene accettata, mentre l'altra viene scartata. Quanto detto si riassume nella seguente definizione.

6.1 Definizione (Ipotesi e test). Dato il modello statistico $\{\mathcal{X}^n, f_n(\mathbf{x}_n; \theta), \theta \in \Theta\}$ e la partizione $\{\Theta_0, \Theta_1\}$ dello spazio parametrico Θ, si chiamano *ipotesi nulla* e *ipotesi alternativa* le seguenti affermazioni sul parametro incognito del modello:

$$H_0 : \theta \in \Theta_0 \quad \text{vs.} \quad H_1 : \theta \in \Theta_1.$$

Si chiama *test* (o test di ipotesi) una *regola decisionale* basata su un campione di dati per scegliere tra ipotesi nulla o ipotesi alternativa. □

La regola decisionale che costituisce un test si implementa nel modo seguente:

1. prima di effettuare l'esperimento che produce i dati $\mathbf{x}_n$, si individua una *partizione dello spazio dei campioni*, $\{A, R\}$, con

   $$A \cup R = \mathcal{X}^n, \quad \text{e} \quad A \cap R = \emptyset$$

   dove

   $$A = \{\mathbf{x}_n \in \mathcal{X}^n : \text{ accetto } H_0\}$$

   è denominata *regione di accettazione* (del test o di H_0) e

   $$R = \{\mathbf{x}_n \in \mathcal{X}^n : \text{ rifiuto } H_0\}$$

   è denominata *regione di rifiuto* o *regione critica* (del test o di H_0);

2. si estrae un campione; se tale campione proviene da A si accetta H_0, se invece proviene da R si rifiuta H_0.

Determinare un test, ovvero una regola di decisione tra H_0 e H_1, consiste quindi nell'individuare - prima ancora di avere osservato i dati - le regioni A e R. Tipicamente le regioni A e R vengono definite attraverso opportune statistiche: in base ai valori assunti da tali statistiche si accetta o si rifiuta H_0. Ad esempio, nel caso di un modello $\mathrm{N}(\theta, 1)$, in cui si è interessati a verificare una ipotesi riguardante θ, è ragionevole utilizzare la media campionaria $\overline{X}_n$ (o una sua funzione), per decidere tra H_0 e H_1. Se, sempre per esempio, le ipotesi tra cui scegliere sono $H_0 : \ \theta \leq \theta_0$ vs. $H_1 : \ \theta > \theta_1$, è ragionevole rifiutare H_0 se il valore osservato di $\overline{X}_n$ è al di sopra di una soglia (da determinare), mentre si accetta se $\overline{X}_n$ assume un valore basso.

6.2 Definizione (Statistica test). Una statistica $W : \mathcal{X}^n \to \mathcal{W}$ è una *statistica test* se esiste una partizione $\{A_W, R_W\}$ di $\mathcal{W}$ (ovvero tale che $A_W \cup R_W = \mathcal{W}$ e $A_W \cap R_W = \emptyset$) per la quale si ha che

$$W(\mathbf{x}_n) \in A_W \Rightarrow \text{ accetto } H_0$$

e

$$W(\mathbf{x}_n) \in R_W \Rightarrow \text{ rifiuto } H_0.$$

A_W e R_W sono, rispettivamente, la *regione di accettazione* e la *regione di rifiuto del test* W. □

Nel seguito indicheremo con la locuzione "test W" la procedura di scelta tra H_0 e H_1 basata sulla statistica test W.

L'idea è la seguente: un metodo (ad esempio quello del rapporto delle verosimiglianze) consente di definire le regioni A e R di accettazione e rifiuto. Attraverso successive elaborazioni matematiche, possiamo esprimere le condizioni di accettazione e rifiuto espresse da A e R attraverso condizioni espresse su W, ovvero attraverso A_W e R_W. Abbiamo cioè che

$$\mathbf{x}_n \in A \quad \Longleftrightarrow \quad W(\mathbf{x}_n) \in A_W$$

e

$$\mathbf{x}_n \in R \quad \Longleftrightarrow \quad W(\mathbf{x}_n) \in R_W.$$

6.3 Esempio Consideriamo il modello $\mathrm{N}(\theta, 1)$ e supponiamo di essere interessati alla scelta tra le ipotesi

$$\theta = 2 \quad \text{vs.} \quad \theta \neq 2.$$

In questo caso è abbastanza naturale considerare, come statistica test, una opportuna funzione di una distanza tra lo stimatore puntuale di θ e il valore ipotizzato da H_0. Ad esempio, possiamo considerare le seguenti statistiche test:

$$W_1(\mathbf{X}_n) = |\overline{X}_n - 2| \qquad \text{e} \qquad W_2(\mathbf{X}_n) = (\overline{X}_n - 2)^2$$

Le regioni di accettazione dei due test saranno quindi

$$A_1 = \{\mathbf{x}_n \in \mathcal{X}^n : \ W_1(\mathbf{x}_n) \leq k_1\} \qquad \text{e} \qquad A_2 = \{\mathbf{x}_n \in \mathcal{X}^n : \ W_2(\mathbf{x}_n) \leq k_2\}.$$

L'idea è quindi, in entrambi i casi, di accettare l'ipotesi nulla se la discrepanza tra il valore osservato $\overline{x}_n$ e il valore ipotizzato 2 è al di sotto di una soglia, k_i. Nel caso del secondo test, ad esempio, possiamo esprimere la regione di accettazione come sottoinsieme dei numeri reali:

$$A_{W_2} = \{(\overline{x}_n - 2) \in \mathbb{R} : \ (\overline{x}_n - 2) \in [-\sqrt{k_2}, \sqrt{k_2}]\}.$$

Abbiamo quindi che $A_{W_2} \subseteq \mathbb{R}^1$.

Si noti che, per rendere operativi i test considerati, è necessario fissare il valore della soglia k_i, $i = 1, 2$. □

Dal precedente semplice esempio possiamo già dedurre due fatti importanti:

- per poter utilizzare un test W è necessario determinare i valori delle soglie che definiscono le regioni A_W e R_W (nell'esempio le soglie sono i valori k_i).
- per lo stesso sistema di ipotesi esiste più di un test (ovvero più di una procedura per scegliere tra le stesse ipotesi).

Errori associati a un test

Come vedremo, per affrontare entrambi i suddetti problemi, l'inferenza statistica frequentista tiene conto degli errori che si possono commettere utilizzando un test W. Infatti, quando si utilizza una qualsiasi statistica test W per scegliere tra le due ipotesi H_0 e H_1 è possibile commettere due tipi di errore:

1. Errore di I tipo: rifiutare H_0 quando $\theta \in \Theta_0$, ovvero quando H_0 è l'ipotesi vera;
2. Errore di II tipo: rifiutare H_1 quando $\theta \in \Theta_1$, ovvero quando H_1 è l'ipotesi vera;

La tabella riporta schematicamente le conseguenze corrispondenti alle scelte effettuate (a favore di H_0 o di H_1) a seconda che l'ipotesi vera sia H_0 oppure H_1.

	si accetta H_0	si rifiuta H_0
H_0 vera	ok	errore I tipo
H_1 vera	errore II tipo	ok

La valutazione di un test W dipende dalle probabilità che, utilizzandolo, si commettano i due suddetti errori. Quindi, in conformità al principio del campionamento ripetuto, un test viene valutato in base alle caratteristiche probabilistiche della statistica test $W(\mathbf{X}_n)$, (ovvero della distribuzione campionaria $f_W(\cdot;\theta)$). In particolare, come vedremo meglio nel seguito, la determinazione esplicita della regione di accettazione del test W si effettua proprio imponendo un vincolo sulla probabilità di rifiutare H_0 quando H_0 è l'ipotesi corretta, e la scelta tra test alternativi per lo stesso sistema di ipotesi si effettua in funzione della probabilità che il test ha di rifiutare l'ipotesi H_0, quando questa è falsa.

A questo punto, solo per motivi didattici, separiamo la trattazione dei test per sistemi di ipotesi semplici da quella per ipotesi miste e composte.

6.2 Ipotesi semplici

Nel caso di scelta tra ipotesi semplici, gli insiemi Θ_0 e Θ_1 coincidono con due valori θ_0 e θ_1. Dato un test statistico basato sulle regioni (A, R) per la scelta tra le due ipotesi, possiamo definire la probabilità di commettere i due tipi di errore. Nel seguito, per un qualsiasi sottoinsieme B di $\mathcal{X}^n$, indicheremo con

$$\mathbb{P}_{\theta_i}(B) = \mathbb{P}_{\theta_i}(\mathbf{X}_n \in B)$$

la probabilità che la v.a. $\mathbf{X}_n$ appartenga al sottoinsieme B di $\mathcal{X}^n$, calcolata con la distribuzione $f_n(\cdot;\theta_i), \quad i = 0, 1$. Nel caso di v.a. assolutamente continue abbiamo che

$$\mathbb{P}_{\theta_i}(B) = \int_B f_{X_1,\ldots,X_n}(x_1,\ldots,x_n;\theta_i)dx_1\ldots dx_n = \int_B f_n(\mathbf{x}_n;\theta_i)d\mathbf{x}_n.$$

6.4 Definizione (Probabilità di errore e potenza di un test). Dato il sistema di sistema di ipotesi semplici

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta = \theta_1,$$

e un test con regioni di accettazione e rifiuto (A, R) la *probabilità di errore di I tipo*[1] (o di I specie) è

$$\alpha = \mathbb{P}_{\theta_0}(R) = \mathbb{P}_{\theta_0}(\mathbf{X}_n \in R);$$

la *probabilità di errore di II tipo* (o di II specie) è

$$\beta = \mathbb{P}_{\theta_1}(A) = \mathbb{P}_{\theta_1}(\mathbf{X}_n \in A).$$

La probabilità di rifiutare l'ipotesi nulla quando questa è falsa

$$\mathbb{P}_{\theta_1}(R) = 1 - \beta$$

si chiama *potenza del test*. □

6.5 Osservazione Nel caso in cui le X_i sono v.a. assolutamente continue si ha che

$$\alpha = \int_R f_n(\mathbf{x}_n; \theta_0) d\mathbf{x}_n$$

e

$$\beta = \int_A f_n(\mathbf{x}_n; \theta_1) d\mathbf{x}_n.$$

La potenza del test è invece

$$1 - \beta = \mathbb{P}_{\theta_1}(R) = \int_R f_n(\mathbf{x}_n; \theta_1) d\mathbf{x}_n.$$

Inoltre, se A e R si possono esprimere in funzione di una statistica W e se A_W e R_W sono le corrispondenti regioni di accettazione e di rifiuto, abbiamo che

$$\alpha = \mathbb{P}^W_{\theta_0}(R_W) = \mathbb{P}_{\theta_0}[W(\mathbf{X}_n) \in R_W]$$

e

$$\beta = \mathbb{P}^W_{\theta_1}(A_W) = \mathbb{P}_{\theta_1}[W(\mathbf{X}_n) \in A_W].$$

Infine, nel caso in cui le X_i sono v.a. assolutamente continue e $W(\mathbf{X}_n)$ è una v.a. a valori in $\mathbb{R}^1$ con densità $f_W(\cdot; \theta)$, si ha che

$$\alpha = \int_R f_n(\mathbf{x}_n; \theta_0) d\mathbf{x}_n = \int_{R_W} f_W(w; \theta_0) dw,$$

e

$$\beta = \int_A f_n(\mathbf{x}_n; \theta_1) d\mathbf{x}_n = \int_{A_W} f_W(w; \theta_1) dw.$$

La potenza del test W (probabilità di rifiutare l'ipotesi nulla quando questa è falsa) è invece

$$1 - \beta = \mathbb{P}_{\theta_1}(R) = \int_R f_n(\mathbf{x}_n; \theta_1) d\mathbf{x}_n = \int_{R_W} f_W(w; \theta_1) dw = \mathbb{P}^W_{\theta_1}(R_W).$$

□

[1] La probabilità di errore di I tipo si indica anche con il termine *ampiezza* del test che, come vedremo, ha una accezione più generale nel caso di test con ipotesi nulla composta.

Implementazione di un test per il confronto tra ipotesi semplici

Un generico test W per il confronto tra ipotesi semplici si implementa nel modo seguente:

- si individuano le *generiche* regioni di accettazione (A_W) e di rifiuto (R_W) del test;
- si individuano le *specifiche* regioni di accettazione (A_W) e di rifiuto (R_W) del test imponendo che il test abbia probabilità di errore di I tipo pari a un valore scelto $\alpha \in (0,1)$;
- si estrae il campione $\mathbf{x}_n$ e si sceglie tra H_0 e H_1 a seconda del valore assunto da $W(\mathbf{x}_n)$: se $W(\mathbf{x}_n) \in A_W$ si accetta l'ipotesi H_0, altrimenti la si rifiuta.

Il test più comunemente utilizzato per la scelta tra ipotesi semplici è il test del rapporto delle verosimiglianze che, come vedremo, gode anche di proprietà ottimali.

6.6 Definizione (Test del rapporto delle verosimiglianze). Dato il modello statistico $\{\mathcal{X}^n, f_n(\mathbf{x}_n;\theta), \theta \in \Theta\}$, un campione $x_1, \ldots, x_n$ e il sistema di ipotesi $H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta = \theta_1$,

1. il *test del rapporto delle verosimiglianze* (test rv) si basa sulle regioni di accettazione e rifiuto A e R definite da

$$A = \{\mathbf{x}_n \in \mathcal{X}^n : \lambda_{01}(\mathbf{x}_n) > k\}$$

e

$$R = \{\mathbf{x}_n \in \mathcal{X}^n : \lambda_{01}(\mathbf{x}_n) \leq k\}$$

dove

$$\lambda_{01}(\mathbf{x}_n) = \frac{L(\theta_0;\mathbf{x}_n)}{L(\theta_1;\mathbf{x}_n)},$$

$L(\cdot;\mathbf{x}_n)$ è la funzione di verosimiglianza di θ associata al campione $\mathbf{x}_n$ e dove k è una costante positiva;

2. fissato un valore di α in $(0,1)$, il test del rv con probabilità di errore di I tipo (o *ampiezza*) pari a α si ottiene ponendo $k = k_\alpha$ in A, con k_α tale che

$$\mathbb{P}_{\theta_0}(R) = \mathbb{P}_{\theta_0}(\lambda_{01}(\mathbf{X}_n) \leq k_\alpha) = \alpha.$$

□

La logica di questo test è abbastanza chiara: si accetta H_0 se l'evidenza fornita dal campione a favore di H_0 rispetto a quella a favore di H_1 è sufficientemente elevata, ovvero superiore a un valore fissato k. Il valore di k nel punto (1) della precedente definizione è del tutto generico. Per rendere operativo il test si richiede che il test abbia una precisa garanzia probabilistica, ovvero che la probabilità di errore di prima specie pari a α, con α valore scelto in $(0,1)$ (in genere un valore ben inferiore a 0.5). Fiisando α viene determinato il valore di k, che indichiamo con k_α.

6.7 Osservazione. Si noti la differenza tra la procedura di test qui proposta e quella considerata nel Capitolo 3. La statistica utilizzata per effettuare il test è la stessa, ma l'implementazione della procedura è completamente diversa. Nel caso del Capitolo 3, λ_{01} misura l'evidenza sperimentale che il campione osservato fornisce a favore di H_0 contro H_1. La soglia k è quindi in genere un valore maggiore di uno. Il test è del tutto post-sperimentale, in linea con l'analisi basata sulla funzione di verosimiglianza sviluppata in quel capitolo. Nel caso del test frequentista, la soglia k si fissa pre-sperimentalmente, in modo tale che la probabilità di errore di I specie sia pari a un valore α scelto da noi. Osservato il

campione $\mathbf{x}_n$, il valore numerico della statistica test λ_{01} viene utilizzato esclusivamente per decidere a favore di H_0 (se $\lambda_{01}(\mathbf{x}_n) \in A$) o di H_1 (se $\lambda_{01}(\mathbf{x}_n) \in R$) e non per quantificare l'evidenza a favore delle ipotesi. È inoltre possibile che la soglia k_α sia inferiore a 1 e che si scelga a favore di H_0 anche se $\lambda_{01}(\mathbf{x}_n) < 1$. Si tratta di un caso esemplare di conflitto tra procedure conformi al principio di verosimiglianza e procedure conformi al principio del campionamento ripetuto. □

6.8 Esempio (Modello normale). Sia $X_1, \ldots, X_n$ un campione casuale da $\mathrm{N}(\theta, \sigma^2)$, con σ^2 noto. Consideriamo il sistema di ipotesi

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta = \theta_1.$$

Ricordando che $L(\theta; \mathbf{x}_n) \propto \exp\{-\frac{n}{2\sigma^2}(\theta - \overline{x}_n)^2\}$, si ottiene

$$\lambda_{01}(\mathbf{x}_n) = \frac{L(\theta_0; \mathbf{x}_n)}{L(\theta_1; \mathbf{x}_n)} = \exp\left\{-\frac{n}{2\sigma^2}\left[(\theta_0 - \overline{x}_n)^2 - (\theta_1 - \overline{x}_n)^2\right]\right\}.$$

Il test del rv ha quindi regione di accettazione

$$A = \left\{\mathbf{x}_n \in \mathcal{X}^n : \lambda_{01}(\mathbf{x}_n) = \exp\left\{-\frac{n}{2\sigma^2}\left[(\theta_0 - \overline{x}_n)^2 - (\theta_1 - \overline{x}_n)^2\right]\right\} > k\right\}$$

Osserviamo che la statistica test dipende dai dati esclusivamente attraverso la media campionaria $\overline{x}_n$. Possiamo quindi esprimere la regione di accettazione del test in funzione di questa statistica. Con semplici calcoli si mostra infatti che:

$$\begin{aligned} A &= \{\mathbf{x}_n \in \mathcal{X}^n : \lambda_{01}(\mathbf{x}_n) > k\} \\ &= \left\{\mathbf{x}_n \in \mathcal{X}^n : \exp\left\{-\frac{n}{2\sigma^2}[2\overline{x}_n(\theta_1 - \theta_0) + (\theta_0^2 - \theta_1^2)]\right\} > k\right\} \\ &= \left\{\mathbf{x}_n \in \mathcal{X}^n : \exp\left\{-\frac{n}{\sigma^2}\overline{x}_n(\theta_1 - \theta_0)\right\} > k'\right\} \end{aligned}$$

dove l'ultima uguaglianza si ottiene ponendo $k' = k \times \exp\{n(\theta_0^2 - \theta_1^2)/2\sigma^2\}$. Abbiamo quindi che

$$A = \left\{\mathbf{x}_n \in \mathcal{X}^n : -\overline{x}_n(\theta_1 - \theta_0) > \frac{\sigma^2}{n}\ln k'\right\} \tag{6.8}$$

A questo punto è necessario distinguere due casi: $\theta_1 > \theta_0$ oppure $\theta_1 < \theta_0$.

