Statistica II
Appunti del corso

Claudio Ortelli

Semestre Autunnale
2012

Indice

Capitolo 1

Introduzione

1.1 Ricapitolazione

Nel corso introduttivo di Statistica I sono stati trattati i concetti fondamentali dell'analisi statistica e della teoria della probabilità. In particolare avete visto
Capitolo 1: introduzione

1. Popolazione di riferimento di un'analisi statistica.
2. Il campione.

Capitolo 2: elementi di statistica descrittiva

1. Le distribuzioni di frequenze (assolute o relative), istogrammi.

Capitolo 3: misure empiriche di centralità e variabilità

1. Indicatori di centralità (moda, media e mediana) e variabilità (range, intervallo interquartile, deviazione standard). Regole di sommatoria e relativa notazione.

$$\overline{x},\ \sum_{i \text{ pari}} x_i,\ S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Capitolo 4: Elementi di probabilità I

1. Spazio campionario Ω: insieme di tutti gli esiti di un esperimento statistico. Evento $E \subset \Omega$: un particolare sottoinsieme di Ω. E^c: il complemento dell'evento E. Prime regole di calcolo:

$$P(A) + P(A^c) = 1.$$

Capitolo 5: Elementi di probabilità II

1. Intersezione di eventi. $A \cap B$: l'insieme degli esiti dell'esperimento che appartengono sia all'evento A che all'evento B.

2. Eventi indipendenti: A e B sono indipendenti se vale la seguente condizione
$$P(A \cap B) = P(A)P(B).$$

3. Probabilità condizionata
$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}.$$

4. Unione di eventi. $A \cup B$: l'insieme degli esiti dell'esperimento che appartengono all'evento A o all'evento B o ad entrambi.

5. Eventi mutualmente esclusivi. A e B sono mutualmente esclusivi se non hanno esiti in comune, ovvero se $A \cap B = \emptyset$.

6. Probabilità di unioni di eventi A e B mutualmente esclusivi:
$$P(A \cup B) = P(A) + P(B)\ .$$

7. Probabilità di unioni di eventi:
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)\ .$$

8. Diagrammi ad albero per la rappresentazione dei possibili esiti di un esperimento e per il calcolo delle probabilità corrispondenti.

9. Teorema di Bayes.

Capitolo 6: Introduzione alle variabili aleatorie

1. Esperimento statistico: procedimento di osservazione di un dato fenomeno. Variabile aleatoria ($V.A.$) X. Definizione di variabili aleatorie discrete e continue. Distribuzioni e funzioni di probabilità.

2. Valore atteso e varianza di una variabile $V.A$ discreta X:
$$\begin{aligned} \mu &= \sum_{i=1}^{n} x_i p(x_i) \\ \sigma^2 &= \sum_{i=1}^{n} (x_i - \mu)^2 p(x_i) \end{aligned}$$
dove p è la funzione di probabilità di X e $x_1, \ldots, x_n$ sono le possibili osservazioni (realizzazioni) di X.

3. Regole di calcolo:

$$\begin{aligned}\mu_{X+bY} &= \mu_X + b\mu_Y \\ \sigma^2_{a+bX} &= b^2\sigma^2_X \\ \sigma^2_X &= \mu_{X^2} + (\mu_X)^2\end{aligned}$$

4. Distribuzione di Bernoulli e distribuzione uniforme.

Capitolo 7: alcune variabili aleatorie discrete

1. Tecniche combinatorie: fattoriali, permutazioni e combinazioni.
2. Esperimenti binomiali e distribuzioni binomiali. Valore atteso e varianza di distribuzioni binomiali.
3. Esperimenti e distribuzioni di Poisson. Valore atteso e varianza di distribuzioni di Poisson.

Capitolo 8: variabili aleatorie continue I

1. Variabili aleatorie continue e funzione di densità.
2. Distribuzione uniforme e triangolare.

Capitolo 9: variabili aleatorie continue II

1. Densità di una distribuzione normale.
2. Valore atteso e varianza di distribuzioni normali.
3. Aree e probabilità di densità normali.

Capitolo 10: variabili aleatorie continue III

1. Unità standard e distribuzione normale standard.
2. Trasformazioni di variabili distribuite in modo normale.
3. Aree sotto la densità normale standard.
4. Aree sotto qualsiasi densità normale.

Capitoli 11 e 12: tecnica di calcolo integrale
Capitolo 13: variabili aleatorie continue (continuazione)

1. Funzione di ripartizione di una *V.A.*

2. Funzione di densità di una $V.A.$ continua.

3. Proprietà della funzione di ripartizione.

4. Valore atteso e varianza.

5. Densità di funzioni biiettive e derivabili di $V.A.$ continue.

Capitolo 14: Teorema del Limite Centrale
Capitolo 15: Distribuzioni Multivariate Discrete

1. Variabili aleatorie multivariate discrete.

2. Funzioni di probabilità congiunte, marginali, condizionate, indipendenza stocastica.

3. Valore atteso e varianza condizionata.

4. Valore atteso di funzioni di variabili aleatorie multivariate, covarianza e correlazione.

$$Cov(X, Y) = E\left[\left(X - E(X)\right)\left(Y - E(Y)\right)\right]$$

5. Covarianze e indipendenza, varianze di somme di $V.A.$.

Seguono infine i Capitoli 16, 17 e 18: distribuzioni multivariate continue I e II, la distribuzione normale bivariata.

1.2 Fondamenti di probabilità

Nella prima parte del corso (Statistica I) sono stati trattati quegli elementi di statistica descrittiva che hanno in seguito permesso di affrontare i temi propri della teoria della probabilità quali ad esempio

- Lo spazio campionario
- La funzione di probabilità
- Le variabili aleatorie di cui avete visto
  - la funzione di ripartizione
  - la funzione di densità
  - alcuni momenti teorici (valore atteso, varianza)

In questa sezione desideriamo rivedere alcuni concetti e definizioni ritenuti fondamentali per la comprensione dei prossimi capitoli. Il primo tema che vogliamo affrontare riguarda il concetto di *esperimento aleatorio*[1].

1.2.1 Esperimento Aleatorio

Definizione 1. Un esperimento aleatorio consiste nell'osservazione di un processo o procedimento aleatorio i cui esiti sono definiti ma non prevedibili.

Definizione 2. L'insieme degli esiti di un esperimento aleatorio è chiamato *spazio campionario* ed è notato Ω.

Definizione 3. Un qualsiasi sottoinsieme E di Ω è chiamato evento.

Esempio 1. Lancio di un dado. $\Omega = \{1, 2, 3, 4, 5, 6\}$. $E = \{\text{osservo un numero pari}\} = \{2, 4, 6\}$.

Esempio 2. Lancio indipendente di due dadi. In questo caso avremo

$$\Omega = \{(1, 1), (1, 2), ..., (1, 6), ..., (6, 1), ..., (6, 6)\}.$$

Un possibile evento: $E = \{\text{Osservo una loro somma superiore a } 9\} = \{(5, 5), (5, 6), (6, 5), (6, 6)\}$.

Tipicamente siamo interessati a calcolare la probabilità di eventi tramite una legge o distribuzione o misura[2] di probabilità P. Ad esempio, quando lo spazio campionario è discreto e finito (cioè Ω contiene un numero finito di esiti) e gli esiti sono equiprobabili avete imparato a calcolare la probabilità di un evento E tramite la formula

$$P(E) = \frac{\#\text{ esiti in } E}{\#\text{ esiti in } \Omega} = \frac{\#\text{ esiti favorevoli}}{\#\text{ esiti totali}}. \tag{1.1}$$

Osservazioni:

[1] Rinominiamo il concetto di *esperimento statistico* in *esperimento aleatorio*. Come avremo modo di discutere più avanti, all'espressione "esperimento statistico" verrà assegnato un significato diverso.

[2] Utilizzeremo questi tre termini quali sinonimi.

1. La formula (1.1) consente di assegnare una probabilità ad uno qualsiasi dei $2^6 = 64$ diversi eventi[3] di Ω dell'Esempio 1.

2. La formula (1.1) ha validità limitata: essa è applicabile solo al caso di uno spazio campionario finito i cui esiti sono equiprobabili.

3. Quando lo spazio campionario Ω ha un numero infinito di esiti le cose si complicano: in generale non è più possibile calcolare la probabilità di un qualsiasi sottoinsieme di Ω.

4. Quando gli esiti non sono equiprobabili la formula (1.1) è inutilizzabile. Ad esempio, nell'esperimento aleatorio definito dal lancio di un dado la probabilità dell'evento $E = \{\text{l'esito è un numero pari}\}$ non è calcolabile tramite la formula 1.1 quando il dado è truccato!

La definizione formale dell'insieme degli eventi $\mathcal{E}$ è la seguente:

Definizione 4. L'insieme degli eventi $\mathcal{E}$ è un insieme i cui elementi sono sottoinsiemi di Ω. Gli elementi $\mathcal{E}$ sono chiamati *eventi*. Devono valere le seguenti proprietà:

1. Lo spazio campionario è elemento di $\mathcal{E}$, cioè $\Omega \in \mathcal{E}$.

2. Dato un qualsiasi elemento E di $\mathcal{E}$ allora anche il suo complemento deve appartenere a $\mathcal{E}$.

3. Se $E_1, E_2, \dots$ è una successione di elementi qualsiasi di $\mathcal{E}$, allora anche la loro unione deve appartenere ad $\mathcal{E}$:

$$E_i \in \mathcal{E}\ \forall i \in \{1, 2, \dots\} \Rightarrow \bigcup_{i=1}^{\infty} E_i \in \mathcal{E}.$$

Esempio 3. Lancio un dado. Lo spazio campionario è $\Omega = \{1, 2, 3, 4, 5, 6\}$. I seguenti insiemi sono tutti esempi validi di insieme degli eventi

1. $\mathcal{E} = \{\Omega, \emptyset\}$.