- I caso: $\theta_1 > \theta_0$. In questo caso la regione A diventa

  $$A = \{\mathbf{x}_n \in \mathcal{X}^n : \overline{x}_n < k''\}$$

  dove $k'' = -\sigma^2 \ln(k')/[n(\theta_1 - \theta_0)]$. Per determinare il valore di k'' tale che il test abbia probabilità di errore di I tipo pari a α osserviamo che[2]

  $$\begin{aligned} \mathbb{P}_{\theta_0}(R) = \alpha &\Leftrightarrow \mathbb{P}_{\theta_0}\left(\lambda_{01}(\mathbf{X}_n) \le k\right) = \alpha \\ &\Leftrightarrow \mathbb{P}_{\theta_0}\left(\overline{X}_n \ge k''\right) = \alpha \\ &\Leftrightarrow \mathbb{P}\left(\frac{\sqrt{n}(\mathbf{X}_n - \theta_0)}{\sigma} \ge \frac{\sqrt{n}(k'' - \theta_0)}{\sigma}\right) = \alpha \\ &\Leftrightarrow \frac{\sqrt{n}(k'' - \theta_0)}{\sigma} = z_{1-\alpha} \\ &\Leftrightarrow k'' = \theta_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha}. \end{aligned}$$

[2] Qui e nel seguito assumeremo che il valore α sia sempre (nettamente) inferiore a 0.5 e quindi che $z_{1-\alpha} > 0$ e $z_\alpha = -z_{1-\alpha} < 0$. La stessa considerazione vale quando, in altri esempi, si considerazioni si usano percentili di distribuzioni diverse dalla normale.

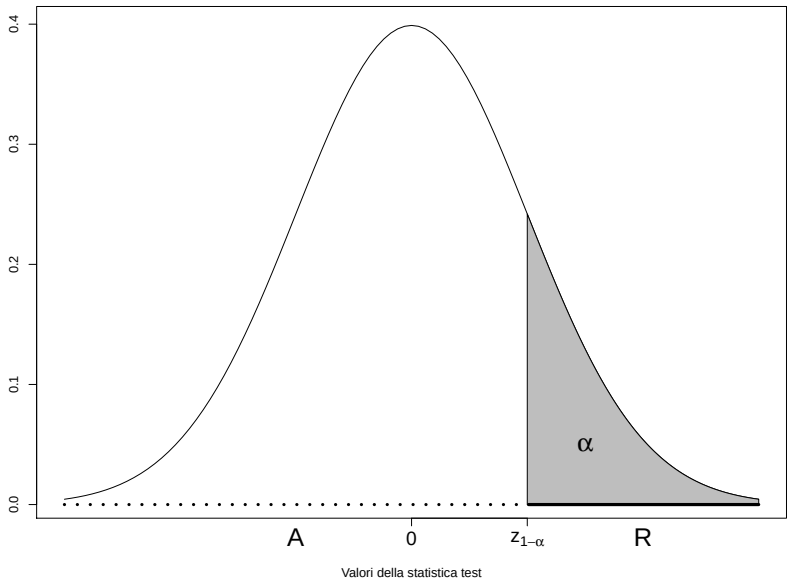

Figura 6.1: `Esempio 6.8, caso` $\theta_1 > \theta_0$. `Rappresentazione della funzione di densità della statistica test` $W(\mathbf{X}_n) = \frac{\sqrt{n}(\overline{X}_n - \theta_0)}{\sigma}$ `che, supponendo vera` H_0, `ha distribuzione` $\mathrm{N}(0,1)$. `Sull'asse delle ascisse si riportanno i valori che può assumere la statistica test e le regioni` A `ed` R.

La regione di accettazione del test di ampiezza α è quindi

$$A_{\overline{X}_n} = \left\{ \mathbf{x}_n \in \mathcal{X}^n : \overline{x}_n < \theta_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha} \right\}.$$

Possiamo esprimere la condizione di accettazione di H_0 in funzione della statistica test

$$W(\mathbf{x}_n) = \frac{\sqrt{n}(\overline{x}_n - \theta_0)}{\sigma},$$

ovvero

$$A_W = \left\{ \overline{x}_n \in \mathbb{R} : \frac{\sqrt{n}(\overline{x}_n - \theta_0)}{\sigma} < z_{1-\alpha} \right\}.$$

Per il test in esame possiamo considerare due rappresentazioni grafiche.

Figura 6.1: riporta il grafico della funzione di densità $\mathrm{N}(0,1)$ della statistica test $W(\mathbf{X}_n)$ sotto l'ipotesi H_0 (ovvero quando si ipotizza che $\theta = \theta_0$). Sull'asse delle ascisse si rappresentano i valori $W(\mathbf{x}_n)$ che la statistica test può assumere. La regione di accettazione A di H_0 coincide con la semiretta $(-\infty, z_{1-\alpha}]$; la regione di rifiuto R con la semiretta $(z_{1-\alpha}, \infty)$. L'area in grigio rappresenta la probabilità di errore di I specie associata al test.

Figura 6.2: rappresenta la funzione di densità della media campionaria $\overline{X}_n$ sotto ipotesi $H_0 : \theta = \theta_0$ e sotto ipotesi $H_1 : \theta = \theta_1$ (caso $\theta_1 > \theta_0$) e le regioni A e R. L'area in grigio rappresenta la probabilità di errore di I specie α; l'area con tratteggio verticale la probabilità di errore di II specie β, l'area con tratteggio orizzontale la potenza del test $1 - \beta$.

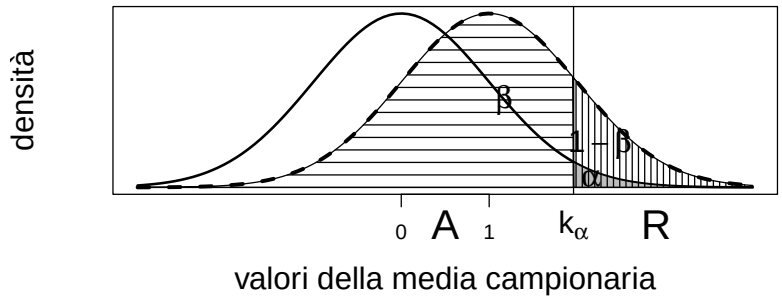

Figura 6.2: `Esempio 6.8, caso` $\theta_1 > \theta_0$. `Rappresentazione delle funzioni di densità della statistica test` $\overline{X}_n$ `sotto ipotesi` $H_0 : \theta = \theta_0$ `(a sinistra) e sotto ipotesi` $H_1 : \theta = \theta_1$ `(a destra) (caso` $\theta_1 > \theta_0$`) e delle regioni` A `e` R. `L'area in grigio rappresenta la probabilità di errore di I specie` α; `l'area con tratteggio verticale la probabilità di errore di II specie` β, `l'area con tratteggio orizzontale la potenza del test` $1 - \beta$.

`Esempio numerico`. Supponiamo che $\theta_0 = 368$, $n = 25$, $\sigma = 15$. Scegliendo per la probabilità di errore di I tipo il valore $\alpha = 0.05$, si ha che $z_{1-\alpha} = z_{0.95} = \texttt{qnorm}(0.95) = 1.645$. La regione di accettazione del test diventa quindi:

$$A_{\overline{X}_n} = \left\{ \overline{x}_n \in \mathbb{R} : \ \frac{\sqrt{25}(\overline{x}_n - 368)}{15} \leq 1.645 \right\}.$$

Supponiamo ora di osservare un campione di 25 osservazioni per il quale si ha $\overline{x}_n = 372$. In questo caso $\sqrt{n}(\overline{x}_n - \theta_0)/\sigma = 1.33 < 1.645 = z_{1-\alpha}$ e quindi accettiamo H_0.

- II caso: $\theta_1 < \theta_0$. Nel caso in cui $\theta_1 < \theta_0$, dalla (6.8) si ottiene che la regione di accettazione del test diventa

$$A = \{ \mathbf{x}_n \in \mathcal{X}^n : \ \overline{x}_n \geq k \}$$

Con passaggi analoghi a quelli del caso precedente, impoinendo che il test abbia ampiezza α si ottiene che

$$A_{\overline{X}_n} = \left\{ \overline{x}_n \in \mathbb{R} : \ \frac{\sqrt{n}(\overline{x}_n - \theta_0)}{\sigma} > -z_{1-\alpha} = z_\alpha \right\}.$$

La Figura 6.3 (caso $\theta_1 < \theta_0$) riporta il grafico della funzione di densità $\mathrm{N}(0, 1)$ della statistica test $W(\mathbf{X}_n)$ sotto H_0 (ovvero quando si ipotizza che $\theta = \theta_0$). Sull'asse delle ascisse si rappresentano i valori $W(\mathbf{x}_n)$ che la statistica test può assumere. La regione di rifiuto R di H_0 coincide con la semiretta $(-\infty, z_\alpha]$; la regione di accettazione A con la semiretta (z_α, ∞). L'area in grigio rappresenta la probabilità di errore di I specie associata al test.

□

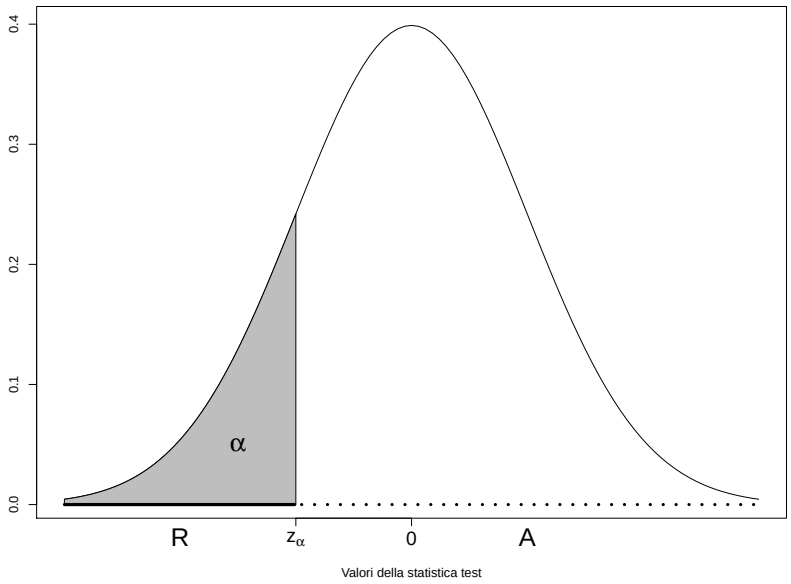

Figura 6.3: *Rappresentazione di A ed R per l'esempio 6.8, caso* $\theta_1 < \theta_0$.

Test del rapporto delle verosimiglianze e statistiche sufficienti

Nel precedente esempio abbiamo visto che le regioni di accettazione e rifiuto del test del rv si possono esprimere attraverso la statistica $\overline{X}_n$ che, nel modello $\mathrm{N}(\theta, \sigma^2)$ è sufficiente minimale. Più in generale vale il seguente terorema.

6.9 Teorema (Test basati su statistiche sufficienti). Si consideri il modello statistico $\{\mathcal{X}^n, f_n(\mathbf{x}_n; \theta), \theta \in \Theta\}$, con $\Theta \subseteq \mathbb{R}$, il sistema di ipotesi semplici

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta = \theta_1.$$

e il test rv. Si supponga inoltre che il modello ammetta statistica sufficiente $T : \mathcal{X}^n \to \mathcal{T} \subseteq \mathbb{R}$ tale che $\lambda_{01}(\mathbf{x}_n) = \varphi(T(\mathbf{x}_n))$, con φ funzione monotona. Si ha allora quanto segue.

- Se φ è una funzione non decrescente di T, allora

$$A = \{\mathbf{x}_n \in \mathcal{X}^n : \lambda_{01}(\mathbf{x}_n) > k\} = \{\mathbf{x}_n \in \mathcal{X}^n : T(\mathbf{x}_n) > k'\},$$

  dove $k' = \varphi^{-1}(k)$ e φ^{-1} è la funzione inversa di φ.

- Se φ è una funzione non crescente di T, allora

$$A = \{\mathbf{x}_n \in \mathcal{X}^n : \lambda_{01}(\mathbf{x}_n) > k\} = \{\mathbf{x}_n \in \mathcal{X}^n : T(\mathbf{x}_n) < k'\}.$$

□

Dimostrazione. Dal criterio di fattorizzazione discende che

$$\lambda_{01}(\mathbf{x}_n) = \frac{L(\theta_0; \mathbf{x}_n)}{L(\theta_1; \mathbf{x}_n)} = \frac{h(\mathbf{x}_n)}{h(\mathbf{x}_n)} \frac{g(T(\mathbf{x}_n), \theta_0)}{g(T(\mathbf{x}_n), \theta_1)} = \frac{g(T(\mathbf{x}_n), \theta_0)}{g(T(\mathbf{x}_n), \theta_1)}.$$

Sia

$$\varphi(T(\mathbf{x}_n)) = \frac{g(T(\mathbf{x}_n), \theta_0)}{g(T(\mathbf{x}_n), \theta_1)}.$$

Si ha allora che

$$\lambda_{01}(\mathbf{x}_n) > k \qquad \Leftrightarrow \qquad \varphi(T(\mathbf{x}_n)) > k.$$

Se φ è monotona non decrescente, si ha

$$\varphi(T(\mathbf{x}_n)) > k \qquad \Leftrightarrow \qquad T(\mathbf{x}_n) > \varphi^{-1}(k) = k';$$

se invece φ è monotona non crescente, allora

$$\varphi(T(\mathbf{x}_n)) > k \qquad \Leftrightarrow \qquad T(\mathbf{x}_n) < \varphi^{-1}(k) = k'.$$

6.10 Esempio (Modello esponenziale negativo). Sia $X_1, \ldots, X_n$ un campione casuale con distribuzione EN(θ). Si consideri il sistema di ipotesi

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta = \theta_1.$$

Il rapporto delle verosimiglianze è

$$\lambda_{01}(\mathbf{x}_n) = \frac{\theta_0^n e^{-\theta_0 n \overline{x}_n}}{\theta_1^n e^{-\theta_1 n \overline{x}_n}} = \left(\frac{\theta_0}{\theta_1}\right)^n e^{-(\theta_0 - \theta_1) n \overline{x}_n}.$$

Consideriamo i due seguenti possibili casi.

- I caso: $\theta_0 < \theta_1$. La statistica λ_{01} è funzione non decrescente di $\overline{x}_n$. Pertanto

$$A = \{\mathbf{x}_n \in \mathcal{X}^n : \overline{x}_n > k'\} = \left\{\mathbf{x}_n \in \mathcal{X}^n : \frac{1}{\overline{x}_n} < k''\right\}, \tag{6.9}$$

  con $k'' = 1/k'$. Si noti che $1/\overline{x}_n$ è la stima di massima verosimiglianza di θ. Risulta quindi sensato accettare H_0 se $1/\overline{x}_n < k$, dal momento che $\theta_0 < \theta_1$. Ovviamente si ottiene lo stesso risultato svolgendo tutti i calcoli, senza cioè riconoscere che λ_{01} è funzione monotona crescente di $\overline{x}_n$. Infatti

$$\begin{aligned}
\lambda_{01}(\overline{x}_n) > k \quad &\Leftrightarrow \quad e^{-(\theta_0 - \theta_1) n \overline{x}_n} > \left(\frac{\theta_1}{\theta_0}\right)^n k = k_1 \\
&\Leftrightarrow \quad -(\theta_0 - \theta_1) n \overline{x}_n > \ln k_1 \\
&\Leftrightarrow \quad \overline{x}_n > \frac{\ln k_1}{n(\theta_1 - \theta_0)} = k_2 \\
&\Leftrightarrow \quad \sum_{i=1}^n x_i > k_3.
\end{aligned}$$

  Dalle precedenti relazioni si ottiene anche che (per l'arbitrarietà di k) possiamo esprimere nel modo seguente la regione di rifiuto del test:

$$R = \left\{\mathbf{x}_n \in \mathcal{X}^n : \sum_{i=1}^n x_i \leq k\right\}. \tag{6.10}$$

  Per fissare il valore k in (6.10), imponiamo che la probabilità di errore di I tipo sia pari ad $\alpha \in (0, 1)$. Ricordiamo che, nel caso del modello EN(θ), $\sum_{i=1}^n X_i \sim \text{Ga}(n, \texttt{rate} = \theta)$. Pertanto,

$$\mathbb{P}_{\theta_0}(R) = \mathbb{P}_{\theta_0}\left(\sum_{i=1}^n X_i \leq k\right) = \alpha \Leftrightarrow k = q_\alpha(n, \theta_0)$$

  dove $q_\alpha(n, \theta_0)$ indica il percentile[3] di livello α di una v.a. Ga($n, \texttt{rate} = \theta_0$).

[3] Con `R` il valore di $q_\alpha(n, \theta_0)$ si trova con la funzione qgamma($\alpha, \texttt{shape} = n, \texttt{rate} = \theta_0$).

- II caso: $\theta_0 > \theta_1$. In modo del tutto analogo, essendo λ_{01} funzione non crescente di $\overline{x}_n$ e quindi di $\sum_{i=1}^n x_i$, si mostra che

$$R = \left\{ \mathbf{x}_n \in \mathcal{X}^n : \sum_{i=1}^n x_i \geq k \right\}. \tag{6.11}$$

  Analogamente a quanto visto prima, il test di ampiezza α si ottiene ponendo in (6.11)

$$\mathbb{P}_{\theta_0}(R) = \mathbb{P}_{\theta_0}\left(\sum_{i=1}^n X_i \geq k \right) = \alpha \Leftrightarrow k = q_{1-\alpha}(n, \theta_0).$$

□

6.11 Esempio (Modello bernoulliano). Consideriamo un campione casuale da un modello $\text{Ber}(\theta)$ e un sistema di ipotesi semplici. In questo caso il rapporto delle verosimiglianze è funzione della statistica sufficiente $y_n = \sum_{i=1}^n x_i$:

$$\lambda_{01}(\mathbf{x}_n) = \frac{\theta_0^{y_n}}{\theta_1^{y_n}} \frac{(1-\theta_0)^{n-y_n}}{(1-\theta_1)^{n-y_n}} = \left(\frac{\theta_0}{\theta_1} \right)^{y_n} \left(\frac{1-\theta_0}{1-\theta_1} \right)^{n-y_n}.$$

- I caso: $\theta_0 > \theta_1$. La precedente espressione è il prodotto di due funzioni non decrescenti di y_n e quindi

$$A = \{ \mathbf{x}_n \in \mathcal{X}^n : \ y_n > k' \}.$$

  Il test di ampiezza α si ottiene imponendo che

$$\mathbb{P}_{\theta_0}(R_{\overline{x}_n}) = \mathbb{P}_{\theta_0}(Y_n \leq k') = \alpha \Leftrightarrow k' = q_\alpha(n, \theta_0)$$

  dove $q_\alpha(n, \theta_0)$ è il percentile di livello α di una v.a. $\text{Bin}(n, \theta_0)$.