2. $\mathcal{E} = \{\Omega, \emptyset, \{1\}, \{2, 3, 4, 5, 6\}\}$.

[3] Il numero di eventi diversi fra loro di uno spazio campionario finito Ω è dato dalla formula

$$2^{\#\ \text{esiti in}\ \Omega}.$$

3. $\mathcal{E} = \{\Omega, \emptyset, \{1,3,5\}, \{2,4,6\}\}$.
4. $\mathcal{E} = \{\Omega, \emptyset, \{1,2,3\}, \{4,5\}, \{6\}, \{4,5,6\}, \{1,2,3,6\}, \{1,2,3,4,5\}\}$.

I seguenti insiemi non sono esempi validi di insiemi degli eventi (perché?):

1. $\mathcal{E} = \{\Omega, \emptyset, \{1\}, \{2,3\}, \{1,2,3\}, \{2,3,4,5,6\}\}$.
2. $\mathcal{E} = \{\Omega, \emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\}$.
3. $\mathcal{E} = \{\Omega, \emptyset, \{1\}, \{1,2\}, \{1,2,3\}, \{1,2,3,4\}, \{1,2,3,4,5\}\}$.

Lasciando i dettagli tecnici ai matematici indicheremo semplicemente con $\mathcal{E}$ l'insieme degli eventi e senza preoccuparci oltre misura della sua struttura daremo per scontato che $\mathcal{E}$ soddisfa le condizioni enunciate nella Definizione 4. Se tale condizioni sono verificate diremo che l'insieme degli eventi $\mathcal{E}$ è misurabile, nel senso che è possibile assegnare una probabilità a ciascun evento in maniera "consistente". La proprietà importante da ricordare è che l'insieme degli eventi $\mathcal{E}$ contiene dei *sottoinsiemi* di Ω: ogni elemento E di $\mathcal{E}$ è un sottoinsieme di Ω, vedi Figura 1.1.

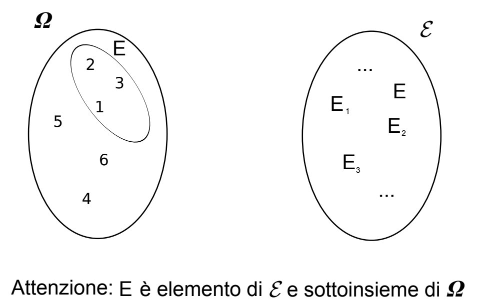


Figura 1.1: Lancio di un dado. Spazio campionario e insieme degli eventi.

Il terzo ed ultimo oggetto fondamentale che vogliamo discutere è la legge di probabilità P per la quale utilizziamo la seguente definizione.

Definizione 5. La legge di probabilità P è una funzione il cui dominio è $\mathcal{E}$, l'insieme degli eventi, con le seguenti proprietà:

1. $P(E) \geq 0$ per ogni evento $E \in \mathcal{E}$.

2. $P(\Omega) = 1$.

3. Se $E_1, E_2, \ldots$ è una successione di eventi di $\mathcal{E}$ fra loro incompatibili, cioè $E_i \cap E_j = \varnothing$ per $i \neq j$, allora

$$P(E_1 \cup E_2 \cup \ldots) = \sum_{i=1}^{\infty} P(E_i).$$

Della Definizione 5 è importante notare una cosa: gli argomenti da inserire in P non sono esiti (elementi di Ω) ma eventi (elementi di $\mathcal{E}$)!

Per concludere: un esperimento aleatorio è caratterizzato da tre oggetti: lo spazio campionario Ω, l'insieme degli eventi $\mathcal{E}$ e la legge di probabilità P. La tripla $(\Omega, \mathcal{E}, P)$ è chiamata *spazio di probabilità*.

Esempio 4. Lancio di un dado non truccato. Abbiamo visto in precedenza che

- $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- $\mathcal{E}$ è l'insieme che contiene i $2^6 = 64$ possibili sottoinsiemi (eventi) di Ω. Il loro elenco è lasciato come esercizio.
- P è la misura o legge di probabilità definita su $\mathcal{E}$. Per qualsiasi elemento (evento) E di $\mathcal{E}$ avremo

$$P(E) = \frac{\#\text{ esiti in } E}{6}.$$

La tripla $(\Omega, \mathcal{E}, P)$ è lo spazio di probabilità definito dall'esperimento aleatorio "lancio di un dato non truccato".

Esempio 5. Lancio di un dado truccato. Gli esiti dell'esperimento non cambiano rispetto al lancio di un dado non truccato. Avremo quindi

- $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Anche gli eventi rimangono immutati. Come in precedenza varrà che

- $\mathcal{E}$ è l'insieme che contiene i $2^6 = 64$ possibili sottoinsiemi (eventi) di Ω.

La legge di probabilità di questo esperimento aleatorio, notata $\tilde{P}$, è diversa dalla precedente. Infatti, mentre in precedenza un qualsiasi evento elementare quale ad esempio $E = \{1\}$ aveva probabilità uguale a $\frac{1}{6}$, ora a causa della "manomissione" del dado i sei eventi elementari $E_i = \{i\}$, $i = 1, \ldots, 6$ avranno probabilità $\tilde{p}_i$ delle quali almeno due saranno diverse da $\frac{1}{6}$. Dato un qualsiasi evento E di $\mathcal{E}$, la sua probabilità sarà calcolata come la somma delle probabilità dei singoli esiti in esso contenuti. Per capire ciò, prendiamo l'evento $E = \{\text{osservo un numero pari}\} = \{2, 4, 6\}$. Notiamo che $E = E_2 \cup E_4 \cup E_6$. Ma i tre eventi elementari sono disgiunti e quindi per la proprietà 3 di una legge di probabilità avremo che

$$\tilde{P}(E) = \tilde{P}(E_2) + \tilde{P}(E_4) + \tilde{P}(E_6) = \tilde{p}_2 + \tilde{p}_4 + \tilde{p}_6.$$

Per un qualsiasi evento E avremo quindi che

$$\tilde{P}(E) = \sum_{i \in E} \tilde{P}(\{i\}) = \sum_{i \in E} \tilde{p}_i.$$

La tripla $(\Omega, \mathcal{E}, \tilde{P})$ è lo spazio di probabilità che corrisponde al "lancio di un dato truccato". Rispetto all'esempio precedente i primi due termini della tripla sono identici, è cambiata unicamente la legge di probabilità.

1.2.2 Funzione di ripartizione e di densità

Abbiamo visto che per ottenere un esperimento aleatorio è necessario definire lo spazio campionario Ω nonché l'insieme degli eventi $\mathcal{E}$ sul quale sarà definta la legge di probabilità P. Consideriamo ora un esperimento aleatorio $(\Omega, \mathcal{E}, P)$ in cui lo spazio campionario Ω è sottoinsieme di $\mathbb{R}$. Questo significa che gli esiti dell'esperimento saranno dei numeri[4]. In tal caso è possibile definire la funzione di ripartizione della legge di probabilità P.

Definizione 6. (Funzione di ripartizione di un esperimento aleatorio). La funzione di ripartizione $F : \mathbb{R} \to [0, 1]$ della legge di probabilità P è definita come

$$F(x) := P((-\infty, x]) \text{ con } x \in \mathbb{R}. \tag{1.2}$$

La funzione di ripartizione $F(x)$ è dunque la probabilità dell'evento $E = (-\infty, x]$ di osservare un esito inferiore o al massimo uguale al valore x.

[4] Non tutti gli esperimenti aleatori soddisfano questa condizione. Pensate all'esperimento aleatorio di osservare il colore della prima autovettura che circola su una determinata tratta stradale a partire da un certo orario. L'esito non è numerico ma è un colore. Per contro, l'esempio del lancio di un dado soddisfa questa condizione in quanto lo spazio campionario $\Omega = \{1, 2, 3, 4, 5, 6\} \subset \mathbb{R}$.

Osservazione 1. Facciamo notare che P ed F sono funzioni diverse. F è una funzione definita su $\mathbb{R}$ mentre P è una funzione definita su $\mathcal{E}$, l'insieme degli eventi. È possibile dimostrare come P ed F siano "le due facce della stessa moneta" nel senso che da P segue univocamente F e viceversa.

Esempio 6. Distribuzione uniforme discreta con $\Omega = \{0.2, 0.4, 0.6, 0.8, 1\}$.

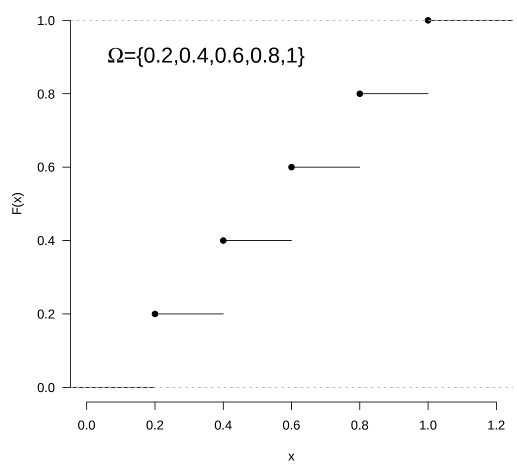


Figura 1.2: Distribuzione uniforme discreta

Esempio 7. Distribuzione uniforme continua $U[0, 1]$.

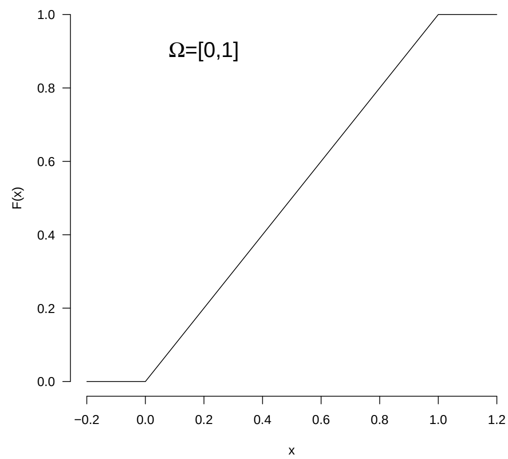


Figura 1.3: Distribuzione uniforme continua

Definizione 7. Diciamo che F è assolutamente continua se esiste una funzione reale f non negativa tale per cui

$$F(x) = \int_{-\infty}^{x} f(u)du \text{ per ogni } x \in \mathbb{R}.$$

La funzione f è chiamata funzione di densità di F. Vale inoltre la relazione

$$F'(x) = f(x).$$

1.3 Statistica descrittiva e teoria della probabilità

Nel corso di Statistica I avete studiato che il punto di partenza di qualunque analisi statistica è la definizione di una popolazione obiettivo (o popolazione di riferimento) nonché della caratteristica della popolazione a cui si è interessati. A titolo di esempio prendiamo quale popolazione di riferimento l'insieme di studenti che lo scorso anno ha seguito il corso di Statistica II. Quali caratteristiche sotto esame scegliamo il sesso e il numero di scarpe di ciascun studente. La Figura 1.4 riporta il grafico delle frequenze relative per le due caratteristiche.

Notiamo che la seconda caratteristica è numerica. In questo semplice esempio non abbiamo definito delle classi di valore: tutti i numeri di scarpe osservati

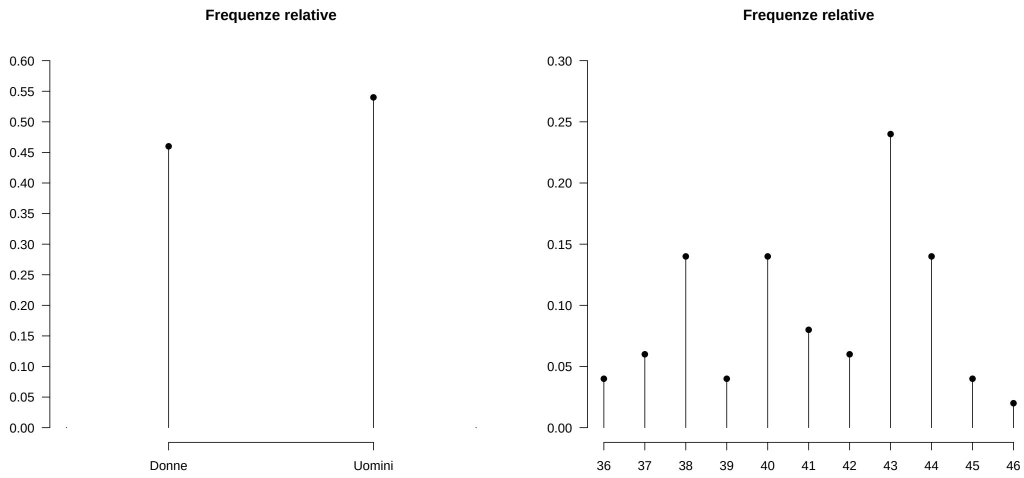


Figura 1.4: Frequenze relative popolazione studenti Statistica II - 2009

figurano nell'istogramma. Tuttavia se avessimo studiato un'altra caratteristica quale ad esempio il peso o l'altezza ecco che molto probabilmente avremmo costruito delle classi in cui inserire ciascun individuo della popolazione. La costruzione di classi può però costituire un problema. Innanzi tutto è necessario trovare una regola che a partire dalle singole osservazioni indichi quante e quali classi costruire. Secondariamente il passaggio dall'insieme della popolazione al suo istogramma di frequenze relative o assolute genera una perdita di informazione. Ad esempio, anche se sapessimo che 23 individui hanno un peso compreso tra 70 e 75 chilogrammi non saremmo in grado di stabilire quanti di essi hanno un peso superiore a 72 chilogrammi. Per tale motivo quando abbiamo a che fare con caratteristiche numeriche preferiamo studiare la funzione di ripartizione della popolazione obiettivo.

Definizione 8. (Funzione di ripartizione della popolazione obiettivo). La funzione di ripartizione, notata F, di una caratteristica x della popolazione obiettivo è una funzione di variabile reale definita su tutto $\mathbb{R}$ e tale che per ogni $x \in \mathbb{R}$

$$F(x) = \frac{\#\text{ unità con caratteristica } \leq x}{\#\text{ unità della popolazione}}. \tag{1.3}$$

L'interpretazione è semplice ed è molto simile all'interpretazione della funzione di ripartizione di un esperimento aleatorio. Infatti, tornando al nostro esempio iniziale della popolazione di studenti che hanno seguito il corso di Statistica II, $F(38) = 0.24$ significa che il 24% della popolazione possiede

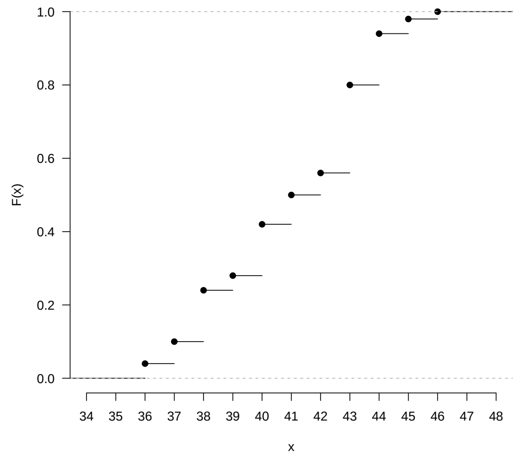


Figura 1.5: Funzione di ripartizione del numero di scarpe della popolazione studenti Statistica II - 2009

un numero di scarpa inferiore o uguale a 38. Come per un esperimento aleatorio, la funzione di ripartizione racchiude tutta l'informazione disponibile sulla popolazione (confronta la Definizione (1.2)). Nel caso della funzione di ripartizione di una popolazione obiettivo non si parla di probabilità ma di frequenze relative in quanto non c'è nulla di aleatorio nella popolazione. La popolazione esiste e l'attributo è osservabile. F descrive come l'attributo in esame è distribuito all'interno della popolazione. La conoscenza di F è equivalente all'osservazione dell'attributo su tutta la popolazione obiettivo.

Lo scopo di molti studi empirici in economia (micro- e macroeconomia, marketing ma anche in altre scienze sociali e non) è quello di determinare la funzione di ripartizione di un determinato attributo o caratteristica di una popolazione obiettivo[5] ed in seguito utilizzare questa informazione per giungere a delle conclusioni di carattere generale. Per tale motivo uno dei temi dei prossimi capitoli sarà proprio quello di studiare come ottenere un'approssimazione della distribuzione dell'intera popolazione partendo da un sottoinsieme di osservazioni della stessa (campione).

Le definizioni (1.2) e (1.3) di funzione di ripartizione caratterizzano rispettivamente l'aspetto probabilistico di un esperimento aleatorio e la struttura

[5]Parleremo semplicemente di *distribuzione della popolazione* quando è chiaro quali siano l'attributo e la popolazione obiettivo in esame.

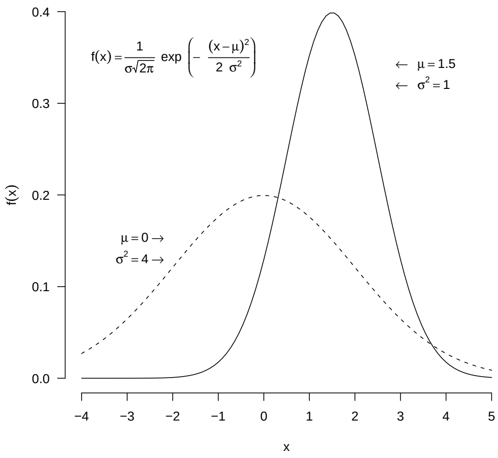


Figura 1.6: Due funzioni di densità della famiglia parametrica Normale

della popolazione obiettivo. Le due definizioni, anche se concettualmente molto simili, non devono essere confuse.

1.4 Famiglie parametriche di distribuzioni

Nel corso di Statistica I sono state presentate alcune particolari distribuzioni di probabilità con le relative funzioni di ripartizione, di densità o di probabilità. La distribuzione Normale (caso continuo) e la distribuzione di Poisson (caso discreto) sono due esempi a voi noti. Per quanto riguarda la distribuzione Normale, essa è caratterizzata da due *parametri* che come già sapete corrispondono al valore atteso ed alla varianza della distribuzione.