- II caso: $\theta_0 < \theta_1$. La precedente espressione è il prodotto di due funzioni non crescenti di y_n e quindi

$$A = \{ \mathbf{x}_n \in \mathcal{X}^n : \ y_n < k' \}.$$

  Il test di ampiezza α si ottiene imponendo che

$$\mathbb{P}_{\theta_0}(R_{\overline{x}_n}) = \mathbb{P}_{\theta_0}(Y_n \geq k') = \alpha \Leftrightarrow k' = q_{1-\alpha}(n, \theta_0)$$

  dove $q_{1-\alpha}(n, \theta_0)$ è il percentile di livello $1-\alpha$ di una v.a. $\text{Bin}(n, \theta_0)$.

Si osservi che in questo caso, essendo la v.a. Y_n una $\text{Bin}(n, \theta_0)$ (discreta), il valore di α non può essere un qualsiasi valore in $(0, 1)$, come avviene invece nel caso di statistiche test che sono v.a. assolutamente continue. □

6.2.1 Ottimalità: Lemma di Neyman e Pearson

I test classici si implementano attraverso il controllo della probabilità di errore di I specie, ovvero del comportamento della statistica test quando $\theta = \theta_0$. Tuttavia il confronto tra test diversi deve tenere conto anche del comportamento del test nel caso in cui $\theta = \theta_1$. È evidente che tra due test con uguale probabilità di errore di I specie è preferibile quello per il quale la probabilità di errore di II specie è più bassa e che il test ideale è quello per il quale sono minime sia α che β. Tuttavia, le probabilità di errore non sono indipendenti tra loro: quanto più basso è il valore di una, tanto più elevato è il valore dell'altra. Informalmente, il ragionamento è il seguente. Per un test con regioni di accettazione e rifiuto A_W e R_W,

affinchè il valore di $\alpha = \mathbb{P}_{\theta_0}(R_W)$ sia piccolo, è necessario che sia piccola la dimensione di R_W. Ma poichè, $A_W = R_W^c$, ciò vuol dire che la dimensione di A_W è grande. Ma quanto più grande è la dimensione di A_W, tanto più alto è il valore di $\beta = \mathbb{P}_{\theta_1}(A_W)$, ovvero la probabilità di errore di II tipo. È quindi necessario trovare un compromesso tra α e β. Si procede allora in questo modo: si impone un limite superiore al valore di α e si cerca il test che minimizza β, ovvero per il quale la potenza $(1 - \beta)$ è massima.

6.12 Definizione (Test più potente di ampiezza fissata). Dato un modello statistico e un sistema di ipotesi semplici, si chiama *test più potente di ampiezza* α un test (A^*, R^*) che ha probabilità di errore di primo tipo pari a α e tale che, per ogni altro test (A', R') con probabilità di errore di I tipo pari a $\alpha' \leq \alpha$, si ha

$$1 - \beta^* = \mathbb{P}_{\theta_1}(R^*) \geq \mathbb{P}_{\theta_1}(R') = 1 - \beta'.$$

□

Il seguente fondamentale teorema mostra che, per il confronto tra ipotesi semplici, il test rv di ampiezza α è il test più potente tra tutti i test di livello α.

6.13 Teorema (Lemma di Neyman e Pearson). Dato il sistema di ipotesi semplici $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$, il test di ampiezza α e potenza $1 - \beta$, basato sulla regione di rifiuto

$$R_\lambda = \{\mathbf{x}_n \in \mathcal{X}^n : \lambda_{01}(\mathbf{x}_n) \leq k\}, \tag{6.12}$$

(dove λ_{01} indica il rapporto delle verosomiglianze e k è un opportuno valore positivo che garantisce l'ampiezza fissata) è tale che, per ogni altro test (A', R') di ampiezza $\alpha' \leq \alpha$ e potenza $1 - \beta'$, si ha

$$1 - \beta \geq 1 - \beta'.$$

□

Dimostrazione. Consideriamo un qualsiasi test con regione di rifiuto R' di ampiezza $\alpha' \leq \alpha$. Dobbiamo dimostrare che

$$1 - \beta = \int_{R_\lambda} f_n(\mathbf{x}_n; \theta_1) d\mathbf{x}_n \geq \int_{R'} f_n(\mathbf{x}_n; \theta_1) d\mathbf{x}_n = 1 - \beta'.$$

A tal fine consideriamo l'integrale

$$J = \int_{\mathcal{X}^n} (1_{R_\lambda}(\mathbf{x}_n) - 1_{R'}(\mathbf{x}_n)) \left(f_n(\mathbf{x}_n; \theta_1) - \frac{1}{k} f_n(\mathbf{x}_n; \theta_0) \right) d\mathbf{x}_n.$$

Osservando quanto segue:

$$R_\lambda = \left\{ \mathbf{x}_n \in \mathcal{X}^n : f_n(\mathbf{x}_n; \theta_1) - \frac{1}{k} f_n(\mathbf{x}_n; \theta_0) \geq 0 \right\}$$

si verifica facilmente che

- se $\mathbf{x}_n \in R_\lambda$, i due fattori delle funzione integranda in J sono entrambi non negativi;
- se $\mathbf{x}_n \notin R_\lambda$, i due fattori delle funzione integranda in J sono entrambi non positivi.

Pertanto $J \geq 0$ in corrispondenza di qualunque campione in $\mathcal{X}^n$. Svolgendo il prodotto interno all'integrale si ottiene che

$$\begin{aligned} J &= \int_{R_\lambda} f_n(\mathbf{x}_n;\theta_1)d\mathbf{x}_n - \int_{R'} f_n(\mathbf{x}_n;\theta_1)d\mathbf{x}_n - \frac{1}{k}\left(\int_{R_\lambda} f_n(\mathbf{x}_n;\theta_0)d\mathbf{x}_n - \int_{R'} f_n(\mathbf{x}_n;\theta_0)d\mathbf{x}_n\right) \\ &= (1-\beta) - (1-\beta') - \frac{1}{k}(\alpha - \alpha') \geq 0 \end{aligned}$$

Pertanto, avendo assunto che $\alpha' \leq \alpha$, si ha che $(\alpha - \alpha')/k \geq 0$ e quindi che

$$(1-\beta) \geq (1-\beta').$$

6.14 Esempio (Modello normale: potenza del test) Riprendiamo in considerazione l'Esempio 6.8 e calcoliamo la potenza del test del rv di ampiezza α (sempre assumendo $\theta_0 < \theta_1$). Per quanto ottenuto e ricordando che, sotto l'ipotesi H_1 si ha

$$\overline{X}_n|\theta_1 \sim \mathrm{N}\left(\theta_1, \frac{\sigma^2}{n}\right) \qquad \text{e} \qquad Z = \frac{\sqrt{n}(\overline{X}_n - \theta_1)}{\sigma} \sim \mathrm{N}(0,1),$$

abbiamo allora che

$$\begin{aligned} 1-\beta = \mathbb{P}_{\theta_1}(R_\lambda) &= \mathbb{P}_{\theta_1}\left\{\overline{X}_n > \theta_0 + z_{1-\alpha}\frac{\sigma}{\sqrt{n}}\right\} \\ &= \mathbb{P}\left(\frac{\sqrt{n}(\overline{X}_n - \theta_1)}{\sigma} > \frac{\sqrt{n}(\theta_0 - \theta_1)}{\sigma} + z_{1-\alpha}\right) \\ &= 1 - \Phi\left(-\frac{\sqrt{n}(\theta_1 - \theta_0)}{\sigma} + z_{1-\alpha}\right). \end{aligned}$$

Si osservi che la potenza $1-\beta$ è

- funzione crescente (tendente al limite a 1) della dimensione campionaria n, per valori fissati di α, σ e $(\theta_1 - \theta_0)$;
- funzione crescente (tendente al limite a 1) di α, per valori fissati di n, σ e $(\theta_1 - \theta_0)$;
- funzione crescente (tendente al limite a 1) di $(\theta_1 - \theta_0)$, per valori fissati di α, σ e n;
- funzione decrescente (tendente al limite ad α) di σ, per valori fissati di n, α e $(\theta_1 - \theta_0)$.

Il lemma di Neymann e Pearson ci assicura che qualunque altro test di stessa ampiezza comporta una potenza inferiore a quella appena determinata. □

6.3 Ipotesi composte

Consideriamo ora un generico sistema di ipotesi composte

$$H_0 : \theta \in \Theta_0 \quad \text{vs.} \quad H_1 : \theta \in \Theta_1$$

e un test (A, R). Estendiamo a questo caso generale le nozioni di probabilità di errore di I e II tipo e quella di potenza del test. Introduciamo la seguente notazione. Dato un sottoinsieme B di $\mathcal{X}^n$, indichiamo con

$$\mathbb{P}_\theta(B; H_i) = \mathbb{P}_\theta[\mathbf{X}_n \in B; H_i], \qquad \theta \in \Theta_i$$

la probabilità che $\mathbf{X}_n$ assuma valori in B, calcolata con la legge di probabilità $f_n(\cdot;\theta)$ in cui θ è un qualsiasi valore in Θ_i. Si noti che, dal momento che l'insieme Θ_i contiene più possibili valori di θ (quelli specificati da H_i) la probabilità $\mathbb{P}_\theta(B; H_i)$ non è un numero, ma una *funzione* di θ, che varia al variare di θ in Θ_i. Se le v.a. X_i sono assolutamente continue, abbiamo inoltre che

$$\mathbb{P}_\theta(B; H_i) = \int_B f_n(\mathbf{x}_n;\theta) d\mathbf{x}_n, \quad \theta \in \Theta_i.$$

6.15 Definizione (Funzioni di probabilità di errore). Dato un test statistico (A, R) per la scelta tra le ipotesi

$$H_0 : \theta \in \Theta_0 \quad \text{vs.} \quad H_1 : \theta \in \Theta_1,$$

la funzione di *probabilità di errore di I tipo* è

$$\alpha(\theta) = \mathbb{P}_\theta(R; H_0) = \mathbb{P}_\theta[\mathbf{X}_n \in R; H_0], \quad \theta \in \Theta_0;$$

la funzione di *probabilità di errore di II tipo* è

$$\beta(\theta) = \mathbb{P}_\theta(A; H_1) = \mathbb{P}_\theta[\mathbf{X}_n \in A; H_1], \quad \theta \in \Theta_1.$$

□

Nel caso in cui le X_i sono v.a. assolutamente continue si ha che

$$\alpha(\theta) = \int_R f_n(\mathbf{x}_n;\theta) d\mathbf{x}_n, \qquad \theta \in \Theta_0$$

e

$$\beta(\theta) = \int_A f_n(\mathbf{x}_n;\theta) d\mathbf{x}_n, \qquad \theta \in \Theta_1.$$

Le caratteristiche probabilistiche di un test, rilevanti alla sua valutazione, sono sintetizzate dalla funzione di θ introdotta nella seguente definizione.

6.16 Definizione (Funzione di potenza del test). Dato un modello statistico, il sistema di ipotesi $H_0 : \theta \in \Theta_0$ vs. $H_1 : \theta \in \Theta_1$ e un campione di dimensione n, si chiama *funzione di potenza* del test (A, R) la funzione di θ definita da

$$\eta(\theta) = \mathbb{P}_\theta(R) = \mathbb{P}_\theta(\mathbf{X}_n \in R), \qquad \theta \in \Theta.$$

□

Si noti che

$$\eta(\theta) = \begin{cases} \alpha(\theta) & \theta \in \Theta_0 \\ 1 - \beta(\theta) & \theta \in \Theta_1 \end{cases}. \tag{6.13}$$

Al pari di $\alpha(\theta)$ e $\beta(\theta)$, la funzione di potenza si calcola con la distribuzione campionaria di $\mathbf{X}_n$ (o, equivalentemente, con quella di $W(\mathbf{X}_n)$, come vedremo tra poco) che, ovviamente, dipende dal parametro incognito θ. Per questo η è una funzione di θ. Nel caso in cui le X_I sono assolutamente continue, abbiamo che

$$\eta(\theta) = \int_R f_n(\mathbf{x}_n;\theta) d\mathbf{x}_n.$$

6.17 Osservazione Nel caso di ipotesi semplici le funzioni $\alpha(\theta)$ e $\beta(\theta)$ assumono ciascuna un solo valore, $\alpha(\theta_0)$ e $\beta(\theta_1)$, coincidenti con le probabilità di I e II tipo, α e β. La funzione di potenza diventa

$$\eta(\theta) = \begin{cases} \alpha & \theta = \theta_0 \\ 1 - \beta & \theta = \theta_1 \end{cases}.$$

Come visto, nel caso di ipotesi semplici si chiama potenza del test la quantità $1-\beta$. □

La funzione potenza $\eta(\theta)$ descrive, al variare di θ, la probabilità di rifiutare H_0 utilizzando il test con (A, R). Un buon test deve avere funzione di potenza che assume valori vicini a zero quando $\theta \in \Theta_0$ e valori vicini ad uno quando $\theta \in \Theta_1$. La funzione di potenza del test ideale è quindi

$$\eta_{id}(\theta) = 1_{\Theta_1}(\theta) = \begin{cases} 0 & \theta \in \Theta_0 \\ 1 & \theta \in \Theta_1 \end{cases},$$

ovvero quella in cui $\alpha(\theta) = 0$ per ogni $\theta \in \Theta_0$ e $\beta(\theta) = 0$ per ogni $\theta \in \Theta_1$. Nelle situazioni concrete, un test è tanto migliore quanto più sono bassi i valori di $\alpha(\theta)$ e quanto più sono elevati i valori di $(1-\beta(\theta))$. Il comportamento della funzione di potenza in θ_0 è sintetizzato dalle quantità introdotte nella seguente definizione.

6.18 Definizione (Ampiezza e livello di un test). Un test (A, R) con funzione di potenza $\eta(\cdot)$ è di *ampiezza* α, con $\alpha \in (0, 1)$, se

$$\sup_{\theta \in \Theta_0} \eta(\theta) = \sup_{\theta \in \Theta_0} \alpha(\theta) = \alpha.$$

Il test è di *livello* α se

$$\sup_{\theta \in \Theta_0} \eta(\theta) = \sup_{\theta \in \Theta_0} \alpha(\theta) \leq \alpha.$$

□

L'ampiezza del test è quindi il massimo valore della probabilità di errore di I tipo del test. Si noti che, se le v.a. sono assolutamente continue, è possibile determinare un test di ampiezza α per qualunque valore scelto per α in $(0, 1)$. Se invece le v.a. sono discrete, ciò non è possibile. Questo fatto motiva la necessità di introdurre il concetto di *livello* di un test. La nozione di ampiezza un test (ovvero il comportamento di η in Θ_0) viene utilizzata per determinare le regioni A e R.

Implementazione di un test per il confronto tra ipotesi composte

La procedura generale per implementare un test è la seguente:

1. si sceglie un test con regioni A ed R (generiche);
2. si determinano le regioni A e R in modo tale che il test sia di ampiezza α;
3. si estrae un campione $\mathbf{x}_n$: se $\mathbf{x}_n \in A$ si accetta H_0, altrimenti si accetta H_1.

6.19 Osservazione Le nozioni introdotte fin qui possono essere espresse anche in termini di una statistica test $W : \mathcal{X}^n \to \mathcal{W}$. Dato un sottoinsieme B_W di $\mathcal{W}$, sia

$$\mathbb{P}_\theta^W(B_W; H_i) = \mathbb{P}_\theta^W[W(\mathbf{X}_n) \in B_W; H_i], \qquad \theta \in \Theta_i.$$

Se le v.a. X_i sono assolutamente continue, abbiamo che

$$\mathbb{P}_\theta^W(B_W; H_i) = \int_{B_W} f_W(w; \theta) dw, \quad \theta \in \Theta_i.$$

Se $W(\mathbf{X}_n)$ è una v.a. a valori in $\mathbb{R}^1$ (ad esempio $|\overline{X}_n - 2|$ in un esempio precedente) con densità $f_W(\cdot; \theta)$, si ha che

$$\mathbb{P}_\theta^W(B_W; H_i) = \int_{B_W} f_W(w; \theta) dw, \quad \theta \in \Theta_i,$$

e il calcolo di $\mathbb{P}_\theta(B_W; H_i)$ richiede quindi un integrale unidimensionale.

Indichiamo ora con A_W e R_W le regioni di accettazione e rifiuto del test espresse in funzione di W. Abbiamo allora che:

$$\alpha(\theta) = \mathbb{P}_\theta(R; H_0) = \mathbb{P}_\theta^W(R_W; H_0) = \mathbb{P}_\theta^W[W(\mathbf{X}_n) \in R_W; H_0], \quad \theta \in \Theta_0;$$

e

$$\beta(\theta) = \mathbb{P}_\theta(A; H_1) = \mathbb{P}_\theta^W(A_W; H_1) = \mathbb{P}_\theta^W[W(\mathbf{X}_n) \in A_W; H_1], \quad \theta \in \Theta_1.$$

Nel caso in cui le X_i sono v.a. assolutamente continue e $W(\mathbf{X}_n)$ è una v.a. a valori in $\mathbb{R}^1$, si ha che

$$\alpha(\theta) = \int_R f_n(\mathbf{x}_n; \theta) d\mathbf{x}_n = \int_{R_W} f_W(w; \theta) dw, \quad \theta \in \theta_0$$

e

$$\beta(\theta) = \int_A f_n(\mathbf{x}_n; \theta) d\mathbf{x}_n = \int_{A_W} f_W(w; \theta) dw, \quad \theta \in \theta_1.$$

La funzione di potenza del test W è quindi

$$\eta_W(\theta) = \mathbb{P}_\theta^W(R_W) = \mathbb{P}_\theta[W(\mathbf{X}_n) \in R_W].$$

Come sopra, abbiamo che

$$\eta_W(\theta) = \begin{cases} \alpha(\theta) & \theta \in \Theta_0 \\ 1 - \beta(\theta) & \theta \in \Theta_1 \end{cases}. \tag{6.14}$$

Poichè assumiamo che W assume valori in $\mathbb{R}$, abbiamo che

$$\eta_W(\theta) = \int_{R_W} f_W(w; \theta) dw.$$

Inoltre, un test W è di *ampiezza* α, con $\alpha \in (0, 1)$, se

$$\sup_{\theta \in \Theta_0} \eta_W(\theta) = \sup_{\theta \in \Theta_0} \alpha(\theta) = \alpha.$$

Un test W è di *livello* α se

$$\sup_{\theta \in \Theta_0} \eta_W(\theta) = \sup_{\theta \in \Theta_0} \alpha(\theta) \leq \alpha.$$

La procedura generale per implementare un test W è la seguente:

1. si sceglie un test W;
2. si determinano le regioni A_W e R_W in modo tale che il test sia di ampiezza α;
3. si estrae un campione $\mathbf{x}_n$ e si calcola $W(\mathbf{x}_n)$: se $W(\mathbf{x}_n) \in A_W$ si accetta H_0, altrimenti si accetta H_1.