Osservazione 2. Attenzione a non generalizzare questa caratteristica della distribuzione Normale. Durante questo corso incontreremo diverse nuove distribuzioni di probabilità. Come nel caso della distribuzione Normale o di Poisson la loro "forma" dipenderà da uno o più parametri la cui interpretazione varierà da distribuzione a distribuzione.

La Figura 1.6 esplicita quanto affermato mostrando la funzione di densità della distribuzione Normale per due diversi valori di (μ, σ^2). Nella Figura 1.6 la formula della funzione di densità è invariata. Cambiano i valori dei due parametri μ e σ^2. Per ogni possibile valore dei parametri si ottiene una diversa distribuzione della famiglia Normale. Si è soliti indicare col termine famiglia parametrica Normale l'insieme di tutte le distribuzioni che si possono ottenere facendo variare i due parametri μ e σ^2. In generale si parlerà di famiglia parametrica di distribuzioni per indicare un preciso insieme di distribuzioni la cui funzione di probabilità, funzione di ripartizione o funzione di densità è identica a meno di un numero finito di parametri. I parametri assumono valori in $\mathbb{R}$ o in sottoinsiemi di $\mathbb{R}$. Nel caso della famiglia parametrica Normale $\mu \in \mathbb{R}$ mentre $\sigma^2 > 0$. L'insieme dei valori che i parametri di una famiglia parametrica di distribuzioni possono assumere è chiamato spazio parametrico ed è indicato con il simbolo Θ. Per la famiglia parametrica Normale $\Theta = \mathbb{R} \times \mathbb{R}^+$. Riportiamo alcuni esempi di famiglie parametriche di distribuzioni discrete e continue. Nella loro definizione appare la *funzione indicatrice* $I_A(x)$, dove A rappresenta un sottoinsieme di $\mathbb{R}$.

Ricordiamo che per qualsiasi sottoinsieme A di $\mathbb{R}$ la funzione $I_A : \mathbb{R} \rightarrow \{0, 1\}$ è definita nel seguente modo

$$I_A(x) = \begin{cases} 1 & \text{se } x \in A \\ 0 & \text{altrimenti} \end{cases} .$$

Prendendo quale esempio la distribuzione *discreta* uniforme, l'insieme A corrisponde all'insieme dei numeri interi compresi da 1 a n, ovvero $\{1, ..., n\}$. Se poniamo $n = 5$ avremo che $I_{\{1,...,5\}}(20) = 0$, $I_{\{1,...,5\}}(2.6) = 0$, $I_{\{1,...,5\}}(4) = 1$.

	Uniforme	**Bernoulli**
$f(x) =$	$\frac{1}{n} I_{\{1,...,n\}}(x)$	$p^x(1-p)^{1-x} I_{\{0,1\}}(x)$
$F(x) =$	$\min(\frac{\lfloor x \rfloor}{n}, 1) I_{[1,\infty)}(x)$	$(1-p) I_{[0,\infty)}(x) + p I_{[1,\infty)}(x)$
Spazio parametrico Θ	$n = 1, 2, \ldots$	$0 \leq p \leq 1$
Valore atteso	$\frac{n+1}{2}$	p
Varianza	$\frac{n^2-1}{12}$	$p(1-p)$

	Binomiale	**Binomiale negativa**
$f(x) =$	$\binom{n}{x} p^x (1-p)^{n-x} I_{\{0,...,n\}}(x)$	$\binom{r+x-1}{x} p^r (1-p)^x I_{\{0,1,...\}}(x)$
$F(x) =$	$\sum_{i=0}^{\lfloor x \rfloor} f(i)$	$\sum_{i=0}^{\lfloor x \rfloor} f(i)$
Spazio parametrico Θ	$0 \leq p \leq 1$; $n = 1, 2, \ldots$	$0 < p \leq 1$; $r = 1, 2, \ldots$
Valore atteso	np	$\frac{r(1-p)}{p}$
Varianza	$np(1-p)$	$\frac{r(1-p)}{p^2}$

	Geometrica	**Ipergeometrica**
$f(x) =$	$p(1-p)^x I_{\{0,\ldots,n\}}(x)$	$\dfrac{\binom{K}{x}\binom{M-K}{n-x}}{\binom{M}{n}} I_{\{0,1,\ldots,n\}}(x)$
$F(x) =$	$\left(1-(1-p)^{\lfloor x+1 \rfloor}\right) I_{[0,\infty)}(x)$	$\sum_{i=0}^{\lfloor x \rfloor} f(i)$
Spazio parametrico Θ	$0 < p \leq 1$	$M = 1, 2, \ldots\ ;\ K = 0, \ldots, M\ ;\ n = 1, 2, \ldots$
Valore atteso	$\frac{1-p}{p}$	$n\frac{K}{M}$
Varianza	$\frac{1-p}{p^2}$	$n\frac{K}{M}\frac{M-K}{M}\frac{M-n}{M-1}$

	Uniforme	**Normale**
$f(x) =$	$\frac{1}{b-a} I_{[a,b]}(x)$	$\frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2)$
$F(x) =$	$\min(\frac{x-a}{b-a}, 1) I_{[a,\infty)}(x)$	$\int_{-\infty}^{x} f(u)du$
Spazio parametrico Θ	$-\infty < a < b < \infty$	$-\infty < \mu < \infty\ ;\ \sigma > 0$
Valore atteso	$(a+b)/2$	μ
Varianza	$(b-a)^2/12$	σ^2

	Logistica	**Pareto**
$f(x) =$		$f(x) = \frac{\theta k^\theta}{x^{\theta+1}} I_{(k,\infty)}(x)$
$F(x) =$	$[1 + \exp(-(x-\alpha)/\beta]^{-1}$	
Spazio parametrico Θ	$\beta > 0\ ;\ -\infty < \alpha < \infty$	$k > 0\ ;\ \theta > 0$
Valore atteso	$\alpha + \beta\gamma$ con $\gamma \approx 0.577216$	$\frac{\theta k}{\theta - 1}$ quando $\theta > 1$
Varianza	$\frac{\pi^2 \beta^2}{6}$	$\frac{\theta k^2}{(\theta-1)^2(\theta-2)}$ quando $\theta > 2$

Le famiglie appena presentate sono solo alcune delle numerose famiglie parametriche di distribuzioni. Come potete verificare voi stessi il numero di parametri e la loro interpretazione varia da famiglia a famiglia.

1.5 Contenuto del corso

Nella seconda parte del corso (Statistica II) verranno affrontati i temi propri dell'induzione statistica, ovvero

- La teoria del campionamento

  La teoria del campionamento trae la sua ragion d'essere dalla necessità di raccogliere in maniera appropriata le informazioni (dati) necessari allo studio del fenomeno a cui siamo interessati.

- La teoria della stima

  La teoria della stima utilizza le informazioni così raccolte (più eventuali altre informazioni) per trarre delle conclusioni o dare delle risposte alle domande relative al fonomeno sotto esame.

- La verifica d'ipotesi

  A causa dell'incompleta informazione e alla natura aleatoria di molti fenomeni studiati in economia e nelle scienze sociali tali conclusioni non saranno vere in assoluto. Tuttavia, se lo studio è stato condotto seguendo certi principi, il grado di incertezza potrà essere quantificato. La verifica d'ipotesi tramite dei test statistici e la construzione di intervalli di confidenza relativi ai parametri stimati permettono di misurare l'attendibilità dei risultati.

Esempio 8. Pino, il responsabile del settore marketing di una grossa azienda, desidera conoscere il grado di visibilità sul mercato di un proprio prodotto. Il motivo di questo interesse consiste nel fatto che si vuole capire se lo scarso successo commerciale sia dovuto alla scarsa qualità del prodotto o al fatto che il prodotto non sia sufficientemente conosciuto. La dirigenza dell'azienda decide di quantificare il grado di visibilità tramite la percentuale p di consumatori che conoscono (ma non necessariamente acquistano o hanno acquistato in passato) il prodotto in questione. Essa valuta il livello attuale di p secondo la seguente tabella

$p < 30\%$	$30\% \leq p < 60\%$	$60\% \leq p$
insufficiente	discreto	buono

La quantità in esame è dunque p. Poiché è praticamente impossibile (e troppo oneroso) intervistare tutta la popolazione dei consumatori sarà necessario selezionare ed intervistare un sottoinsieme di tale popolazione che chiameremo campione.

- Come si dovrà costruire il campione da intervistare?
- Sulla base di quali criteri dovranno essere selezionate le persone (o unità) da intervistare: l'età, il reddito, la provenienza geografica, il sesso?
- Quante persone dovranno essere intervistate?

La teoria del campionamento si occupa di trovare una risposta a queste domande.

Una volta effettuato il campionamento si potrà stimare

$$0 \leq p \leq 1.$$

La teoria della stima consente di effettuare due tipi di stime: una stima puntuale di p ed una stima per intervallo. La stima puntuale di p, notata semplicemente $\widehat{p}$, fornisce quella che noi riteniamo essere la migliore approssimazione o alternativa al valore sconosciuto p. Poiché il calcolo è effettuato sulla base di un campione di consumatori e non sull'intera popolazione (l'informazione è incompleta), il valore di $\widehat{p}$ sarà molto probabilmente diverso da p. Si è dunque confrontati con un errore di stima.

La stima per intervallo affronta il problema di stima in maniera diversa. Essa infatti fornisce un intervallo di valori nel quale noi confidiamo che p sia incluso. Tale intervallo è chiamato *intervallo di confidenza*. La forza o intensità con cui noi crediamo nella nostra affermazione è chiamata *livello di confidenza*. Il livello di confidenza corrisponde alla probabilità ex-ante che l'intervallo costruito sulla base delle osservazioni disponibili contenga p. Un esempio di una realizzazione di intervallo di confidenza potrebbe essere:

$$[0.2, 0.4].$$

La teoria della stima si occupa dunque di come calcolare $\widehat{p}$ o come costruire l'intervallo di confidenza utilizzando le informazioni disponibili (dati).

Esempio 9. Continuiamo l'Esempio 8. Supponiamo che, nel caso in cui p fosse insufficiente, Pino abbia intenzione di condurre una vasta campagna pubblicitaria e che il valore stimato di p sia $\widehat{p} = 29\%$. A questo punto dobbiamo chiederci se la campagna pubblicitaria (ed i relativi costi) sia veramente necessaria. Infatti, potrebbe essere che il vero valore di p sia superiore al 30% ma per semplice "sfortuna" nella scelta del campione la nostra stima risulti inferiore a questa soglia. Il grado di visibilità è realmente insufficiente? Ci troviamo qui confrontati col terzo tema dell'inferenza statistica: la verifica d'ipotesi. Valori di $\widehat{p}$ inferiori al 30% sono evidenza a favore dell'ipotesi di un p insufficiente e contro l'ipotesi di un p discreto o buono. Tenendo conto dell'aleatorietà nella stima di p e del relativo errore, quanto bassa deve essere la stima $\widehat{p}$ per decidere di effettuare la campagna pubblicitaria e quindi rifiutare l'ipotesi che p sia discreto o buono? Per dare una risposta a questa domanda si costruirà un test statistico. Sarà innanzi tutto necessario una formalizzazione matematica della nostra ipotesi. Per Pino, che non ha seguito nessun corso di statistica ed è uomo d'azione, $\widehat{p} = 29\%$ è un valore più che sufficiente per eseguire la campagna pubblicitaria e spendere un sacco di soldi.

Sappiamo che tanto più la campagna pubblicitaria sarà efficace, tanto più aumenterà il grado di visibilità. Il manager dell'azienda non è soddisfatto del modo come Pino ha condotto la campagna pubblicitaria e desidera verificarne immediatamente l'efficacia. Ordina quindi ad una società indipendente e specializzata in sondaggi di intervistare un nuovo campione da cui stimare il nuovo grado di visibilità del prodotto che indicheremo ora con

$$0 \leq p_{new} \leq 1.$$

p_{new} è il grado di visibilità *dopo* la campagna pubblicitaria. Il manager utilizzerà la seguente regola per decidere la sorte di Pino:

$$\begin{array}{cc} p_{new} = p & p_{new} \geq p \\ \text{licenzio Pino} & \text{non licenzio pino} \end{array}$$

La stima puntuale di p_{new}, notata $\widehat{p}_{new}$ risulta essere del 34%, un valore di poco superiore al precedente 29% calcolato prima della campagna pubblicitaria. Pino è felice. Il suo capo gli fa però notare che l'esiguo aumento del 5% rispetto alla precedente stima potrebbe essere semplicemente dovuto al caso. Pino vi chiede allora di quantificare sulla base dei dati a disposizione il livello di evidenza per cui l'ipotesi $p = p_{new}$ (licenzio Pino) possa essere rigettata. È compito della verifica d'ipotesi fornire una risposta a questa domanda.

Esempio 10. Un produttore di mele deve fissare anticipatamente la quantità di mele da consegnare al suo acquirente. Sappiamo che la quantità q prodotta dipende tra le altre cose da fattori climatici, difficilmente prevedibili ed assai variabili. Da ricerche effettuate si sa che la produzione di mele q è distribuita secondo la distribuzione Normale[6] che sapete essere univocamente identificata dai parametri μ e σ^2. Purtroppo i due parametri μ e σ^2 in questo caso sono sconosciuti e dovranno essere stimati sulla base dei valori della produzione osservati negli ultimi 20 anni. È questo un problema legato alla teoria della stima. Infine, potremmo chiederci se l'assunzione della distribuzione normale si addice alla variabile aleatoria q o se tale assunzione è fondamentalmente sbagliata. In questo frangente si tratta quindi di verificare l'ipotesi concernente la distribuzione di una variabile aleatoria per la quale si osservano un certo numero di realizzazioni.

[6]Per la precisione la distribuzione Normale assegna probabilità positive all'evento di produrre una quantità negativa di mele. Tuttavia tale probabilità è trascurabile per dati valori di sigma.

1.6 Domande di fine capitolo

Domanda 1. Cosa rappresenta il simbolo Ω? Esplicitate il suo contenuto facendo un nuovo esempio non visto in classe.

Domanda 2. Che relazione sussiste tra Ω, $\mathcal{E}$ e E?

Domanda 3. Quali sono i tre elementi fondamentali che costituiscono un esperimento aleatorio?

Domanda 4. Costruite un esperimento aleatorio per cui $\Omega \nsubseteq \mathbb{R}$ (Ω *non* è l'insieme dei numeri reali o un suo qualsiasi sottoinsieme).

Domanda 5. Cos'è la funzione di ripartizione? Dato un qualsiasi esperimento aleatorio $(\Omega, \mathcal{E}, P)$ è sempre possibile definire la funzione di ripartizione della legge di probabilità P? Se non fosse sempre possibile costruite un semplice esempio di esperimento aleatorio per cui non è possibile definire la funzione di ripartizione.

Domanda 6. Che differenza c'è tra P ed F?

Domanda 7. Che differenza c'è tra la funzione di probabilità e la funzione di densità?

Domanda 8. Cos'è una famiglia parametrica di distribuzioni? Cosa rappresenta Θ?

Domanda 9. Per due famiglie parametriche di distribuzioni discrete e due famiglie parametriche di distribuzioni continue a scelta indicate

1. il numero di parametri
2. la funzione di probabilità o di densità o di ripartizione
3. lo spazio parametrico

che le contraddistingue. Scegliete in seguito dei valori appropriati per i parametri ed eseguite il grafico della funzione di probabilità o di densità o di ripartizione a dipendenza dalla scelta effettuata.

Domanda 10. Ripetete lo stesso esercizio della domanda precedente con altre due famiglie parametriche di distribuzioni (una discreta e l'altra continua) che non siano già elencate in questo capitolo.

Capitolo 2

Induzione statistica

2.1 Introduzione

Il processo di apprendimento e di creazione di nuova conoscenza è possibile grazie a procedimenti di carattere deduttivo e induttivo. L'acquisizione per via deduttiva permette, partendo da delle premesse generali la cui validità è garantita, di giungere a delle conclusioni particolari. Un esempio di ragionamento deduttivo è il seguente:

- Premessa maggiore: Tutti gli uomini sono mortali
- Premessa minore: Socrate è un uomo
- Conclusione: Socrate è mortale

Per contro, l'apprendimento per via induttiva consiste nel trovare delle leggi universalmente valide partendo da osservazioni particolari. Il processo di estensione dal particolare al generale è chiamato inferenza induttiva. Questo tipo di estensione della conoscenza è tuttavia soggetto ad incertezza: l'induzione non dimostra niente, essa è corretta solo nella totalità dei casi in cui è confermata la sua validità.
In statistica ci troviamo tipicamente nella situazione di dover inferire qualcosa circa la

1. *distribuzione* di una o più caratteristiche di una certa popolazione obiettivo, dove per *popolazione obiettivo* si intende la totalità degli elementi (individui, aziende, famiglie, cantoni, regioni, ...) per i quali desideriamo ottenere tali informazioni. In questo caso la popolazione obiettivo è finita ed indicheremo con la lettera N il numero dei suoi elementi.

2. legge o distribuzione di probabilità di un esperimento aleatorio.

Per quanto riguarda il caso di una popolazione finita, avete visto nel corso di statistica I come è possibile ottenere la distribuzione di un determinato carattere della popolazione in termini assoluti o relativi (percentuali) a partire dall'osservazione di ogni singolo elemento o unità della popolazione stessa. La conoscenza della distribuzione statistica è fondamentale in quanto essa descrive in maniera univoca e completa la caratteristica sotto esame. Purtroppo però solo in rare occasioni l'informazione disponibile è rappresentata dalla distribuzione statistica di *tutta* la popolazione obiettivo. Anzi è interessante osservare come ci siano situazioni dove la numerosità stessa della popolazione obiettivo non è conosciuta. Si pensi ad esempio ad un'analisi sul numero di sigarette che un fumatore consuma giornalmente. Quanto vale N, il numero di fumatori? Non solo la distribuzione del numero giornaliero di sigarette consumate da un fumatore è sconosciuta ma anche il numero stesso di fumatori che in questo esempio costituiscono la popolazione obiettivo.

Nel caso di un esperimento aleatorio formalizzato tramite uno spazio di probabilità $(\Omega, \mathcal{E}, P)$ l'interesse è posto sulla legge o distribuzione di probabilità P. Essa racchiude l'informazione disponibile sull'esperimento e le sue caratteristiche aleatorie. Vista la natura casuale dell'esperimento il risultato o esito $\omega \in \Omega$ dell'esperimento non è conosciuto a priori: quello che possiamo fare è assegnare delle probabilità agli eventi $E \in \mathcal{E}$. La conoscenza della legge di probabilità P permette di quantificare l'incertezza sul risultato dell'esperimento.