6.3.1 Test del rapporto delle verosimiglianze massimizzate

Il metodo più comune per ottenere un test per il confronto di ipotesi composte è una generalizzazione del test del rapporto delle verosimiglianze, già introdotta nel Capitolo 3. □

6.20 Definizione (Test del rapporto delle verosimiglianze massimizzate). Dato il modello statistico $\{\mathcal{X}^n, f_n(\mathbf{x}_n; \theta), \theta \in \Theta\}$, un campione $X_1, \ldots, X_n$ e il sistema di ipotesi

$$H_0 : \theta \in \Theta_0 \quad \text{vs.} \quad H_1 : \theta \in \Theta_1,$$

1. il test del rapporto delle verosimiglianze massimizzate (rvm) si basa sulla regione di accettazione
$$A = \{\mathbf{x}_n \in \mathcal{X}^n : \lambda_{01}^m(\mathbf{x}_n) \geq k\}$$
dove
$$\lambda_{01}^m(\mathbf{x}_n) = \frac{\sup_{\theta\in\Theta_0} L(\theta;\mathbf{x}_n)}{\sup_{\theta\in\Theta} L(\theta;\mathbf{x}_n)},$$
$L(\theta;\mathbf{x}_n)$ è la funzione di verosimiglianza di θ associata al campione $\mathbf{x}_n$ e dove k è una costante in $[0,1]$.

2. Il test rvm di *ampiezza* α (con $\alpha \in (0,1)$) si ottiene imponendo che
$$\sup_{\theta\in\theta_0} \eta(\theta) = \alpha,$$
dove $\eta(\theta) = \mathbb{P}_\theta(R)$ è la funzione di potenza del test rvm e $R = A^C$ la regione di rifiuto del test.

□

6.21 Esempio (Ipotesi alternativa unilaterale, modello normale). Si consideri un campione casuale dal modello $\mathrm{N}(\theta,\sigma^2)$ (varianza nota).

(a) Consideriamo il sistema di ipotesi unilaterale
$$H_0 : \theta \leq \theta_0, \quad \text{vs.} \quad H_1 : \theta > \theta_0.$$
Ricordando che $L(\theta;\mathbf{x}_n) \propto \exp\{-\frac{n}{2\sigma^2}(\theta-\overline{x}_n)^2\}$ e che $\widehat{\theta}_{mv} = \overline{x}_n$, si ottiene che
$$\lambda_{01}^m(\mathbf{x}_n) = \frac{\sup_{\theta\in\Theta_0} L(\theta;\mathbf{x}_n)}{\sup_{\theta\in\Theta} L(\theta;\mathbf{x}_n)} = \frac{\sup_{\theta\leq\theta_0} L(\theta;\mathbf{x}_n)}{L(\widehat{\theta}_{mv};\mathbf{x}_n)} = \frac{\sup_{\theta\leq\theta_0} L(\theta;\mathbf{x}_n)}{1}.$$
Inoltre si ha che
$$\lambda_{01}^m(\mathbf{x}_n) = \sup_{\theta\leq\theta_0} L(\theta;\mathbf{x}_n) = \begin{cases} L(\widehat{\theta}_{mv};\mathbf{x}_n) & \text{se } \widehat{\theta}_{mv} < \theta_0 \\ L(\theta_0;\mathbf{x}_n) & \text{se } \widehat{\theta}_{mv} > \theta_0 \end{cases} =$$
$$= \begin{cases} 1 & \text{se } \widehat{\theta}_{mv} < \theta_0 \\ L(\theta_0;\mathbf{x}_n) & \text{se } \widehat{\theta}_{mv} > \theta_0 \end{cases}.$$
Si ha quindi che la regione di rifiuto del test è
$$R = \{\mathbf{x}_n \in \mathcal{X}^n : \lambda_{01}^m(\mathbf{x}_n) < k\} = \{\mathbf{x}_n \in \mathcal{X}^n : \overline{x}_n > k'\},$$
dove k' è soluzione dell'equazione (in $\overline{x}_n$) $L(\theta_0;\mathbf{x}_n) = k$. La funzione di potenza del test trovato è quindi
$$\eta(\theta) = \mathbb{P}_\theta(R) = \mathbb{P}_\theta(\overline{X}_n > k') = 1 - \Phi\left(\frac{\sqrt{n}(k'-\theta)}{\sigma}\right). \tag{6.15}$$
La funzione $\eta(\theta)$ è crescente in θ, pertanto
$$\sup_{\theta\leq\theta_0} \eta(\theta) = \eta(\theta_0).$$

Il test di ampiezza α si ottiene quindi determinando il valore k' tale che

$$\sup_{\theta \leq \theta_0} \eta(\theta) = \eta(\theta_0) = 1 - \Phi\left(\frac{\sqrt{n}(k' - \theta_0)}{\sigma}\right) = \alpha,$$

ovvero ponendo

$$\frac{\sqrt{n}(k' - \theta_0)}{\sigma} = z_{1-\alpha}$$

da cui si ottiene

$$k'_\alpha = \theta_0 + z_{1-\alpha}\frac{\sigma}{\sqrt{n}}.$$

Si ottiene quindi che

$$R = \left\{\mathbf{x}_n \in \mathcal{X}^n : \ \overline{x}_n > \theta_0 + z_{1-\alpha}\frac{\sigma}{\sqrt{n}}\right\} = \left\{\mathbf{x}_n \in \mathcal{X}^n : \ W(\mathbf{x}_n) > z_{1-\alpha}\right\}, \tag{6.16}$$

dove

$$W(\mathbf{x}_n) = \frac{\sqrt{n}(\overline{x}_n - \theta_0)}{\sigma}. \tag{6.17}$$

È importante notare che la statistica test e le regioni A ed R determinate nell'esempio precedente coincidono con quelle determinate nell'esempio (6.8) per il confronto tra ipotesi semplici relativamente al caso $\theta_1 > \theta_0$. La Figura 6.1 rappresenta quindi le regioni A ed R anche per questo esempio.
Possiamo ora determinare la funzione di potenza per il test di ampiezza α sostituendo in (6.15) l'espressione $k'_\alpha = \theta_0 + z_{1-\alpha}\frac{\sigma}{\sqrt{n}}$, ottenendo

$$\eta(\theta) = \mathbb{P}_\theta(R) = 1 - \Phi\left(z_{1-\alpha} + \frac{\sqrt{n}(\theta_0 - \theta)}{\sigma}\right) \tag{6.18}$$

Si noti che la funzione di potenza (6.18) calcolata in θ_1 coincide esattamente con l'espressione della potenza $1-\beta$ del test determinata nell'esempio 6.14, sempre per il confronto tra ipotesi puntuali. Si osservi inoltre che $\eta(\theta) \to 0$ per $\theta \to -\infty$ e che $\eta(\theta) \to 1$ per $\theta \to +\infty$. La Figura 6.4 riporta il grafico della funzione di potenza del test al variare di θ per $\theta_0 = 368$, $\sigma = 15$, $\alpha = 0.05$ e per tre numerosità campionarie ($n = 25, 50, 100$). La retta verticale è tracciata in corrispondenza del valore $\theta_0 = 368$, dove $\eta(\theta)$ assume il valore pari a 0.05. La retta orizzontale ètracciata in corrispondenza del valore $\eta(368) = 0.05$. Per ogni punto $\theta > \theta_0$ la funzione η cresce al crescere di n: al crescere di n aumenta la probabilitá di rifiutare l'ipotesi nulla quando è vera l'ipotesi alternativa.

(b) Consideriamo ora il sistema di ipotesi miste

$$H_0 : \ \theta = \theta_0, \quad \text{vs.} \quad H_1 : \ \theta > \theta_0. \tag{6.19}$$

In questo caso

$$\lambda^m_{01}(\mathbf{x}_n) = \frac{L(\theta_0; \mathbf{x}_n)}{\sup_{\theta \geq \theta_0} L(\theta; \mathbf{x}_n)}.$$

Poichè

$$\sup_{\theta > \theta_0} L(\theta; \mathbf{x}_n) = \begin{cases} L(\theta_0; \mathbf{x}_n) & \text{se} \quad \widehat{\theta}_{mv} < \theta_0 \\ 1 & \text{se} \quad \widehat{\theta}_{mv} > \theta_0 \end{cases}$$

abbiamo che

$$\lambda^m_{01} = \begin{cases} 1 & \text{se} \quad \widehat{\theta}_{mv} < \theta_0 \\ L(\theta_0; \mathbf{x}_n) & \text{se} \quad \widehat{\theta}_{mv} > \theta_0 \end{cases}.$$

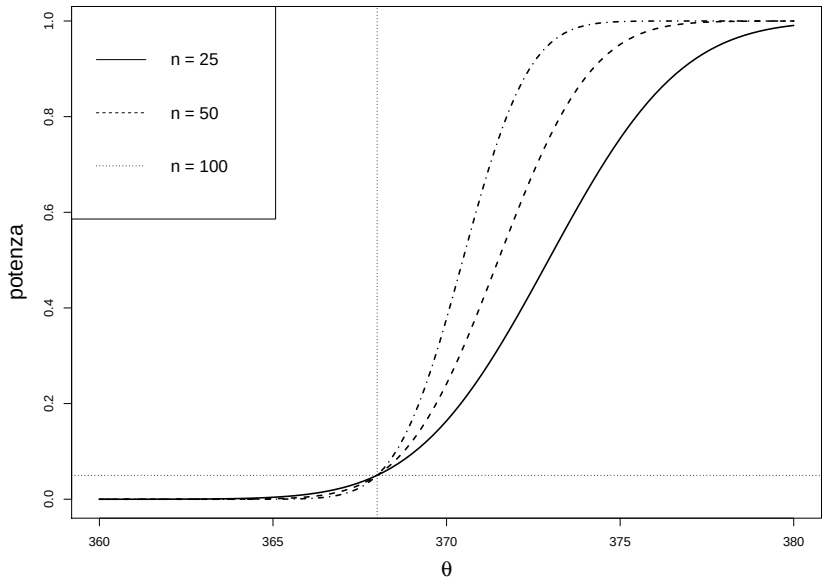

Figura 6.4: `Funzione di potenza per il test dell'esempio 6.21` ($H_0 : \theta \leq \theta_0$ `vs.` $H_1 : \theta > \theta_0$) `per tre valori della numerosità campionaria` n.

La statistica test λ_{01}^m coincide quindi con quella trovata per il test con ipotesi nulla composta ($\theta \leq \theta_0$) e quindi anche le generiche regioni A e R coincidono con quelle del caso (a). Inoltre, essendo l'ipotesi nulla puntuale, imporre che il test abbia ampiezza α coincide con il richiedere, esattamente come nel caso (a), che

$$\eta(\theta_0) = \alpha.$$

Discende pertanto che la regione di accettazione test del rvm di ampiezza α per le ipotesi 6.19 coincide con la regione (6.16), trovata per il confronto tra le ipotesi composte in (a). □

6.22 Esempio (Ipotesi alternativa unilaterale, modello normale - cont.). In modo analogo a quanto illustrato nell'esempio precedente, è semplice verificare che il test rvm di ampiezza α per il confronto tra le ipotesi composte

$$H_0 : \ \theta \geq \theta_0, \ \text{vs.} \ H_1 : \ \theta < \theta_0.$$

e per le ipotesi miste

$$H_0 : \ \theta = \theta_0, \ \text{vs.} \ H_1 : \ \theta < \theta_0$$

ha regione di rifiuto

$$R = \{\mathbf{x}_n \in \mathcal{X}^n : \ W(\mathbf{x}_n) < z_\alpha\},$$

dove l'espressione della statistica test $W(\mathbf{x}_n)$ è data dalla (6.17). In questo caso la statistica test e le regioni A ed R determinate nell'esempio precedente coincidono con quelle determinate nell'esempio (6.8) per il confronto tra ipotesi semplici relativamente, questa volta, al caso $\theta_1 < \theta_0$. La Figura 6.3 rappresenta le regioni A ed R anche per questo esempio. Inoltre

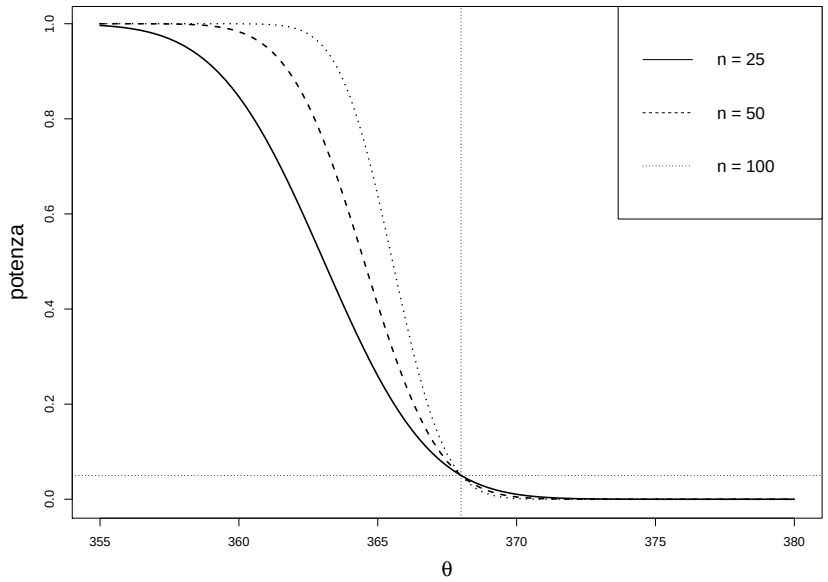

Figura 6.5: `Funzione di potenza per il test dell'esempio 6.22` ($H_0 : \theta \geq \theta_0$ `vs.` $H_1 : \theta < \theta_0$) `per tre valori della numerosità campionaria` n.

è semplice verificare che la funzione di potenza del test di ampiezza α è

$$\eta(\theta) = \mathbb{P}_\theta(R) = \Phi\left(z_\alpha + \frac{\sqrt{n}(\theta_0 - \theta)}{\sigma}\right) \tag{6.20}$$

Si noti che la funzione di potenza (6.20) calcolata in θ_1 coincide esattamente con l'espressione della potenza $1 - \beta$ del test per il confronto tra ipotesi puntuali (caso $\theta_1 < \theta_0$). Si osservi inoltre che $\eta(\theta) \to 1$ per $\theta \to -\infty$ e che $\eta(\theta) \to 0$ per $\theta \to +\infty$. La Figura 6.5 riporta il grafico della funzione di potenza del test al variare di θ per $\theta_0 = 368$, $\sigma = 15$, $\alpha = 0.05$ e per tre numerosità campionarie ($n = 25, 50, 100$). La retta verticale è tracciata in corrispondenza del valore $\theta_0 = 368$, dove $\eta(\theta)$ assume il valore pari a 0.05. La retta orizzontale tracciata in corrispondenza del valore $\eta(368) = 0.05$. Per ogni punto $\theta < \theta_0$ la funzione η cresce al crescere di n: al crescere di n aumenta la probabiliá di rifiutare l'ipotesi nulla quando è vera l'ipotesi alternativa.

□

6.23 Esempio (Ipotesi alternativa bilaterale, modello normale). Consideriamo il sistema di ipotesi miste

$$H_0 : \ \theta = \theta_0, \ \text{vs.} \ H_1 : \ \theta \neq \theta_0.$$

In questo caso il rvm coincide con la funzione di verosimiglianza relativa che, per il modello normale è

$$\lambda_{01}^m = \frac{L(\theta_0; \mathbf{x}_n)}{\sup_{\theta \in \Theta} L(\theta; \mathbf{x}_n)} = \overline{L}(\theta_0; \mathbf{x}_n) = \exp\left\{-\frac{n}{2\sigma^2}(\theta_0 - \overline{x}_n)^2\right\}$$

Si verifica facilmente che (per k arbitrario)

$$\lambda_{01}^m \geq k \quad \Leftrightarrow \quad \exp\left\{-\frac{n}{2\sigma^2}(\overline{x}_n - \theta_0)^2\right\} \geq k \quad \Leftrightarrow \quad \frac{n(\overline{x}_n - \theta_0)^2}{\sigma^2} \leq k'$$

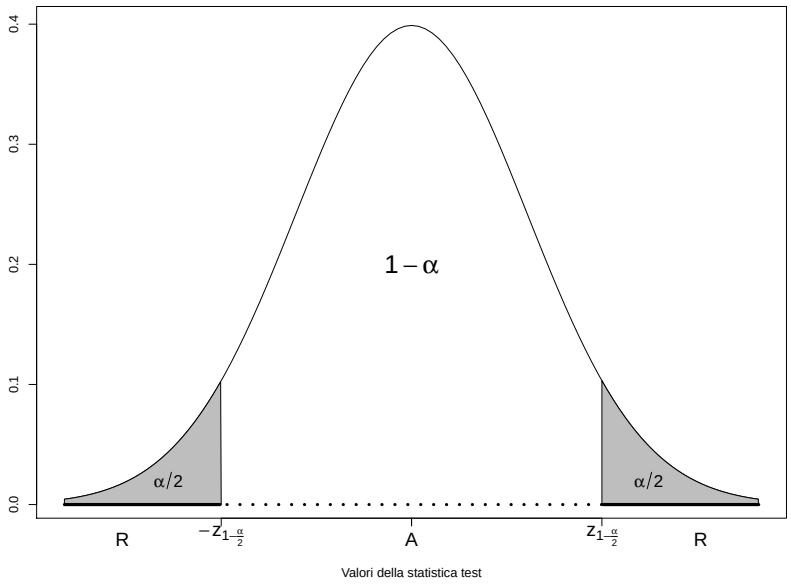

Figura 6.6: *Rappresentazione di A ed R per l'esempio 6.23.*

e quindi che la regione di accettazione del test è

$$\begin{aligned} A &= \left\{ \mathbf{x}_n \in \mathcal{X}^n : \; -k'' \leq \frac{\sqrt{n}(\overline{x}_n - \theta_0)}{\sigma} \leq k'' \right\} \\ &= \left\{ \mathbf{x}_n \in \mathcal{X}^n : \; |W(\mathbf{x}_n)| < k'' \right\}, \end{aligned}$$

dove

$$W(\mathbf{x}_n) = \frac{\sqrt{n}(\overline{x}_n - \theta_0)}{\sigma}.$$

Un test di ampiezza α si ottiene determinando un valore di k'' tale che $\eta(\theta_0) = \alpha$. Dobbiamo determinare quindi un valore di k'' tale che

$$\eta(\theta_0) = \mathbb{P}_{\theta_0}(R) = 1 - \mathbb{P}_{\theta_0}(A) = 1 - \mathbb{P}_{\theta_0}\left(-k'' \leq \frac{\sqrt{n}(\overline{X}_n - \theta_0)}{\sigma} \leq k'' \right) = \alpha$$

ovvero tale che

$$\mathbb{P}_{\theta_0}\left(-k'' \leq \frac{\sqrt{n}(\overline{X}_n - \theta_0)}{\sigma} \leq k'' \right) = 1 - \alpha.$$