Per studiare la legge di probabilità di un dato esperimento aleatorio è necessario ripetere l'esperimento ed osservarne un certo numero di risultati . La teoria della stima si occupa di come utilizzare l'informazione disponibile per risalire alla legge di probabilità P o equivalentemente[1] alla funzione di ripartizione F. Se invece l'attenzione è rivolta allo studio di una certa caratteristica di una popolazione obiettivo, allora sarà necessario osservare un certo numero di individui/unità della popolazione ed in seguito stimare la distribuzione dell'intera popolazione obiettivo. È questo dunque il carattere induttivo della statistica inferenziale: da un numero *limitato* di osservazioni ottenere delle conclusioni di carattere *generale*.

2.2 Modelli parametrici

La stima di F risulta particolarmente difficile in quanto occorre stimare un'intera funzione: è necessario conoscere non solo il valore che F assume in un

[1] Nell'arco di questo corso tratteremo unicamente spazi di probabilità ed esperimenti aleatori per i quali l'esistenza di una funzione di ripartizione è assicurata.

determinato punto, ma in tutti i punti dello spazio campione! È questa un'impresa estremamente difficile. Sappiamo che F ha una caratteristica particolare: è una funzione monotona crescente. Tuttavia esite un'infinità di funzioni del genere. Per tale motivo lo statistico spesso restringe la sua ricerca ad un sottoinsieme ben preciso di funzioni di ripartizione. Ad esempio, si potrebbe pensare di limitarsi a quelle funzioni di ripartizione F che sono continue, o a quelle che posseggono una funzione di densità f (chiamate distribuzioni assolutamente continue, vedi Definizione 7)

$$F(x) = \int_{-\infty}^{x} f(t)dt \quad \text{con } f \text{ non negativa.} \tag{2.1}$$

Questo in parte può aiutare a semplificare la scelta (stima) di F ma non risolve tutti i problemi. La funzione di ripartizione di esperimento aleatorio il cui spazio campione è discreto, ad esempio, non è continua e tanto meno può essere rappresentata tramite un integrale del tipo (2.1). Inoltre, pur restringendo l'insieme dal quale stimare F, la scelta risulta ancora assai vasta. Una strada diversa e a volte complementare a quella appena esposta consiste nel limitare la ricerca di F ad un particolare insieme di distribuzioni parametriche. Come trattato nella Sezione 1.4, una famiglia parametrica di distribuzioni è un insieme di distribuzioni. Da ora in poi noteremo con la lettera $\mathcal{P}$ una qualsiasi famiglia parametrica di distribuzioni. I parametri sconosciuti associati a tale famiglia saranno indicati[2] con θ_1, θ_2 ... θ_K. Utilizzeremo semplicemente la notazione θ quando desideriamo indicare l'insieme dei K parametri posti in una lista ordinata che chiameremo semplicemente vettore $\theta = (\theta_1, \theta_2, ...\theta_K)$. La funzione di ripartizione verrà quindi indicata con $F(\cdot;\theta)$ e se esiste la funzione di densità (caso continuo) o di probabilità (caso discreto) scriveremo analogamente $f(\cdot;\theta)$. In generale si assumerà che a due diversi parametri θ e $\widetilde{\theta}$ corrispondano due diverse leggi di probabilità. Infine, l'insieme dei valori ammissibili per θ è indicato col simbolo Θ. Come osservato in precedenza, nel corso introduttivo di Statistica I avete incontrato diverse famiglie di distribuzioni parametriche sia discrete che continue (si vedano le dispense riguardanti alcune delle distribuzioni di probabilità parametriche più comuni). Utilizzeremo la seguente notazione per definire una famiglia parametrica di distribuzioni

$$\mathcal{P} = \{F(\cdot;\theta) : \ \theta \in \Theta\}$$

oppure, nel caso in cui la famiglia parametrica venga definita indirettamente tramite la funzione di probabilità (caso discreto) o la funzione di densità

[2]In generale e a dipendenza della distribuzione si utilizzano diversi altri simboli, quali ad esempio μ, σ o λ.

(caso continuo)

$$\mathcal{P} = \{f(\cdot;\theta) : \ \theta \in \Theta\} .$$

Esempio 11. La distribuzione continua uniforme è caratterizzata da due parametri che notiamo con a e b e che corrispondono rispettivamente all'estremo inferiore e superiore della funzione di densità. Formalmente questa famiglia può essere scritta come

$$\mathcal{P} = \left\{ f(x;\theta) = \frac{1}{b-a} I_{[a,b]}(x);\ \theta = (a,b) \in \mathbb{R}^2, \text{ con } a < b \right\} .$$

In questo esempio a qualsiasi coppia di numeri reali (a, b) tali per cui $a < b$ associamo una funzione di densità definita da

$$f(x;\theta) = \begin{cases} \frac{1}{b-a} & \text{se } x \in [a,b] \\ 0 & \text{altrimenti} \end{cases} . \tag{2.2}$$

Nel caso della distribuzione uniforme $A = [a, b]$. È facile dimostrare applicando la definizione di funzione indicatrice che la funzione di densità (2.2) è riscrivibile più semplicemente come

$$f(x;\theta) = \frac{1}{b-a} I_{[a,b]}(x) .$$

L'insieme dei possibili valori di θ, notato in precedenza con Θ, è dunque uguale al sottoinsieme dei punti del piano la cui prima coordinata è minore della seconda, formalmente

$$\Theta = \left\{ (a,b) \in \mathbb{R}^2 \mid a < b \right\} .$$

Esercizio 1. Consideriamo la distribuzione parametrica di Poisson. Rispondete alle seguenti domande.

1. La distribuzione di Poisson è una distribuzione discreta o continua?

2. Qual è o quali sono i parametri che la caratterizzano. Che valori possono assumere i/il parametro/i di una distribuzione di Poisson? Definite il contenuto di θ nonché l'insieme Θ. Definite $\mathcal{P}$.

3. Assegnate dei valori al/ai parametro/i. In Excel o Openoffice eseguite i grafici della funzione di densità e della funzione di ripartizione. Calcolate la probabilità degli eventi $E_1 = [0, 1]$ ed $E_2 = (1, 2)$.

4. Scegliete nuovi valori per il/i parametri e col medesimo programma confrontate i grafici delle due funzioni di densità e delle due funzioni di ripartizione. Sia S il sottoinsieme di $\mathbb{R}$ definito da quei punti per cui la funzione di densità è maggiore di 0, cioè

$$S = \{x \in \mathbb{R} \mid f(x) > 0\} .$$

   Rispetto al valore assegnato al punto precedente cambia l'insieme S con i nuovi valori? Siete in grado di generalizzare la vostra conclusione?

5. Cosa implica un cambiamento dei valori assegnati al/ai parametro/i rispetto alle probabilità degli eventi E_1 ed E_2 definiti in precedenza? Notate delle differenze? Spiegate?

Esercizio 2. Consideriamo la distribuzione parametrica di Pareto. Rispondete alle seguenti domande.

1. La distribuzione di Pareto è una distribuzione discreta o continua?

2. Qual è o quali sono i parametri che la caratterizzano. Che valori possono assumere i/il parametro/i di una distribuzione di Pareto? Definite il contenuto di θ nonché l'insieme Θ. Definite $\mathcal{P}$.

3. Assegnate dei valori al/ai parametro/i ed eseguite i grafici della funzione di densità e della funzione di ripartizione. Calcolate la probabilità degli eventi $E_1 = [0, 3]$ ed $E_2 = \{-1, 0, 2, 3\}$.

4. Trovate l'insieme S per i valori dei parametri da voi scelti.

5. È possibile trovare dei nuovi valori dei parametri in maniera tale da modificare l'insieme S? Trovate un esempio del genere e confrontate i grafici delle due funzioni di densità e delle due funzioni di ripartizione.

2.3 Esperimento statistico

È stato detto che la teoria della stima si occupa di giungere a delle conclusioni generali riguardo

- alla legge di probabilità di un esperimento aleatorio, oppure
- alla distribuzione di una variabile aleatoria, oppure
- alla distribuzione statistica di una popolazione obiettivo

partendo da un numero limitato di osservazioni di un certo esperimento aleatorio o di elementi/unità della popolazione. Da un punto di vista formale, ripetere n volte in maniera indipendente un esperimento aleatorio $(\Omega, \mathcal{E}, P)$ costituisce un nuovo tipo di esperimento a cui è dato il nome di *esperimento statistico.*

Definizione 9. (Esperimento statistico). Un esperimento statistico di ampiezza n è costituito da un numero n di ripetizioni indipendenti dello stesso esperimento aleatorio $(\Omega, \mathcal{E}, P)$.

Quali saranno i possibili esiti di un esperimento statistico? Sappiamo che ad ogni ripetizione dell'esperimento aleatorio lo spazio campione è sempre Ω. Potremmo pensare di raccogliere gli esiti delle n ripetizioni e creare una *lista ordinata* di valori. Il primo valore della lista corrisponde all'esito della prima esecuzione dell'esperimento aleatorio, il secondo valore corrisponde all'esito della seconda esecuzione e via dicendo. La nostra lista conterrà gli n esiti che indicheremo con $(x_1, x_2, ..., x_n)$. Attenzione a non confondere la lista così costruita con lo spazio campione Ω. Infatti, per ciascun $i = 1, ..., n$ l'i-esimo elemento x_i della lista è un elemento di Ω ma $(x_1, x_2, ..., x_n) \neq \Omega$. Lo spazio campione può contenere un numero infinito non numerabile di esiti. Pensate ad esempio all'esperimento aleatorio di registrare il tempo d'attesa ad uno sportello postale. In tal caso $\Omega = [0, \infty\}$ mentre $(x_1, x_2, ..., x_n)$ sarà sempre un insieme finito.

Esempio 12. L'esperimento statistico corrisponde a lanciare 2 volte un dado. L'esperimento aleatorio è il lancio di un dado. $\Omega = \{1, ..., 6\}$ mentre P è la distribuzione uniforme discreta (dado non truccato). Gli esiti dell'esperimento statistico possono essere rappresentati graficamente come segue:

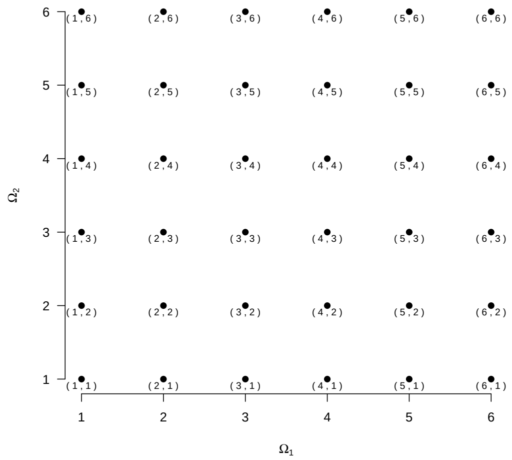


Figura 2.1: Esiti dell'esperimento statistico: lancio di due dadi.

Indicando con Ω_1 lo spazio campionario della prima esecuzione dell'esperimento aleatorio e con Ω_2 lo spazio campionario della seconda esecuzione, l'esperimento statistico avrà quale spazio campionario l'insieme delle coppie di esiti (x_1, x_2) con $x_1 \in \Omega_1$ e $x_2 \in \Omega_2$. Un esito di questo particolare esperimento statistico potrebbe essere $(3, 4)$. Poiché l'ordine conta, l'esito $(3, 4)$ è da considerarsi diverso dall'esito $(4, 3)$.

Definizione 10. Chiamiamo *spazio campionario prodotto* l'insieme di tutte le coppie (x_1, x_2) dove $x_1 \in \Omega_1$ e $x_2 \in \Omega_2$. Esso verrà notato con $\Omega \times \Omega$ o Ω^2. Nel caso di un numero n qualsiasi di ripetizioni scriveremo semplicemente Ω^n e un esito dell'esperimento statistico sarà rappresentato tramite la lista $(x_1, x_2, ..., x_n)$.

Avendo definito lo spazio campionario di un esperimento statistico dovremo ora definire gli eventi ad esso associati. In maniera del tutto identica alla definizione di evento di un esperimento aleatorio, l'evento di un esperimento statistico è un sottoinsieme di Ω^n. Indicheremo con $\bigotimes_{i=1}^{n} \mathcal{E}_i$ l'insieme degli eventi definiti su Ω^n.

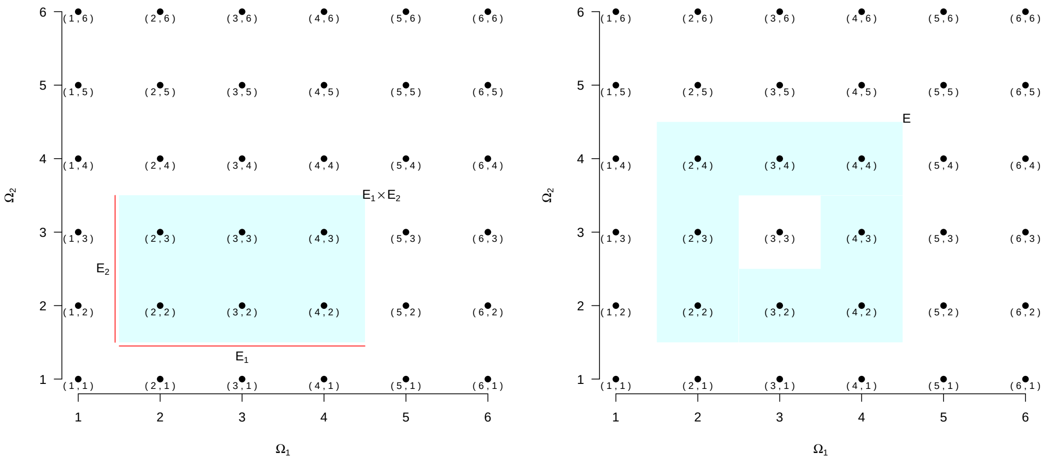


Figura 2.2: Esempi di eventi dello spazio campionario prodotto

Le due immagini della Figura 2.2 rappresentano graficamente due possibili eventi dell'esperimento statistico discusso nell'Esempio 12. Come potete notare nella prima immagine, l'evento definito dall'area colorata corrisponde al prodotto cartesiano $E_1 \times E_2$ dei due eventi $E_1 = \{2, 3, 4\}$ ed $E_2 = \{2, 3\}$ definiti rispettivamente sul primo e sul secondo lancio. Lo spazio $\bigotimes_{i=1}^{n} \mathcal{E}_i$ degli eventi definiti su Ω^n è molto vasto. La seconda immagine ci mostra infatti che esso comprende eventi che non sono semplicemente riconducibili ad un prodotto cartesiano $E_1 \times E_2$ di eventi rispettivamente di Ω_1 e Ω_2. Ancora una volta evitiamo di addentrarci in dettagli tecnici. Considereremo lo spazio degli eventi $\bigotimes_{i=1}^{n} \mathcal{E}_i$ come un "particolare" insieme i cui *elementi* sono *sottoinsiemi* di Ω^n.

Utilizzando la proprietà di indipendenza tra le varie ripetizioni dell'esperimento aleatorio varrà che la legge di probabilità di un esperimento statistico è semplicemente il prodotto delle singole leggi di probabilità P. Infatti, indicando con Q la legge di probabilità di un esperimento statistico (definita quindi su $\bigotimes_{i=1}^{n} \mathcal{E}_i$), avremo per qualsiasi evento

$$E_1 \times E_2 \times \ldots \times E_n$$

appartenente a $\overset{n}{\underset{i=1}{\otimes}}\mathcal{E}_i$ che la sua probabilità è uguale

$$Q(E_1 \times E_2 \times \ldots \times E_n) = P\left(E_1\right)\ P\left(E_2\right)\ \cdots\ P\left(E_n\right)\ .$$

In maniera del tutto simile ad un esperimento aleatorio, un esperimento statistico può quindi essere rappresentato dalla tripla

$$\left(\Omega^n, \overset{n}{\underset{i=1}{\otimes}}\mathcal{E}_i, Q = \overset{n}{\underset{i=1}{\otimes}}P_i\right). \tag{2.3}$$

Un esperimento statistico possiede la medesima struttura di un esperimento aleatorio e soddisfa quindi le caratteristiche di uno spazio di probabilità. Vediamo ora un esempio di esperimento statistico.

Esempio 13. Consideriamo l'esperimento aleatorio che riguarda la misurazione della durata di vita di una lampadina: $\Omega = \mathbb{R}$ mentre quale distribuzione parametrica per la legge di probabilità P utilizziamo la distribuzione esponenziale

$$\mathcal{P} = \left\{f(x;\lambda) = \lambda e^{-\lambda x} I_{(0,\infty)}(x)\ :\lambda > 0\right\}.$$

Definiamo ora l'esperimento statistico costituito dalla misurazione della durata di vita di $n = 3$ lampadine prodotte in maniera indipendente l'una dall'altra. Il nuovo esperimento avrà quale spazio campionario $\mathbb{R}^3$. Un evento di questo spazio potrebbe essere il seguente sottoinsieme di $\mathbb{R}^3$

$$E_1 \times E_2 \times E_3$$

dove

- $E_1 = \{$"La prima lampadina si spegne dopo 3 ore"$\} = [3, \infty)$,
- $E_2 = \{$"La seconda lampadina dura meno di 20 minuti"$\} = [0, \frac{1}{3})$,
- $E_3 = \{$"La terza lampadina prima o poi si spegne"$\} = [0, \infty)$.

Grazie all'ipotesi di indipendenza nell'esecuzione dei tre esperimenti la legge di probabilità su $\overset{3}{\underset{i=1}{\otimes}}\mathcal{E}_i$ non è altro che il prodotto di P sui rispettivi eventi

$$Q(E_1 \times E_2 \times E_3) = P(E_1)P(E_2)P(E_3).$$

Le singole probabilità possono essere calcolate individualmente tramite l'integrale della funzione di densità, ad esempio

$$P(E_1) = \int_3^\infty \lambda e^{-\lambda x} dx$$

Il modello parametrico della legge di probabilità Q sarà semplicemente (indicheremo con $f(x_1, x_2, x_3)$ la funzione di densità di Q)

$$\begin{aligned}\mathcal{P} &= \{f(x_1, x_2, x_3; \lambda) = f(x_1; \lambda)\ f(x_2; \lambda)\ f(x_3; \lambda) : \ \lambda > 0\} \\ &= \left\{f(x_1, x_2, x_3; \lambda) = \lambda^3 e^{-\lambda(x_1+x_2+x_3)} I_{(0,\infty)}(x_1) I_{(0,\infty)}(x_2) I_{(0,\infty)}(x_3) : \ \lambda > 0\right\}.\end{aligned}$$

Come potete constatare, la difficoltà è più di carattere formale (forma e notazione) che concettuale.