Poichè, per $\theta = \theta_0$ (in gergo "sotto ipotesi nulla") la statistica test

$$W(\mathbf{X}_n) = \frac{\sqrt{n}(\overline{X}_n - \theta_0)}{\sigma} \sim \mathrm{N}(0, 1)$$

un test di ampiezza α si ottiene ponendo $k'' = z_{1-\frac{\alpha}{2}}$. La Figura 6.6 mostra la funzione di densità della statistica test $W(\mathbf{X}_n)$ e delle regioni A ed R, delimitate dai percentili della v.a. normale standardizzata $\pm z_{1-\frac{\alpha}{2}}$.
In questo caso la funzione di potenza del test di ampiezza α è

$$\eta(\theta) = \mathbb{P}_\theta(R) = 1 - \left[\Phi\left(\frac{\sqrt{n}(\theta_0 - \theta)}{\sigma} + z_{1-\frac{\alpha}{2}} \right) - \Phi\left(\frac{\sqrt{n}(\theta_0 - \theta)}{\sigma} - z_{1-\frac{\alpha}{2}} \right) \right].$$

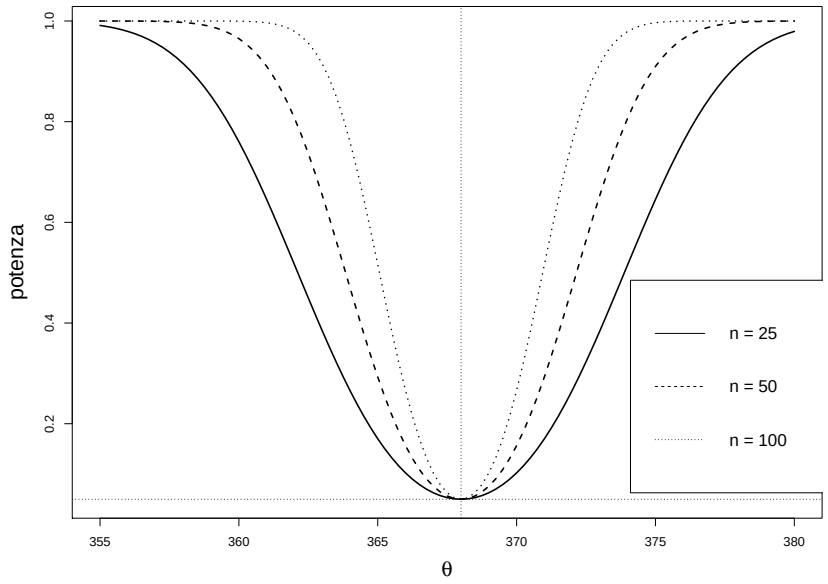

Figura 6.7: `Funzione di potenza per il test dell'esempio 6.23` ($H_0 : \theta = \theta_0$ `vs.` $H_1 : \theta \neq \theta_0$) `per tre valori della numerosità campionaria` n.

Si osservi che $\eta(\theta) \to 1$ per $|\theta| \to +\infty$ e che $\eta(\theta) \to \alpha$ per $\theta \to \theta_0$. La Figura 6.7 riporta il grafico della funzione di potenza del test al variare di θ per $\theta_0 = 368$, $\sigma = 15$, $\alpha = 0.05$ e per tre numerosità campionarie ($n = 25, 50, 100$). La retta verticale è tracciata in corrispondenza del valore $\theta_0 = 368$, dove $\eta(\theta)$ assume il valore pari a 0.05. La retta orizzontale ẗracciata in corrispondenza del valore $\eta(368) = 0.05$. Per ogni punto $\theta \neq \theta_0$ (valore nel quale η è vincolata ad assumere il valore $\alpha = 0.05$) la funzione η cresce al crescere di n: al crescere di n aumenta la probabiliá di rifiutare l'ipotesi nulla quando è vera l'ipotesi alternativa. □

Test del rvm con parametri di disturbo

Il test del rvm consente di trattare in modo piuttosto semplice il problema di verifica di ipotesi per parametri vettoriali. Supponiamo che il parametro $\boldsymbol{\theta}$ sia un vettore di dimensione $k = r + s$:

$$\boldsymbol{\theta} = (\boldsymbol{\theta}_r, \boldsymbol{\theta}_s),$$

dove

$$(\boldsymbol{\theta}_r, \boldsymbol{\theta}_s) = (\theta_1, \ldots, \theta_r, \theta_{r+1}, \ldots, \theta_{r+s}).$$

Consideriamo il sistema di ipotesi

$$H_0 : \ \theta_1 = \theta_{10}, \ldots, \theta_r = \theta_{r0}, \qquad \text{vs.} \qquad H_1 : \ \theta_1 \neq \theta_{10}, \ldots, \theta_r \neq \theta_{r0}$$

che coinvolge esclusivamente la componente $\boldsymbol{\theta}_r$ del vettore dei parametri, mentre $\boldsymbol{\theta}_s$ è un vettore di parametri di disturbo. Indichiamo con $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\theta}}_r, \widehat{\boldsymbol{\theta}}_s)$ il vettore delle stime di massima verosimiglianza di θ e $\widehat{\widehat{\boldsymbol{\theta}}}_s$ la stima di mv del parametro di disturbo sotto H_0. Il

rapporto delle verosimiglianze diventa in questo caso

$$\lambda_{01}^m(\mathbf{x}_n) = \frac{L(\boldsymbol{\theta}_{r0}, \widehat{\widehat{\boldsymbol{\theta}}}_s; \mathbf{x}_n)}{L(\widehat{\boldsymbol{\theta}}_r, \widehat{\boldsymbol{\theta}}_s; \mathbf{x}_n)}$$

La regione di rifiuto del test di ampiezza α è, al solito

$$R = \{\mathbf{x}_n \in \mathcal{X}^n : \lambda_{01}^m(\mathbf{x}_n) < k_\alpha\},$$

con k_α tale che la probabilità di R calcolata sotto ipotesi nulla è pari ad α:

$$\mathbb{P}(R; H_0) = \mathbb{P}(\lambda_{01}^m(\mathbf{X}_n) < k_\alpha) = \alpha.$$

6.24 Esempio (Modello normale, varianza incognita). Si consideri un campione casuale da una popolazione $\mathrm{N}(\theta, \sigma^2)$ in cui la varianza σ^2 non è nota (parametro di disturbo). Consideriamo il sistema di ipotesi $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$. Sotto H_0 la smv di σ^2 è $S_0^2 = \sum_{i=1}^n (x_i - \theta_0)^2/n$. Sotto H_1 la smv di (θ, σ^2) è $(\overline{X}_n, \widehat{\sigma}_n^2)$, dove $\widehat{\sigma}_n^2 = \sum_{i=1}^n (x_i - \overline{x}_n)^2/n$. Ricordando che

$$L(\theta, \sigma^2; \mathbf{x}_n) = \frac{1}{\left(\sigma\sqrt{2\pi}\right)^n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \theta)^2\right\}$$

e osservando che

$$S_0^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \theta_0)^2 = \frac{1}{n}\sum_{i=1}^n (x_i \pm \overline{x}_n - \theta_0)^2 = \widehat{\sigma}_n^2 + (\overline{x}_n - \theta_0)^2,$$

il rvm diventa

$$\begin{aligned}\lambda_{01}^m(\mathbf{x}_n) &= \frac{L(\theta_0, S_0^2; \mathbf{x}_n)}{L(\overline{x}_n, \widehat{\sigma}_n^2; \mathbf{x}_n)} = \left(\frac{\widehat{\sigma}_n^2}{S_0^2}\right)^{n/2} \frac{\exp\{-\frac{1}{2}nS_0^2/S_0^2\}}{\exp\{-\frac{1}{2}n\widehat{\sigma}_n^2/\widehat{\sigma}_n^2\}} = \left(\frac{\widehat{\sigma}_n^2}{\widehat{\sigma}_n^2 + (\overline{x}_n - \theta_0)^2}\right)^{n/2} \\ &= \frac{1}{\left[1 + \frac{T(\mathbf{x}_n)^2}{n-1}\right]^{n/2}},\end{aligned}$$

dove

$$T(\mathbf{x}_n) = \frac{\sqrt{n}(\overline{x}_n - \theta_0)}{S_n}$$

e dove S_n^2 è la varianza campionaria corretta. Il rvm λ_{01}^m è una funzione monotona decrescente di T^2. La regione di accettazione del test è quindi (per k arbitrario)

$$A = \{\mathbf{x}_n \in \mathcal{X}^n : T(\mathbf{x}_n)^2 \leq k\} = \{\mathbf{x}_n \in \mathcal{X}^n : -\sqrt{k} \leq T(\mathbf{x}_n) \leq \sqrt{k}\}.$$

Il test di ampiezza α si ottiene determinando il valore di k tale che $\mathbb{P}_{\theta_0}(R) = \alpha$. ovvero che $\mathbb{P}_{\theta_0}(A) = 1 - \alpha$. Osservando che, sotto ipotesi nulla,

$$\frac{\sqrt{n}(\overline{X}_n - \theta_0)}{S_n} \sim t_{n-1}$$

abbiamo che (ponendo $k' = \sqrt{k}$)

$$\mathbb{P}_{\theta_0}(A) = \mathbb{P}_{\theta_0}\left(-k' \leq \frac{\sqrt{n}(\overline{x}_n - \theta_0)}{S_n} \leq k'\right) = 1 - \alpha$$

per $k' = t_{n-1;1-\frac{\alpha}{2}}$, dove $t_{n-1;\epsilon}$ indica il percentile di livello ϵ di una v.a. t di Student con $n-1$ gradi di libertà, le regioni di accettazione e rifiuto del test di ampiezza α sono quindi:

$$A = \left\{ \mathbf{x}_n \in \mathcal{X}^n : \ -t_{n-1;1-\frac{\alpha}{2}} \leq \frac{\sqrt{n}(\overline{x}_n - \theta_0)}{S_n} \leq t_{n-1;1-\frac{\alpha}{2}} \right\}$$

e

$$R = \left\{ \mathbf{x}_n \in \mathcal{X}^n : \ \frac{\sqrt{n}|\overline{x}_n - \theta_0|}{S_n} > t_{n-1;1-\frac{\alpha}{2}} \right\}.$$

□

6.3.2 Valore-p

La procedura frequentista per la verifica di ipotesi illustrata fin qui ha come obiettivo la scelta tra le due ipotesi formulate. I dati entrano in gioco solo per calcolare il valore della statistica test e stabilire se questa cade nella regione di accettazione o di rifiuto e non per stabilire con quanta *forza* l'ipotesi nulla viene rifiutata oppure accettata. Nella pratica è però necessario quantificare, attraverso i dati osservati, l'*evidenza* a favore o contraria alle ipotesi considerate. Usualmente si cerca di misurare l'evidenza a favore di H_0 attraverso una quantità detta *valore-p* o *livello di significatività osservato* (in inglese, *p-value*). Introduciamo questo concetto con un esempio.

6.25 Esempio Consideriamo il sistema di ipotesi $H_0 : \theta = \theta_0$ vs. $H_1 : \theta > \theta_0$. Il test del rapporto delle verosimiglianze di ampiezza $\alpha = 0.05$ prevede il rifiuto dell'ipotesi nulla se

$$W(\mathbf{x}_n) = \frac{\sqrt{n}(\overline{x}_n - \theta_0)}{\sigma} > z_{1-\alpha} = 1.64.$$

Supponiamo ora di avere a disposizione due campioni $\mathbf{x}_n^1$ e $\mathbf{x}_n^2$, in corrispondenza dei quali si ha $W(\mathbf{x}_n^1) = 1.90$ e $W(\mathbf{x}_n^2) = 2.85$. In entrambi i casi i campioni osservati ci consentono di rifiutare H_0, dal momento che i valori osservati della statistica test superano la soglia 1.64. Tuttavia, il secondo campione testimonia una discrepanza (opportunamente standardizzata) tra dati ($\overline{x}_n$) e valore di θ fissato nell'ipotesi nulla (θ_0) superiore a quella del primo campione: il valore 2.85 è più addentrato nella regione di rifiuto del valore 1.90. Quindi, se è vero che entrambi i campioni portano alla stessa decisione (rifiutare H_0), il secondo campione ci fa rifiutare l'ipotesi nulla con maggiore forza del primo. In generale, quanto maggiore è $W(\mathbf{x}_n)$ rispetto a 1.64, tanto più forte è l'evidenza di $\mathbf{x}_n$ contro l'ipotesi nulla. Per quantificare la forza con cui i due campioni ci consentono di rifiutare H_0 possiamo allora calcolare le probabilità di osservare valori di $W(\mathbf{X}_n)$ ancora più in contrasto con H_0 di quelli osservati, ovvero :

$$\mathbb{P}_{\theta_0}\left(\frac{\sqrt{n}(\overline{X}_n - \theta_0)}{\sigma} > 1.90 \right) = 1 - \Phi(1.90) = 0.0287$$

e

$$\mathbb{P}_{\theta_0}\left(\frac{\sqrt{n}(\overline{X}_n - \theta_0)}{\sigma} > 2.85 \right) = 1 - \Phi(2.85) = 0.0022,$$

Più piccola è la probabilità, maggiore è l'evidenza del campione contro H_0. In questo caso, come ci aspettavamo, il campione $\mathbf{x}_n^2$ è quello che ci fa rifiutare con maggiore forza H_0 ($0.0287/0.0022 \simeq 13$). □

Introduciamo ora in modo un pò più formale (ma non del tutto generale) la nozione di valore-p. Per semplicità, consideriamo v.a. e statistiche test assolutamente continue.

Valore-p per ipotesi nulla semplice

Supponiamo per il momento che l'ipotesi nulla sia puntuale: $H_0 : \theta = \theta_0$. Data una statistica test W e un campione osservato $\mathbf{x}_n$, il valore-p è la probabilità (calcolata ipotizzando vera H_0) di osservare un valore di $W(\mathbf{X}_n)$ più in contrasto con H_0 di quanto sia il valore osservato $W(\mathbf{x}_n)$. L'idea è piuttosto intuitiva. Se in corrispondenza di un campione $\mathbf{x}_n$ si ottiene un valore-p elevato, ciò vuol dire che, assumendo vera H_0, è molto probabile osservare un valore di W che indica un contrasto con H_0 più elevato di quello effettivamente osservato. In altri termini, il valore osservato $W(\mathbf{x}_n)$ non indica un contrasto significativo contro H_0. Se invece si osserva un valore-p piuttosto basso, ciò vuol dire che, sotto H_0, è difficile osservare una evidenza contraria a H_0 più forte di quella che indica $W(\mathbf{x}_n)$. In altre parore:

- valore-p basso $\Rightarrow$ $W(\mathbf{x}_n)$ è evidenza sperimentale contro H_0
- valore-p alto $\Rightarrow$ $W(\mathbf{x}_n)$ non è evidenza sperimentale contro H_0

Indichiamo con F_{θ_0} la funzione di ripartizione di $W(\mathbf{X}_n)$ sotto H_0 e consideriamo i tre seguenti casi[4].

1. Nel caso in cui la regione di rifiuto di un test sia
$$R = \{\mathbf{x}_n \in \mathcal{X}^n : \ W(\mathbf{x}_n) > \xi\}$$
il valore-p è
$$p_{\theta_0}(\mathbf{x}_n) = \mathbb{P}_{\theta_0}[W(\mathbf{X}_n) > W(\mathbf{x}_n)] = 1 - F_{\theta_0}[W(\mathbf{x}_n)]$$
2. Nel caso in cui la regione di rifiuto di un test sia
$$R = \{\mathbf{x}_n \in \mathcal{X}^n : \ W(\mathbf{x}_n) < \xi\}$$
il valore-p è
$$p_{\theta_0}(\mathbf{x}_n) = \mathbb{P}_{\theta_0}[W(\mathbf{X}_n) < W(\mathbf{x}_n)] = F_{\theta_0}[W(\mathbf{x}_n)].$$
3. Nel caso in cui la regione di rifiuto di un test sia
$$R = \{\mathbf{x}_n \in \mathcal{X}^n : \ W(\mathbf{x}_n) < -\xi \text{ oppure } W(\mathbf{x}_n) > \xi\}$$
il valore-p è
$$p_{\theta_0}(\mathbf{x}_n) = \mathbb{P}_{\theta_0}[|W(\mathbf{X}_n)| > |W(\mathbf{x}_n)|] = 2(1 - F_{\theta_0}[|w^{oss}|]).$$

6.26 Esempio (Modello normale, varianza nota). Nel caso del modello normale con varianza σ^2 nota abbiamo visto che la statistica test per i tre più comuni sistemi di ipotesi è
$$W(\mathbf{X}_n) = \frac{\sqrt{n}(\overline{X}_n - \theta_0)}{\sigma}.$$
Se, ad esempio, consideriamo il sistema di ipotesi $H_0 : \theta = \theta_0$ vs $H_1 : \theta > \theta_0$, la cui regione di rifiuto è $R = \{\mathbf{x}_n \in \mathcal{X}^n : \ W(\mathbf{x}_n) > k\}$, la statistica test $W(\mathbf{X}_n)$ ha distribuzione $\mathrm{N}(0, 1)$ e il valore-p sarà
$$p_{\theta_0}(\mathbf{x}_n) = \mathbb{P}_{\theta_0}[W(\mathbf{X}_n) > W(\mathbf{x}_n)] = 1 - \Phi[W(\mathbf{x}_n)] = 1 - \Phi\left[\frac{\sqrt{n}(\overline{x}_n - \theta_0)}{\sigma}\right],$$

[4]Si ricordi che, per semplicità, consideriamo v.a. assolutamente continue

dove $\Phi(\cdot)$ indica, al solito, la funzione di ripartizione della v.a. normale standard. Analogamente, se consideriamo il sistema di ipotesi $H_0 : \theta = \theta_0$ vs $H_1 : \theta < \theta_0$, la cui regione di rifiuto è $R = \{\mathbf{x}_n \in \mathcal{X}^n : \ W(\mathbf{x}_n) < k\}$, si ha che

$$p_{\theta_0}(\mathbf{x}_n) = \mathbb{P}_{\theta_0}[W(\mathbf{X}_n) < W(\mathbf{x}_n)] = \Phi[W(\mathbf{x}_n)] = \Phi\left[\frac{\sqrt{n}(\overline{x}_n - \theta_0)}{\sigma}\right],$$

Infine, se il sistema di ipotesi è $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$, la cui regione di rifiuto è $R = \{\mathbf{x}_n \in \mathcal{X}^n : \ |W(\mathbf{x}_n)| > k\}$, si ha che

$$p_{\theta_0}(\mathbf{x}_n) = \mathbb{P}_{\theta_0}[|W(\mathbf{X}_n)| > |W(\mathbf{x}_n)|] = 2(1 - \Phi[|W(\mathbf{x}_n)|]) = 2\left(1 - \Phi\left[\frac{\sqrt{n}|\overline{x}_n - \theta_0|}{\sigma}\right]\right),$$

□

Valore-p per ipotesi composte

Nei caso in cui l'ipotesi nulla è composta ($H_0 : \theta \in \Theta_0$), la probabilità di osservare una evidenza più contraria a H_0 di quella osservata è, in generale, una funzione di θ, che indichiamo con $p_\theta(\mathbf{x}_n)$. Dato un campione osservato $\mathbf{x}_n$ di $\mathcal{X}^n$ definiamo valore-p il numero

$$p(\mathbf{x}_n) = \sup_{\theta \in \Theta_0} p_\theta(\mathbf{x}_n). \tag{6.21}$$

Il valore-p è quindi la massima evidenza (per θ che varia in Θ_0) che il campione dà a favore di H_0.

6.27 Esempio (Modello normale, varianza nota - cont.). Sempre con riferimento al modello $\mathrm{N}(\theta, \sigma^2)$ (varianza nota), consideriamo ora il sistema di ipotesi composte

$$H_0 : \theta \leq \theta_0 \quad \text{vs} \quad H_1 : \theta > \theta_0,$$

la cui regione di rifiuto è

$$R = \{\mathbf{x}_n \in \mathcal{X}^n : \ W(\mathbf{x}_n) > k\}, \qquad W(\mathbf{x}_n) = \frac{\sqrt{n}(\overline{x}_n - \theta_0)}{\sigma}.$$

In questo caso si ha che

$$\begin{aligned}
p(\mathbf{x}_n) = \sup_{\theta \in \Theta_0} p_\theta(\mathbf{x}_n) &= \sup_{\theta \leq \theta_0} \mathbb{P}_\theta\left[W(\mathbf{X}_n) > W(\mathbf{x}_n)\right] \\
&= \sup_{\theta \leq \theta_0} \mathbb{P}_\theta\left(\frac{\sqrt{n}(\overline{X}_n - \theta_0)}{\sigma} > W(\mathbf{x}_n)\right) \\
&= \sup_{\theta \leq \theta_0} \mathbb{P}_\theta\left(\frac{\sqrt{n}(\overline{X}_n - \theta)}{\sigma} > \frac{\sqrt{n}(\theta_0 - \theta)}{\sigma} + W(\mathbf{x}_n)\right) \\
&= \sup_{\theta \leq \theta_0} \mathbb{P}_\theta\left(Z > \frac{\sqrt{n}(\theta_0 - \theta)}{\sigma} + W(\mathbf{x}_n)\right) \\
&= \sup_{\theta \leq \theta_0} \left(1 - \Phi\left[\frac{\sqrt{n}(\theta_0 - \theta)}{\sigma} + W(\mathbf{x}_n)\right]\right) \\
&= 1 - \Phi\left[W(\mathbf{x}_n)\right] \\
&= 1 - \Phi\left[\frac{\sqrt{n}(\overline{x}_n - \theta_0)}{\sigma}\right] \\
&= p_{\theta_0}(\mathbf{x}_n),
\end{aligned}$$

dove la terzultima uguaglianza si giustifica osservando che $\left(1 - \Phi\left[\frac{\sqrt{n}(\theta_0-\theta)}{\sigma} + W(\mathbf{x}_n)\right]\right)$ è una funzione crescente di θ che, per $\theta \leq \theta_0$ assume valore massimo in θ_0 e dove $p_{\theta_0}(\mathbf{x}_n)$ indica il valore-p nel caso di ipotesi nulla semplice. Abbiamo quindi verificato che in questo caso il valore-p coincide con quello in cui l'ipotesi nulla è puntuale. In modo del tutto analogo si verifica che, per il sistema di ipotesi

$$H_0 : \theta \geq \theta_0 \text{ vs } H_1 : \theta < \theta_0,$$

il valore-p è

$$\Phi\left[\frac{\sqrt{n}(\overline{x}_n - \theta_0)}{\sigma}\right].$$

La seguente tabella riporta l'espressione del valore-p per i sistemi d'ipotesi più comuni.

H_0	H_1	rif. H_0	Valore-p
$\theta = \theta_0$	$\theta \neq \theta_0$	$\lvert W(\mathbf{x}_n)\rvert > \lvert\xi\rvert$	$2\left[1 - \Phi\left(\frac{\sqrt{n}\lvert\overline{x}_n-\theta_0\rvert}{\sigma}\right)\right]$
$\theta \leq (=)\, \theta_0$	$\theta > \theta_0$	$W(\mathbf{x}_n) > \xi$	$1 - \Phi\left[\frac{\sqrt{n}(\overline{x}_n-\theta_0)}{\sigma}\right]$
$\theta \geq (=)\, \theta_0$	$\theta < \theta_0$	$W(\mathbf{x}_n) < \xi$	$\Phi\left[\frac{\sqrt{n}(\overline{x}_n-\theta_0)}{\sigma}\right]$

□

Relazione tra valore-p e test di ampiezza α

Consideriamo il caso di ipotesi nulla puntuale e supponiamo che i test basati su W siano di ampiezza α, con regione di rifiuto R_α e di accettazione A_α. Se $p(\mathbf{x}_n)$ è il valore-p basato su W si ha che

$$p_{\theta_0}(\mathbf{x}_n) < \alpha \qquad \Leftrightarrow \qquad \mathbf{x}_n \in R_\alpha$$

e

$$p_{\theta_0}(\mathbf{x}_n) \geq \alpha \qquad \Leftrightarrow \qquad \mathbf{x}_n \in A_\alpha$$

ovvero che un valore-p minore (maggiore) di α indica il rifiuto (accettazione) dell'ipotesi nulla nel test (basato sulla stessa statistica test) di ampiezza α. In questo caso il valore-p è una vera e propria statistica test.

Per verificare la prima delle precedenti relazioni, consideriamo ad esempio il caso in cui $R_\alpha = \{\mathbf{x}_n \in \mathcal{X}^n : W(\mathbf{x}_n) > k_\alpha\}$, dove k_α è il valore che garantisce che il test sia di ampiezza α. Si ha allora che

$$\begin{aligned} \mathbf{x}_n \in R_\alpha \quad &\Leftrightarrow \quad W(\mathbf{x}_n) > k_\alpha \\ &\Leftrightarrow \quad p_{\theta_0}(\mathbf{x}_n) = \mathbb{P}_{\theta_0}[W(\mathbf{X}_n) > W(\mathbf{x}_n)] \leq \mathbb{P}_{\theta_0}[W(\mathbf{X}_n) > k_\alpha] = \alpha. \end{aligned}$$

La seconda relazione si verifica in modo analogo.

6.28 Esempio (Modello normale, varianza incognita: t-test). Nel caso del modello normale $\mathrm{N}(\theta, \sigma^2)$ con varianza σ^2 incognita, la statistica test per i tre più comuni sistemi di ipotesi su θ è

$$W(\mathbf{X}_n) = \frac{\sqrt{n}(\overline{X}_n - \theta_0)}{S_n},$$

dove $S_n^2 = \sum_{i=1}^n (X_i - \overline{X}_n)^2/(n-1)$. Sotto ipotesi nulla $W(\mathbf{X}_n)$ ha distibruzione t di Student con $n-1$ gradi di libertà, la cui funzione di ripartizione è $\mathrm{F}_{n-1}(\cdot)$. La seguente tabella riassuntiva[5] riporta l'espressione del valore-p per i sistemi d'ipotesi più comuni.

H_0	H_1	rif. H_0	Valore-p
$\theta = \theta_0$	$\theta \neq \theta_0$	$\lvert W(\mathbf{x}_n)\rvert > \lvert\xi\rvert$	$2\left[1 - \mathrm{F}_{n-1}\left(\frac{\sqrt{n}\lvert\overline{x}_n - \theta_0\rvert}{S_n}\right)\right]$
$\theta \leq (=)\, \theta_0$	$\theta > \theta_0$	$W(\mathbf{x}_n) > \xi$	$1 - \mathrm{F}_{n-1}\left[\frac{\sqrt{n}(\overline{x}_n - \theta_0)}{S_n}\right]$
$\theta \geq (=)\, \theta_0$	$\theta < \theta_0$	$W(\mathbf{x}_n) < \xi$	$\mathrm{F}_{n-1}\left[\frac{\sqrt{n}(\overline{x}_n - \theta_0)}{S_n}\right]$

□

6.3.3 Ottimalità: test uniformemente più potenti

Generalizziamo ora al caso di ipotesi composte la nozione di ottimalità (massima potenza) considerata per il caso di ipotesi semplici. Dobbiamo ora ricordare che le probabilità di errore di I e II tipo (e quindi la potenza di un test) sono funzioni di θ e non semplici numeri, come nel caso di ipotesi semplici. Per confrontare due test è quindi necessario tenere conto del comportamento complessivo (in Θ) delle rispettive funzioni di potenza.

6.29 Definizione (Test uniformemente più potente). Dato il sistema di ipotesi composte

$$H_0 : \theta \in \Theta_0 \quad \text{vs.} \quad H_1 : \theta \in \Theta_1$$

il test W^*, con funzione di potenza η_{W^*}, si dice *uniformemente più potente* (UMP=*uniformly most powerful*) di ampiezza α se

$$\sup_{\theta \in \Theta_0} \eta_{W^*}(\theta) = \alpha \tag{6.22}$$

e se, per ogni test W con funzione di potenza η_W tale che $\sup_{\theta\in\Theta_0} \eta_W(\theta) = \alpha' \leq \alpha$, si ha

$$\eta_{W^*}(\theta) \geq \eta_W(\theta), \quad \forall \theta \in \Theta_1. \tag{6.23}$$

□

La definizione data richiede che la funzione di potenza di un test UMP assuma valori più elevati di quella di ogni altro test in Θ_1: si richiede quindi che, usando il test W^*, la probabilità di rifiutare correttamente l'ipotesi nulla (quando cioè questa è falsa), non sia mai inferiore a quella di qualsiasi altro test il cui massimo valore della funzione di probabilità di errore di I specie (ovvero della funzione di potenza in θ_0), pari a α', è inferiore o uguale a quello del test W^* (pari ad α). Ciò vuol dire che il comportamento del test UMP può anche essere peggiore in (anche tutto) Θ_0 di quello di altri test, ma è richiesto che sia uniformemente migliore (o equivalente) in Θ_1.
Si noti inoltre che, se le ipotesi sono semplici, la definizione introdotta coincide con quella di *test più potente* .

[5] Per i dettagli, vedere Casella-Berger (2002, p. 398).

Test UMP per sistemi di ipotesi unilaterali

Possiamo ora mostrare che, nel caso di sistemi di ipotesi composte unilaterali, sotto opportune condizioni è garantita l'esistenza di test UMP e che la loro determinazione è alquanto semplice. Introduciamo ora la proprietà che, se posseduta da un modello statistico, consente di individuare test UMP per ipotesi riguardanti il parametro di questo modello.

6.30 Definizione (Rapporto di verosimiglianza monotono). Il modello statistico $\{\mathcal{X}^n, f_n(\mathbf{x}_n;\theta), \theta \in \Theta\}$ (con $\Theta \subseteq \mathbb{R}^1$) ha *rapporto delle verosimiglianze monotono* se esiste una statistica T con valori in $\mathbb{R}^1$ tale che, comunque fissati $\theta_1 < \theta_2$, si ha:

$$\lambda_{21}(\mathbf{x}_n) = \frac{L(\theta_2, \mathbf{x}_n)}{L(\theta_1; \mathbf{x}_n)} = \frac{f_n(\mathbf{x}_n;\theta_2)}{f_n(\mathbf{x}_n;\theta_1)} = \varphi\left(T(\mathbf{x}_n)\right), \tag{6.24}$$

dove $\varphi(\cdot)$ è una funzione monotona non decrescente. □

6.31 Teorema (Famiglie esponenziali e rv monotono). Un modello statistico uniparametrico che sia famiglia esponenziale e con funzione di massa o di densità di probabilità

$$f_X(x;\theta) = h(x)\exp\{\omega(\theta)T(x) - B(\theta)\}$$

ha il rapporto delle verosimiglianze monotono rispetto alla statistica sufficiente $T_n(\mathbf{x}_n) = \sum_{i=1}^n T(X_i)$ se $w(\cdot)$ è una funzione non decrescente. □

Dimostrazione. Nel caso di una campione casuale si ha che

$$f_n(\mathbf{x}_n;\theta) = \prod_{i=1}^n f_X(x_i;\theta) = h_n(\mathbf{x}_n)\exp\{\omega(\theta)T_n(\mathbf{x}_n) - B_n(\theta)\},$$

dove $h_n(\mathbf{x}_n) = \prod_{i=1}^n h(x_i)$, $T_n(\mathbf{x}_n) = \sum_{i=1}^n T(x_i)$, $B_n(\theta) = nB(\theta)$ Per ogni coppia di valori θ_1, θ_2 tale che $\theta_1 < \theta_2$ si ha quindi che

$$\frac{f_n(\mathbf{x}_n;\theta_2)}{f_n(\mathbf{x}_n;\theta_1)} = \exp\{T_n(\mathbf{x}_n)[\omega(\theta_2) - \omega(\theta_1)] - B_n(\theta_2) + B_n(\theta_1)\}$$

che è funzione non decrescente di T_n se $\omega(\cdot) > 0$.
Possiamo ora enunciare il celebre teorema di Karlin-Rubin[6], che garantisce l'esistenza di test UMP nel caso di ipotesi unilaterali. Per la dimostrazione si rimanda a Piccinato (2009) o a Casella-Berger (2002).

6.32 Teorema (Karlin-Rubin). Se il modello statistico $\{\mathcal{X}^n, f_n(\mathbf{x}_n;\theta), \theta \in \Theta\}$ (con $\Theta \subseteq \mathbb{R}^1$) ha il rapporto delle verosimiglianze monotono rispetto a una statistica sufficiente T_n, si ha quanto segue.
(a) La classe dei test per il confronto delle ipotesi unilaterali

$$H_0 : \theta \leq \theta_0 \quad \text{vs.} \quad H_1 : \theta > \theta_0$$

basata sulla regione di rifiuto

$$R = \{\mathbf{x}_n \in \mathcal{X}^n : T_n(\mathbf{x}_n) \geq c\}, \qquad c \in \mathbb{R}, \tag{6.25}$$

è costituita da test uniformemente più potenti (di ampiezza dipendente dal valore c).
(b) La classe dei test per il confronto delle ipotesi unilaterali

$$H_0 : \theta \geq \theta_0 \quad \text{vs.} \quad H_1 : \theta < \theta_0.$$

[6]Karlin S. e Rubin H. (1956). The theory of decision procedures for distributions with monotone likelihood ratio. *Annals of Mathematical Statistics*, Vol. 27, N. 2, 272-299.

basata sulla regione di rifiuto

$$R = \{\mathbf{x}_n \in \mathcal{X}^n : \ T_n(\mathbf{x}_n) \leq c\}, \qquad c \in \mathbb{R}. \tag{6.26}$$

è costituita da test uniformemente più potenti (di ampiezza dipendente dal valore c). □

6.33 Osservazione.

1. Il test basato sulla regione (6.25) è UMP anche per il sistema di ipotesi $H_0 : \ \theta = \theta_0$ vs. $H_1 : \ \theta > \theta_0$; il test basato sulla regione (6.26) è UMP anche per il sistema di ipotesi $H_0 : \ \theta = \theta_0$ vs. $H_1 : \ \theta < \theta_0$.
2. Nel caso in cui il modello statistico ha rapporto delle verosimiglianze monotono *non crescente* in $T(\mathbf{x}_n)$, si accetta l'ipotesi nulla $H_0 : \theta \leq \theta_0$ se $T_n \geq c$ e, analogamente, si accetta l'ipotesi nulla $H_0 : \theta \geq \theta_0$ se $T_n \leq c$.

□

6.34 Esempio (Test sul valore atteso, modello normale). Si considerino il modello e i sistemi di ipotesi (a) e (b) dell'Esempio 6.21. In questo caso è semplice verificare che $\forall(\theta_1, \theta_2)$ con $\theta_1 < \theta_2$,

$$\lambda_{21}(\mathbf{x}_n) = \exp\left\{\overline{x}_n \frac{n}{\sigma^2}(\theta_2 - \theta_1) - \frac{n}{2\sigma^2}(\theta_2^2 - \theta_1^2)\right\},$$

che, essendo $\theta_2 > \theta_1$ è una funzione monotona di $T(\mathbf{x}_n) = \overline{x}_n$. Applicando il teorema di Karlin-Rubin si trovano pertanto le regioni di accettazione e rifiuto di H_0 ottenute già nell'Esempio 6.21. □

6.35 Esempio (Test sulla varianza, modello normale). Consideriamo il modello $\mathrm{N}(\theta_0, \sigma^2)$ (θ_0 noto) e il sistema di ipotesi $H_0 : \ \sigma^2 = \sigma_0^2$ vs. $H_1 : \ \sigma^2 > \sigma_0^2$, dove σ_0^2 è un valore noto per la varianza. Sappiamo che, dato un campione casuale, la statistica $S_0^2 = \sum(X_i - \theta_0)^2/n$ è sufficiente per il modello. Inoltre, trattandosi di famiglia esponenziale, il modello ha rapporto delle verosimiglianze monotono (non decrescente). Per il teorema di Karlin-Rubin la regione di rifiuto del test UMP è quindi

$$R = \{\mathbf{x}_n \in \mathcal{X}^n : \ S_0^2 > c\}.$$

Il test di ampiezza α si ottiene determinando il valore di c tale che $\mathbb{P}(R; H_0) = \alpha$. Osservando che, se è vera H_0,

$$\frac{nS_0^2}{\sigma_0^2} \sim \chi_n^2$$

e indicando con $\chi^2_{n;\epsilon}$ il percentile di livello ϵ di una v.a. χ_n^2, si ha che

$$\mathbb{P}(R; H_0) = \mathbb{P}_{\sigma_0^2}(S_0^2 > c) = \mathbb{P}_{\sigma_0^2}\left(\frac{nS_0^2}{\sigma_0^2} > c'\right) = \alpha$$

se

$$c' = \chi^2_{n;1-\alpha}.$$

La regione critica del test di ampiezza α è quindi

$$R = \{\mathbf{x}_n \in \mathcal{X}^n : \ \frac{nS_0^2}{\sigma_0^2} > \chi^2_{n;1-\alpha}\} = \{\mathbf{x}_n \in \mathcal{X}^n : \ S_0^2 > \frac{\sigma_0^2}{n}\chi^2_{n;1-\alpha}\}.$$

Il test determinato vale anche nel caso in cui l'ipotesi nulla è del tipo $\sigma^2 \leq \sigma_0^2$. Inoltre, in modo del tutto analogo si ottiene il test per la verifica del sistema di ipotesi $H_0 : \ \sigma^2 = \sigma_0^2$ vs. $H_1 : \ \sigma^2 < \sigma_0^2$ □

Non esistenza del test UMP: ipotesi bilaterali

Il teorema di Karlin-Rubin garantisce l'esistenza di test UMP solo nel caso di ipotesi composte. Il caso di ipotesi bilaterali $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$ resta pertanto escluso. È anzi possibile mostrare (si vedano gli esempi 8.3.19 e 8.3.20 in Casella-Berger (2002)) che per questo sistema di ipotesi non esistono test UMP. Per trovare test con proprietà di ottimalità è necessario aggiungere un ulteriore condizione.

6.36 Definizione (Test non distorti). Dato un generico sistema di ipotesi composte $H_0 : \theta \in \Theta_0$ vs. $H_1 : \theta \in \Theta_1$, un test W di ampiezza α si dice *non distorto* se

$$\eta_W(\theta) \geq \alpha, \qquad \theta \in \Theta_1. \tag{6.27}$$

□

La richiesta è quindi che la probabilità di rifiutare H_0 quando questa ipotesi è falsa non deve mai andare al di sotto di α, la massima probabilità di rifiutare l'ipotesi nulla quando questa è vera.

Si può dimostrare che, nel caso di test sul parametro di un modello che è una famiglia esponenziale uniparametrica, con statistica sufficiente $T_n(\mathbf{x}_n)$, dati due valori $c_1, c_2 \in \mathbb{R}$ (con $c_1 < c_2$), la classe di test non distorti basati sulle regioni di accettazione e di rifiuto del tipo

$$A = \{\mathbf{x}_n \in \mathcal{X}^n : c_1 < T_n(\mathbf{x}_n) < c_2\} \tag{6.28}$$

$$R = \{\mathbf{x}_n \in \mathcal{X}^n : T_n(\mathbf{x}_n) \leq c_1 \quad \text{oppure} \quad T(\mathbf{x}_n) \geq c_2\} \tag{6.29}$$

è costituita da test uniformemente più potenti tra i non distorti, con ampiezza determinata dai valori scelti per c_1 e c_2. Questi test sono definiti test UMPU (*uniformly most powerful unbiased*)

Nel caso del sistema di ipotesi

$$H_0 : \theta = \theta_0 \text{ vs. } H_1 : \theta \neq \theta_0$$

un test non distorto deve quindi avere un punto di minimo in $\theta = \theta_0$. Nel caso di funzione di potenza continua, la condizione necessaria perchè $\eta_W(\theta)$ abbia minimo in questo punto è che

$$\frac{d}{d\theta}\eta_W(\theta) = 0 \quad \text{per } \theta = \theta_0. \tag{6.30}$$

6.37 Esempio Utilizzando la condizione (6.30), si può ad esempio ottenere il test UMPU di ampiezza α per il test $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$ nel modello $\mathrm{N}(\theta, \sigma^2)$ (σ^2 noto). Si può verificare che il test che si ottiene coincide con quello ottenuto nell'esempio 6.23. (Vedi Piccinato es. 7.23, pp. 324-325). □

6.4 Test asintotici

L'implementazione di un test richiede che sia nota la distribuzione campionaria della v.a. $W(\mathbf{X}_n)$ sotto H_0. Ciò è necessario per determinare il valore della soglia che separa la regione di accettazione da quella di rifiuto, in modo tale che il test abbia ampiezza α prefissata. In generale, la conoscenza di $f_W(\cdot;\theta)$ consente di determinare la funzione di potenza del test, η_W. Quando la distribuzione di W non è nota, ma è possibile determinarne una approssimazione asintotica, possiamo derivare un test di ampiezza approssimativamente uguale al livello *nominale* α. L'uso di test asintotici è anche utile nel caso di problemi di verifica di ipotesi per parametri di modelli per v.a. discrete.

6.4.1 Test di Wald

Un metodo piuttosto comune per costruire un test asintotico per un parametro scalare θ di un modello statistico si basa su stimatori asintoticamente normali (vedi Capitolo 4). Sia quindi $(\widehat{\theta}_n,\ n \in \mathbb{N})$ una successione di stimatori asintoticamente normali di θ e $\widehat{\mathbb{V}_\theta(\widehat{\theta}_n)}$ la corrispondente successione di stimatori delle varianze di $\widehat{\theta}_n$, tale che $\forall \theta \in \Theta$, la successione

$$\frac{\widehat{\theta}_n - \theta}{\sqrt{\widehat{\mathbb{V}_\theta(\widehat{\theta}_n)}}} \xrightarrow{d} \mathrm{N}(0,1).$$

Se, ad esempio, consideriamo il sistema di ipotesi

$$H_0:\ \theta = \theta_0, \qquad \text{vs.} \qquad H_1:\ \theta \neq \theta_0,$$

il *test di Wald* di ampiezza (approssimativamente) pari ad α si basa sulla regione di rifiuto

$$R_{\widetilde{W}} = \{\mathbf{x}_n \in \mathcal{X}^n:\ |\widetilde{W}(\mathbf{x}_n)| > z_{1-\frac{\alpha}{2}}\}$$

dove

$$\widetilde{W}(\mathbf{x}_n) = \frac{\widehat{\theta}_n - \theta_0}{\sqrt{\widehat{\mathbb{V}_\theta(\widehat{\theta}_n)}}} \tag{6.31}$$

è la *statistica test di Wald*. Supponendo vera l'ipotesi nulla ($\theta = \theta_0$), si ha che

$$\widetilde{W}(\mathbf{x}_n) \dot{\sim} \mathrm{N}(0,1).$$

Per n finito, scelta un'ampiezza α, il test $\widetilde{W}$ ha ampiezza $\widetilde{\alpha}_n$ solo approssimativamente pari ad α. Inoltre, per $n \to \infty$, $\widetilde{\alpha}_n \to \alpha$. Infatti, indicando con Z una v.a. con distribuzione $\mathrm{N}(0,1)$, abbiamo che, per n che tende a $+\infty$,

$$\widetilde{\alpha}_n = \mathbb{P}_{\theta_0}(R_{\widetilde{W}}) = \mathbb{P}_{\theta_0}(|\widetilde{W}| > z_{1-\frac{\alpha}{2}}) \to \mathbb{P}(|Z| > z_{1-\frac{\alpha}{2}}) = \alpha.$$

Una versione alternativa del test di Wald si basa sulla statistica test

$$\widetilde{W}_0(\mathbf{x}_n) = \frac{\widehat{\theta}_n - \theta_0}{\sqrt{\mathbb{V}_\theta(\widehat{\theta}_n)|_{\theta=\theta_0}}} = \frac{\widehat{\theta}_n - \theta_0}{\mathrm{sd}(\theta_0)} \tag{6.32}$$

dove $\mathbb{V}_\theta(\widehat{\theta}_n)|_{\theta=\theta_0}$ indica la varianza di $\widehat{\theta}_n$ calcolata in $\theta = \theta_0$, $\mathrm{sd}(\theta_0) = \sqrt{\mathbb{V}_\theta(\widehat{\theta}_n)|_{\theta=\theta_0}}$.
La funzione di potenza per questo test è

$$\begin{aligned}
\eta_{\widetilde{W}_0}(\theta) &= \mathbb{P}_\theta(|\widetilde{W}_0| > z_{1-\frac{\alpha}{2}}) = 1 - \mathbb{P}_\theta(|\widetilde{W}_0| < z_{1-\frac{\alpha}{2}}) \\
&= 1 - \mathbb{P}_\theta\left(-z_{1-\frac{\alpha}{2}} < \frac{\widehat{\theta}_n - \theta_0}{\mathrm{sd}(\theta_0)} < z_{1-\frac{\alpha}{2}}\right) \\
&= 1 - \mathbb{P}_\theta\left(\frac{\theta_0 - \theta}{\mathrm{sd}(\theta)} - z_{1-\frac{\alpha}{2}}\frac{\mathrm{sd}(\theta_0)}{\mathrm{sd}(\theta)} < \frac{\widehat{\theta}_n - \theta}{\mathrm{sd}(\theta)} < \frac{\theta_0 - \theta}{\mathrm{sd}(\theta)} + z_{1-\frac{\alpha}{2}}\frac{\mathrm{sd}(\theta_0)}{\mathrm{sd}(\theta)}\right) \\
&\simeq 1 - \mathbb{P}_\theta\left(\frac{\theta_0 - \theta}{\mathrm{sd}(\theta)} - z_{1-\frac{\alpha}{2}}\frac{\mathrm{sd}(\theta_0)}{\mathrm{sd}(\theta)} < Z < \frac{\theta_0 - \theta}{\mathrm{sd}(\theta)} + z_{1-\frac{\alpha}{2}}\frac{\mathrm{sd}(\theta_0)}{\mathrm{sd}(\theta)}\right),
\end{aligned}$$

dove $\mathrm{sd}(\theta) = \sqrt{\mathbb{V}_\theta(\widehat{\theta}_n)}$. Pertanto

$$\eta_{\widetilde{W}_0}(\theta) \simeq \quad 1 - \Phi\left(\tfrac{\theta_0-\theta}{\mathrm{sd}(\theta)} - z_{1-\frac{\alpha}{2}}\tfrac{\mathrm{sd}(\theta_0)}{\mathrm{sd}(\theta)}\right) + \Phi\left(\tfrac{\theta_0-\theta}{\mathrm{sd}(\theta)} + z_{1-\frac{\alpha}{2}}\tfrac{\mathrm{sd}(\theta_0)}{\mathrm{sd}(\theta)}\right).$$

In modo del tutto analogo si ottengono i test di Wald e le funzioni di potenza nel caso di ipotesi unilaterali. La seguente tabella riassume gli elementi fondamentali dei test di Wald $\widetilde{W}_0$ di ampiezza (approsimativamente uguale a) α per i principali sistemi di ipotesi.

H_0	H_1	condiz. di rif. di H_0	Funzione di potenza
$\theta = \theta_0$	$\theta \neq \theta_0$	$\lvert\widetilde{W}_0(\mathbf{x}_n)\rvert > z_{1-\alpha/2}$	$1 - \Phi\left(\tfrac{\theta_0-\theta}{\mathrm{sd}(\theta)} + z_{1-\frac{\alpha}{2}}\tfrac{\mathrm{sd}(\theta_0)}{\mathrm{sd}(\theta)}\right) + \Phi\left(\tfrac{\theta_0-\theta}{\mathrm{sd}(\theta)} - z_{1-\frac{\alpha}{2}}\tfrac{\mathrm{sd}(\theta_0)}{\mathrm{sd}(\theta)}\right)$
$\theta \leq \theta_0$	$\theta > \theta_0$	$\widetilde{W}_0(\mathbf{x}_n) > z_{1-\alpha}$	$1 - \Phi\left(\tfrac{\theta_0-\theta}{\mathrm{sd}(\theta)} + z_{1-\alpha}\tfrac{\mathrm{sd}(\theta_0)}{\mathrm{sd}(\theta)}\right)$
$\theta \geq \theta_0$	$\theta < \theta_0$	$\widetilde{W}_0(\mathbf{x}_n) < -z_{1-\alpha}$	$\Phi\left(\tfrac{\theta_0-\theta}{\mathrm{sd}(\theta)} + z_{\alpha}\tfrac{\mathrm{sd}(\theta_0)}{\mathrm{sd}(\theta)}\right)$

6.38 Esempio (Modello bernoulliano). Consideriamo il modello di Bernoulli di parametro θ e uno dei sistemi di ipotesi della tabella precedente. In questo caso, per un campione casuale di dimensione n, lo stimatore di massima verosimiglianza è $\widehat{\theta}_n = \overline{X}_n$ e $\mathbb{V}_\theta(\widehat{\theta}_n) = \theta(1-\theta)/n$. Si ha quindi che $\widehat{\mathbb{V}_\theta(\widehat{\theta}_n)} = \overline{x}_n(1-\overline{x}_n)/n$ e $\mathbb{V}_\theta(\widehat{\theta}_n)|_{\theta=\theta_0} = \theta_0(1-\theta_0)/n$. Pertanto le statistiche test di Wald (6.31) e (6.32) sono

$$\widetilde{W}(\mathbf{X}_n) = \frac{\sqrt{n}(\overline{X}_n - \theta_0)}{\sqrt{\overline{X}_n(1-\overline{X}_n)}} \qquad \text{e} \qquad \widetilde{W}_0(\mathbf{X}_n) = \frac{\sqrt{n}(\overline{X}_n - \theta_0)}{\sqrt{\theta_0(1-\theta_0)}}.$$

□

6.39 Esempio (Modello di Poisson). Consideriamo il modello di Poisson di parametro θ e uno dei sistemi di ipotesi della tabella precedente. In questo caso, per un campione casuale di dimensione n, lo stimatore di massima verosimiglianza è $\widehat{\theta}_n = \overline{X}_n$ e $\mathbb{V}_\theta(\widehat{\theta}_n) = \theta/n$. Si ha quindi che $\widehat{\mathbb{V}_\theta(\widehat{\theta}_n)} = \overline{x}_n/n$ e $\mathbb{V}_\theta(\widehat{\theta}_n)|_{\theta=\theta_0} = \theta_0/n$. Pertanto le statistiche test di Wald (6.31) e (6.32) sono

$$\widetilde{W}(\mathbf{X}_n) = \frac{\sqrt{n}(\overline{X}_n - \theta_0)}{\sqrt{\overline{X}_n}} \qquad \text{e} \qquad \widetilde{W}_0(\mathbf{X}_n) = \frac{\sqrt{n}(\overline{X}_n - \theta_0)}{\sqrt{\theta_0}}.$$

□

6.40 Esempio (Modello uniforme). Consideriamo un campione casuale da un modello uniforme in $[0, \theta]$ e uno dei sistemi di ipotesi previsti dalla tabella precedente. Abbiamo visto che, in questo caso, lo stimatore dei momenti è $\widehat{\theta}_m(\mathbf{X}_n) = 2\overline{X}_n$, per il quale $2\overline{X}_n \dot{\sim} \mathrm{N}(\theta, \theta^2/(3n))$. Pertanto le statistiche test di Wald (6.31) e (6.32) sono

$$\widetilde{W}(\mathbf{X}_n) = \frac{\sqrt{3n}(2\overline{X}_n - \theta_0)}{2\overline{X}_n} \qquad \text{e} \qquad \widetilde{W}_0(\mathbf{X}_n) = \frac{\sqrt{3n}(2\overline{X}_n - \theta_0)}{2\theta_0}.$$

□

Test di Wald basato su stimatori di massima verosimiglianza
Nel caso in cui la successione considerata è quella degli stimatori di massima verosimiglianza di θ, a denominatore della statistica $\widetilde{W}$ possiamo considerare una stima della varianza asintotica di $\widehat{\theta}_n$ che, come sappiamo, coincide con l'inverso dell'informazione di Fisher, oppure il valore della varianza asintotica calcolato in θ_0. Le due versione del test di Wald che si ottengono sono:

$$\widetilde{W}(\mathbf{X}_n) = \frac{\widehat{\theta}_n - \theta_0}{\sqrt{\widehat{I_n(\theta)}^{-1}}} \qquad \textbf{e} \qquad \widetilde{W}_0(\mathbf{X}_n) = \frac{\widehat{\theta}_n - \theta_0}{\sqrt{I_n(\theta_0)^{-1}}}. \tag{6.33}$$

Ricordare che la stima della varianza asintotica di $\widehat{\theta}_n$ può essere $[I_n(\widehat{\theta}_n)]^{-1}$ (inverso dell'informazione attesa calcolata in $\widehat{\theta}_n$) oppure $[\mathcal{I}_n(\mathbf{x}_n)]^{-1}$ (inverso dell'informazione osservata).

6.41 Esempio (Modello beta). Consideriamo un campione casuale di dimensione n da un modello Beta$(\theta, 1)$, $x \in (0,1)$, $\theta > 0$. Sappiamo che in questo caso lo smv è $\widehat{\theta}_n = -n/(\sum_{i=1}^n \ln X_i)$, che l'informazione attesa è $I_n(\theta) = n/\theta^2$ e l'informazione osservata coincide con $I_n(\widehat{\theta}_n)$. Le statistiche test di Wald (6.31) e (6.32) sono

$$\widetilde{W}(\mathbf{X}_n) = \frac{\sqrt{n}(\widehat{\theta}_n - \theta_0)}{\widehat{\theta}_n} \qquad \text{e} \qquad \widetilde{W}_0(\mathbf{X}_n) = \frac{\sqrt{n}(\widehat{\theta}_n - \theta_0)}{\theta_0}.$$

□

Test di Wald per ipotesi su $g(\theta)$
Il test di Wald può essere anche utilizzato per la verifica di ipotesi su funzioni $g(\theta)$ di θ (nelle condizioni di regolarità previste), utilizzando la statistica

$$\widetilde{W}_g(\mathbf{X}_n) = \frac{g(\widehat{\theta}_n) - g(\theta_0)}{\sqrt{g'(\widehat{\theta}_n)^2 \widehat{I_n(\theta)}^{-1}}}. \tag{6.34}$$

6.42 Esempio (Modello di Poisson). Consideriamo il modello Pois(θ) e supponiamo di voler effettuare un test su $g(\theta) = \mathbb{P}_\theta(X = 0) = e^{-\theta}$. In precedenza abbiamo visto che lo stimatore di massima verosimiglianza di $g(\theta)$ è $\widehat{g(\theta)} = g(\widehat{\theta}_n) = e^{-\overline{X}_n}$, $I_n(\theta) = n/\theta$ e $g'(\theta) = -e^{-\theta}$. Pertanto la varianza asintotica di $g(\widehat{\theta}_n)$ è $v_a[g(\widehat{\theta}_n)] = g'(\theta)^2 I_n(\theta)^{-1} = e^{-2\theta}\theta/n$ e inoltre

$$g(\widehat{\theta}_n) = e^{-\overline{X}_n} \dot{\sim} \mathrm{N}\left(e^{-\theta}, e^{-2\theta}\frac{\theta}{n}\right).$$

Lo stimatore della varianza asintotica di $g(\widehat{\theta}_n)$ è $\widehat{v_a[g(\widehat{\theta}_n)]} = g'(\widehat{\theta}_n)^2 \mathcal{I}_n(\mathbf{X}_n)^{-1} = e^{-2\overline{X}_n}\overline{X}_n/n$. Pertanto la statistica test di Wald è in questo caso

$$\widetilde{W}_g(\mathbf{X}_n) = \frac{\sqrt{n}\left(e^{-\overline{X}_n} - e^{-\theta_0}\right)}{\sqrt{e^{-2\overline{X}_n}\overline{X}_n}}.$$

Analogamente al caso semplice possiamo anche considerare la statistica

$$\widetilde{W}_{g,0}(\mathbf{X}_n) = \frac{\sqrt{n}\left(e^{-\overline{X}_n} - e^{-\theta_0}\right)}{\sqrt{e^{-2\theta_0}\theta_0}}.$$

□

6.4.2 Test di Wilks

L'implementazione del test del rapporto delle verosimiglianze richiede che sia nota la distribuzione di probabilità della statistica $\lambda_{01}^m(\mathbf{X}_n)$ sotto H_0. Ciò serve, infatti, per determinare il valore della soglia k_α che specifica il test di ampiezza α. Fortunatamente, sotto opportune condizioni, il rapporto delle verosimiglianze possiede alcune proprietà asintotiche che consentono di risolvere, almeno approssimativamente, questo problema. Diamo qui solo alcuni cenni di questi risultati, dovuti a Wilks (1962). Consideriamo il caso multiparametrico:

$$\boldsymbol{\theta} = (\boldsymbol{\theta}_r, \boldsymbol{\theta}_s),$$

dove

$$(\boldsymbol{\theta}_r, \boldsymbol{\theta}_s) = (\theta_1, \ldots, \theta_r, \theta_{r+1}, \ldots, \theta_{r+s})$$

e il sistema di ipotesi

$$H_0 : \boldsymbol{\theta}_r = \boldsymbol{\theta}_{r0} \qquad \text{vs.} \qquad H_1 : \boldsymbol{\theta}_r \neq \boldsymbol{\theta}_{r0}$$

ovvero

$$H_0 : \theta_1 = \theta_{10}, \ldots, \theta_r = \theta_{r0}, \qquad \text{vs.} \qquad H_1 : \theta_1 \neq \theta_{10}, \ldots, \theta_r \neq \theta_{r0}.$$

In questo caso

$$\lambda_{01}^m(\mathbf{X}_n) = \frac{L(\boldsymbol{\theta}_{r,0}, \widehat{\widehat{\boldsymbol{\theta}}}_s; \mathbf{X}_n)}{L(\widehat{\boldsymbol{\theta}}_r, \widehat{\boldsymbol{\theta}}_s; \mathbf{X}_n)}.$$

Sotto le condizioni di regolarità che garantiscono le proprietà asintotiche degli stimatori di massima verosimiglianza, si ha che

$$-2 \ln \lambda_{01}^m(\mathbf{X}_n) \dot{\sim} \chi_r^2 \tag{6.35}$$

dove, ricordiamo, $\widehat{\boldsymbol{\theta}}_r, \widehat{\boldsymbol{\theta}}_s$ sono i vettori delle stime di mv di $\boldsymbol{\theta}$ e $\widehat{\widehat{\boldsymbol{\theta}}}_s$ il vettore della stime di mv del parametro di disturbo $\boldsymbol{\theta}_s$ sotto H_0. Il test (approssimato) ha quindi regione di rifiuto

$$R = \{\mathbf{x}_n \in \mathcal{X}^n : \lambda_{01}^m(\mathbf{x}_n) \leq k\} = \{\mathbf{x}_n \in \mathcal{X}^n : -2 \ln \lambda_{01}^m(\mathbf{x}_n) > k\}, \qquad k > 0.$$

Il test di ampiezza α si ottiene determinando k in modo tale che

$$\sup_{\boldsymbol{\theta} \in \Theta_0} \mathbb{P}_{\boldsymbol{\theta}}(-2 \ln \lambda_{01}^m(\mathbf{x}_n) > k) = \alpha$$

ovvero ponendo

$$k = k_\alpha = \chi^2_{r;1-\alpha}$$

dove $\chi^2_{r;1-\alpha}$ è il percentile di livello $1-\alpha$ della v.a. χ_r^2.

Nel caso in cui si ha un unico parametro incognito,

$$-2 \ln \lambda_{01}^m(\mathbf{X}_n) \dot{\sim} \chi_1^2.$$

6.43 Esempio (Modello di Poisson)[7]. Nel caso del modello di Poisson, per il sistema di ipotesi ($\theta = \theta_0$ vs. $\theta \neq \theta_0$) si ottiene

$$-2 \ln \lambda_{01}^m(\mathbf{x}_n) = -2 \ln \left(\frac{e^{-n\theta_0} \theta_0^{\sum_{i=1}^n x_i}}{e^{-n\widehat{\theta}_{mv}} \widehat{\theta}_{mv}^{\sum_{i=1}^n x_i}} \right) = 2n \left[(\theta_0 - \widehat{\theta}_{mv}) - \widehat{\theta}_{mv} \ln(\theta_0 / \widehat{\theta}_{mv}) \right],$$

con $\widehat{\theta}_{mv} = \overline{x}_n$. Rifiutiamo H_0 se $-2 \ln \lambda_{01}^m > \chi^2_{1,1-\alpha}$. □

[7] Esempio tratto da Casella-Berger (2002, p. 489)

6.5 Test per i parametri del modello normale

Riassumiamo in questa sezione i principali test di ampiezza α per i parametri del modello $\mathrm{N}(\mu, \sigma^2)$.

6.5.1 Test per il valore atteso

Varianza nota

H_0	H_1	condiz. di rif. di H_0	Valore-p
$\mu = \mu_0$	$\mu \neq \mu_0$	$\lvert W(\mathbf{x}_n)\rvert > \lvert z_{1-\frac{\alpha}{2}}\rvert$	$2\left[1 - \Phi\left(\frac{\sqrt{n}\lvert\overline{x}_n - \mu_0\rvert}{\sigma}\right)\right]$
$\mu \leq (=)\, \mu_0$	$\mu > \mu_0$	$W(\mathbf{x}_n) > z_{1-\alpha}$	$1 - \Phi\left[\frac{\sqrt{n}(\overline{x}_n - \mu_0)}{\sigma}\right]$
$\mu \geq (=)\, \mu_0$	$\mu < \mu_0$	$W(\mathbf{x}_n) < z_{\alpha}$	$\Phi\left[\frac{\sqrt{n}(\overline{x}_n - \mu_0)}{\sigma}\right]$

- Statistica test: $W(\mathbf{X}_n) = \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{\sigma}$.
- Distribuzione di $W(\mathbf{X}_n)$ per $\mu = \mu_0$: $\mathrm{N}(0, 1)$.
- z_ϵ: percentile di livello ϵ della v.a. $\mathrm{N}(0, 1)$.
- Ottimalità: i test 2 e 3 sono UMP, il test 1 è UMPU.
- I tre test coincidono con i test del rvm.

Varianza incognita

H_0	H_1	condiz. di rif. di H_0	Valore-p
$\mu = \mu_0$	$\mu \neq \mu_0$	$\lvert W(\mathbf{x}_n)\rvert > \lvert t_{n-1;1-\frac{\alpha}{2}}\rvert$	$2\left[1 - \mathrm{F}_{n-1}\left(\frac{\sqrt{n}\lvert\overline{x}_n - \mu_0\rvert}{S_n}\right)\right]$
$\mu \leq (=)\, \mu_0$	$\mu > \mu_0$	$W(\mathbf{x}_n) > t_{n-1;1-\alpha}$	$1 - \mathrm{F}_{n-1}\left[\frac{\sqrt{n}(\overline{x}_n - \mu_0)}{S_n}\right]$
$\mu \geq (=)\, \mu_0$	$\mu < \mu_0$	$W(\mathbf{x}_n) < -t_{n-1;1-\alpha}$	$\mathrm{F}_{n-1}\left[\frac{\sqrt{n}(\overline{x}_n - \mu_0)}{S_n}\right]$

- Statistica test: $W(\mathbf{X}_n) = \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{S_n}$, dove $S_n^2 = \frac{\sum_{i=1}^n (X_i - \overline{X}_n)^2}{n-1}$.
- Distribuzione di $W(\mathbf{X}_n)$ per $\mu = \mu_0$: t_{n-1}.
- $t_{n-1;\epsilon}$: percentile di livello ϵ della v.a. t di Student con $n-1$ gradi di libertà.
- Ottimalità: i tre test sono UMPU (si veda Lehmann (1986)).

6.5.2 Test per la varianza

Valore atteso noto

H_0	H_1	condiz. di rif. di H_0
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$W(\mathbf{x}_n) > \chi^2_{n;1-\frac{\alpha}{2}}$ oppure $W(\mathbf{x}_n) < \chi^2_{n;\frac{\alpha}{2}}$
$\sigma^2 \leq (=) \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$W(\mathbf{x}_n) > \chi^2_{n;1-\alpha}$
$\sigma^2 \geq (=) \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$W(\mathbf{x}_n) < \chi^2_{n;\alpha}$

- Statistica test: $W(\mathbf{X}_n) = \frac{nS_0^2}{\sigma_0^2}$, dove $S_0^2 = \frac{\sum_{i=1}^n (X_i - \mu)^2}{n}$.
- Distribuzione di $W(\mathbf{X}_n)$ per $\sigma^2 = \sigma_0^2$: χ^2_n.
- $\chi^2_{n;\epsilon}$: percentile di livello ϵ della v.a. chi quadrato con n gradi di libertà.
- Ottimalità: i test 2 e 3 sono UMPU, il test 1 è approssimativamente UMPU (si veda Lehmann (1986)).

Valore atteso incognito

H_0	H_1	condiz. di rif. di H_0
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$W(\mathbf{x}_n) > \chi^2_{n-1;1-\frac{\alpha}{2}}$ oppure $W(\mathbf{x}_n) < \chi^2_{n-1;\frac{\alpha}{2}}$
$\sigma^2 \leq (=) \sigma_0$	$\sigma^2 > \sigma_0^2$	$W(\mathbf{x}_n) > \chi^2_{n-1;1-\alpha}$
$\sigma^2 \geq (=) \sigma_0$	$\sigma^2 < \sigma_0^2$	$W(\mathbf{x}_n) < \chi^2_{n-1;\alpha}$

- Statistica test: $W(\mathbf{X}_n) = \frac{(n-1)S_n^2}{\sigma_0^2}$, dove $S_n^2 = \frac{\sum_{i=1}^n (X_i - \overline{X}_n)^2}{n-1}$.
- Distribuzione di $W(\mathbf{X}_n)$ per $\sigma^2 = \sigma_0^2$: χ^2_{n-1}.
- $\chi^2_{n-1;\epsilon}$: percentile di livello ϵ della v.a. chi quadrato con $n-1$ gradi di libertà.

6.5.3 Test con due campioni: confronto tra valori attesi di modelli normali

Consideriamo due campioni casuali indipendenti, $\boldsymbol{x}_{n_a}, \boldsymbol{x}_{n_b}$, costituiti rispettivamente da n_a e n_b osservazioni i.i.d., provenienti da due popolazioni normali di parametri (μ_a, σ_a^2) e (μ_b, σ_b^2):

$$X_i^a \sim \mathrm{N}(\mu_a, \sigma_a^2) \qquad \text{e} \qquad X_i^b \sim \mathrm{N}(\mu_b, \sigma_b^2)$$

Per risultati noti relativi a v.a. normali, si ha che

$$\overline{X}_{n_a} - \overline{X}_{n_b} \sim \mathrm{N}\left(\mu_a - \mu_b, \frac{\sigma_a^2}{n_a} + \frac{\sigma_b^2}{n_b}\right)$$

Per confrontare tra loro i valori attesi μ_a e μ_b possiamo considerare il parametro

$$\mu = \mu_a - \mu_b.$$

ed effettuare test su tale quantità.

Varianze note

H_0	H_1	condiz. di rif. di H_0
$\mu = \delta_0$	$\mu \neq \delta_0$	$\lvert W(\boldsymbol{x}_{n_a}, \boldsymbol{x}_{n_b})\rvert > \lvert z_{1-\frac{\alpha}{2}}\rvert$
$\mu \leq (=)\, \delta_0$	$\mu > \delta_0$	$W(\boldsymbol{x}_{n_a}, \boldsymbol{x}_{n_b}) > z_{1-\alpha}$
$\mu \geq (=)\, \delta_0$	$\mu < \delta_0$	$W(\boldsymbol{x}_{n_a}, \boldsymbol{x}_{n_b}) < z_{\alpha}$

- Statistica test: $W(\boldsymbol{X}_{n_a}, \boldsymbol{X}_{n_b}) = \frac{(\overline{X}_{n_a} - \overline{X}_{n_b}) - \delta_0}{\sqrt{\frac{\sigma_a^2}{n_a} + \frac{\sigma_b^2}{n_b}}}$
- $W(\boldsymbol{X}_{n_a}, \boldsymbol{X}_{n_b}) \sim \mathrm{N}(0,1)$ se $\mu = \delta_0$.

Varianze incognite e uguali

H_0	H_1	condiz. di rif. di H_0
$\mu = \delta_0$	$\mu \neq \delta_0$	$\lvert W(\boldsymbol{x}_{n_a}, \boldsymbol{x}_{n_b})\rvert > \lvert t_{n_a+n_b-2;1-\frac{\alpha}{2}}\rvert$
$\mu \leq (=)\, \delta_0$	$\mu > \delta_0$	$W(\boldsymbol{x}_{n_a}, \boldsymbol{x}_{n_b}) > t_{n_a+n_b-2;1-\alpha}$
$\mu \geq (=)\, \delta_0$	$\mu < \delta_0$	$W(\boldsymbol{x}_{n_a}, \boldsymbol{x}_{n_b}) < t_{n_a+n_b-2;\alpha}$

- Statistica test: $W(\boldsymbol{X}_{n_a}, \boldsymbol{X}_{n_b}) = \frac{(\overline{X}_{n_a} - \overline{X}_{n_b}) - \delta_0}{S_p\sqrt{\frac{1}{n_a} + \frac{1}{n_b}}}$, $\qquad S_p^2 = \frac{(n_a-1)S_{n_a}^2 + (n_b-1)S_{n_b}^2}{n_a+n_b-2}$.
- Distribuzione di $W(\boldsymbol{X}_{n_a}, \boldsymbol{X}_{n_b})$ se $\mu = \delta_0$: $t_{n_a+n_b-2}$.
- $t_{n_a+n_b-2;\epsilon}$: percentile di livello ϵ della v.a. t con $n_a + n_b - 2$ gradi di libertà.

6.5.4 Test con due campioni: confronto tra varianze di modelli normali

Consideriamo, come nel caso precedente, due popolazioni normali, ma supponiamo di essere ora interessati a effettuare un test per il confronto tra le varianze incognite σ_a^2 e σ_b^2. Consideriamo quindi il parametro

$$\psi = \frac{\sigma_a^2}{\sigma_b^2}.$$

Valori attesi noti

La statistica test utilizzata per verificare ipotesi su ψ (ad esempio l'ipotesi $\psi = \psi_0$, dove ψ_0 è uno specifico valore del parametro) è

$$\frac{S_{0_a}^2}{S_{0_b}^2}/\psi_0$$

dove $S^2_{0_i} = \sum_{j=1}^{n_i}(X^i_j - \mu_i)^2/n_i,\ i = a, b$. Tale v.a., sotto l'ipotesi che $\psi = \psi_0$, ha distribuzione F_{n_a,n_b}[8].

H_0	H_1	condiz. di rif. di H_0
$\psi = \psi_0$	$\psi \neq \psi_0$	$W(\boldsymbol{x}_{n_a}, \boldsymbol{x}_{n_b}) > F_{n_a,n_b;1-\frac{\alpha}{2}}$ oppure $W(\boldsymbol{x}_{n_a}, \boldsymbol{x}_{n_b}) < F_{n_a,n_b;\frac{\alpha}{2}}$
$\psi \leq (=) \psi_0$	$\psi > \psi_0$	$W(\boldsymbol{x}_{n_a}, \boldsymbol{x}_{n_b}) > F_{n_a,n_b;1-\alpha}$
$\psi \geq (=) \psi_0$	$\psi < \psi_0$	$W(\boldsymbol{x}_{n_a}, \boldsymbol{x}_{n_b}) < F_{n_a,n_b;\alpha}$

- Statistica test: $W(\boldsymbol{X}_{n_a}, \boldsymbol{X}_{n_b}) = \frac{S^2_{0_a}}{S^2_{0_b}}$
- Distribuzione di $W(\mathbf{X}^a_n, \mathbf{X}^b_n)$ per $\psi = \psi_0$: F_{n_a,n_b}.
- $F_{n_a,n_b;\epsilon}$: percentile di livello ϵ della v.a. F con (n_a, n_b) gradi di libertà.

Valori attesi incogniti

La statistica test utilizzata per verificare ipotesi su ψ è

$$\frac{S^2_{n_a}}{S^2_{n_b}}/\psi_0$$

dove $S^2_{n_i} = \sum_{j=1}^{n_i}(X^i_j - \overline{X}_{n_i})^2/(n_i - 1),\ i = a, b$ e che, sotto l'ipotesi che $\psi = \psi_0$, ha distribuzione F_{n_a-1,n_b-1}.

H_0	H_1	condiz. di rif. di H_0
$\psi = \psi_0$	$\psi \neq \psi_0$	$W(\boldsymbol{x}_{n_a}, \boldsymbol{x}_{n_b}) > F_{n_a-1,n_b-1;1-\frac{\alpha}{2}}$ oppure $W(\boldsymbol{x}_{n_a}, \boldsymbol{x}_{n_b}) < F_{n_a-1,n_b-1;\frac{\alpha}{2}}$
$\psi \leq (=) \psi_0$	$\psi > \psi_0$	$W(\boldsymbol{x}_{n_a}, \boldsymbol{x}_{n_b}) > F_{n_a-1,n_b-1;1-\alpha}$
$\psi \geq (=) \psi_0$	$\psi < \psi_0$	$W(\boldsymbol{x}_{n_a}, \boldsymbol{x}_{n_b}) < F_{n_a-1,n_b-1;\alpha}$

- Statistica test: $W(\mathbf{X}^a_n, \mathbf{X}^b_n) = \frac{S^2_{n_a}}{S^2_{n_b}}$
- Distribuzione di $W(\mathbf{X}^a_n, \mathbf{X}^b_n)$ per $\psi = \psi_0$: F_{n_a-1,n_b-1}.
- $F_{n_a-1,n_b-1;\epsilon}$: percentile di livello ϵ della v.a. F con $(n_a - 1, n_b - 1)$ gradi di libertà.

[8] Si ricordi infatti che, nel campionamento da popolazioni normali,

$$\frac{n_i S^2_{0_i}}{\sigma^2_i} \sim \chi^2_{n_i}, \qquad i = a, b$$

e che il rapporto di due v.a. chi quadrato tra loro indipendenti, ciascuna divisa per i rispettivi gradi di libertà, definisce una v.a. con distribuzione F di Fisher con gradi di libertà coincidenti con quelli delle due v.a. chi quadrato.