2.4 Variabili aleatorie

Talvolta succede che ciò a cui siamo veramente interessati non sia l'esito di un esperimento aleatorio ma una sua particolare funzione. Pensate ad esempio ad una compagnia aerea. Supponiamo che il ritardo di un volo di linea sia una variabile aleatoria. Alla compagnia interessa soprattutto conoscere i costi che un certo ritardo può generare piuttosto che l'ammontare in minuti del ritardo. I costi saranno quindi una certa funzione dei minuti di ritardo. Generalizzando al caso astratto, introduciamo la seguente definizione.

Definizione 11. Dato un esperimento aleatorio $(\Omega, \mathcal{E}, P)$ diremo che la funzione $g : \Omega \to \mathbb{R}$ è misurabile se, preso un qualsiasi evento $E' \subset \mathbb{R}$, l'insieme $g^{-1}(E')$ è un elemento di $\mathcal{E}$.

In altre parole, diremo che la funzione g è misurabile se la preimmagine di un evento di $\mathbb{R}$ è a sua volta un evento di Ω. Le figure 2.3 e 2.5 mostrano astrattamente la situazione[3].

Ricordiamo che per definizione

$$g^{-1}(E') = \{\omega \in \Omega \mid g(\omega) \in E'\}.$$

[3]Per convenienza assumeremo che i valori che la funzione g assume siano numeri reali. Tuttavia ci preme sottolineare che la definizione generale di variabile aleatoria non impone questa restrizione sul codominio di g. Un esempio in cui il codominio di g non è numerico è descritto nella Sezione 3.1.

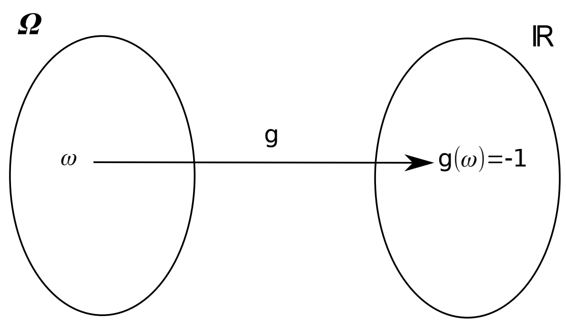


Figura 2.3: Variabile aleatoria - diagramma di Venn

La preimmagine di E' rispetto alla funzione g è un sottoinsieme di Ω e contiene tutti gli esiti la cui immagine attraverso g appartiene all'evento E' di $\mathbb{R}$. La nostra definizione di evento (confronta la Definizione 3) considera evento un qualsiasi sottoinsieme di Ω. Ne segue che per noi qualsiasi funzione da Ω verso $\mathbb{R}$ è misurabile[4].

Esempio 14. Il numero di ore che un traghetto impiega per andare da Genova ad Olbia è una variabile aleatoria in quanto il tempo di percorrenza è influenzato da diversi fattori in parte non prevedibili quali ad esempio le condizioni del mare, l'intensità e direzione del vento, il numero di passeggeri e così via. Indicando con x il tempo impiegato dal traghetto per la traversata, la funzione di costo è data da

$$g(x) = \begin{cases} 10 + 10x, & \text{se } 0 < x \leq 9 \\ 100 + 20(x-9), & \text{se } 9 < x \leq 10 \\ 120 + 40(x-10), & \text{se } x > 10 \end{cases} .$$

Con quale probabilità si osserveranno dei costi compresi nell'intervallo $E' = [110 - 130]$? Per rispondere a questa domanda occorre calcolare la preimmagine dell'evento E' utilizzando la funzione dei costi $g(x)$: $E = g^{-1}(E')$. È possibile aiutarsi col grafico della funzione, vedi Figura 2.4.

[4]Come accennato precedentemente le definizioni di evento e dell'insieme $\mathcal{E}$ andrebbero approfondite. In generale non tutte le funzioni da uno spazio di probabilità Ω verso $\mathbb{R}$ sono misurabili.

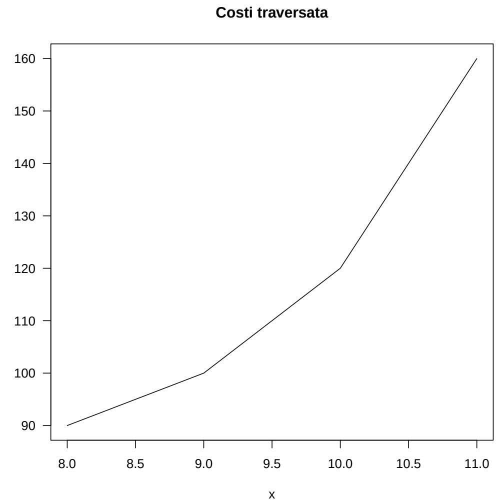


Figura 2.4: Funzione dei costi

Dopo che l'evento E è stato identificato è possibile calcolare la probabilità che si realizzi utilizzando la legge di probabilità P. Tale probabilità corrisponde dunque alla probabilità di osservare dei costi come specificato da E'.

Definizione 12. Sia $(\Omega, \mathcal{E}, P)$ uno spazio di probabilità e $g : \Omega \to \mathbb{R}$ misurabile. Allora g è chiamata variabile aleatoria.

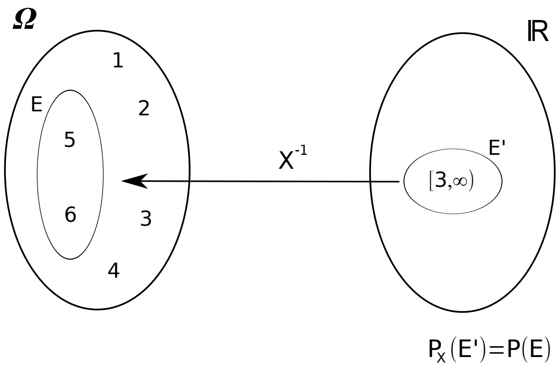


Figura 2.5: Preimmagine

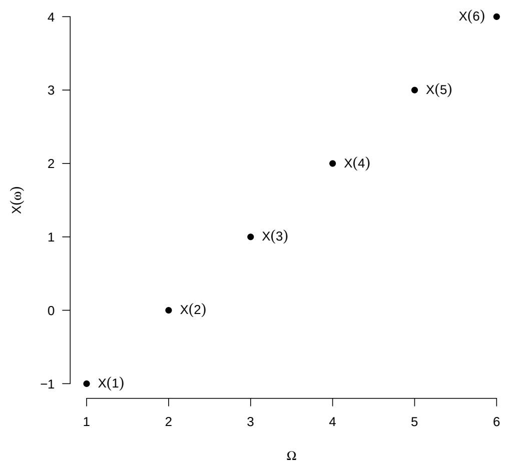


Figura 2.6: Variabile aleatoria

Osservazione 3. In statistica si è soliti indicare una variabile aleatoria con una lettera maiuscola, tipicamente X. Ricordatevi dunque che malgrado il nome ingannevole, la variabile aleatoria X è una funzione! Anziché scrivere ogni volta "la variabile aleatoria X" nell'ambito di questo corso utilizzeremo l'abbreviazione $V.A.$ X.

Lasciamo ai matematici lo studio delle condizioni rispetto alle quali una funzione X è una variabile aleatoria. Per noi questa condizione sarà sempre soddisfatta. Ci domandiamo piuttosto perché una variabile aleatoria debba soddisfare una simile condizione di misurabilità. Il motivo è dato dalla necessità di assegnare una probabilità agli eventi E' di $\mathbb{R}$.

Esempio 15. Consideriamo l'esperimento aleatorio di lanciare una volta un dado e di fare una scommessa sul possibile esito ω dell'esperimento. Supponiamo dunque che la vincita o la perdita in franchi, notata X, dipenda dall'esito dell'esperimento e sia descritta dalla funzione $X(\omega) = \omega - 2$. Siamo interessati all'evento $E' = [3, \infty)$ di guadagnare almeno 3 franchi. Il nostro esperimento aleatorio è formalizzato da $\Omega = \{1, 2, 3, 4, 5, 6\}$ e P distribuzione uniforme discreta. La variabile aleatoria X può essere rappresentata tramite un diagramma di Venn simile alla Figura 2.3 oppure graficamente come in Figura 2.6.

Consideriamo ora la probabilità dell'evento E'. Come si calcola tale probabilità? L'idea è semplicissima. Guardando il diagramma di Venn 2.5 è facile convincersi che l'evento E' si avvera se e solo se l'evento $E = X^{-1}(E')$ si realizza. Nel nostro esempio vinceremo almeno 3 franchi se il lancio del dado risulterà essere un numero superiore o alla peggio uguale a 5. L'evento E nel nostro caso è dato dal sottoinsieme $\{5, 6\}$ di Ω. La probabilità di vincere almeno 3 franchi è dunque uguale alla probabilità che esca un cinque o un sei. Scriveremo

$$P_X(E') := P(X^{-1}(E')) = P(E) = \frac{1}{3}. \tag{2.4}$$

Osservate attentamente l'uguaglianza 2.4. Quando ci si riferisce alla probabilità dell'evento E' occorre farlo aggiungendo la lettera X alla probabilità P in quanto E' non è evento di Ω ma di $\mathbb{R}$, l'insieme di arrivo della *V.A.* X. È possibile dimostrare che P_X, così come definita tramite l'uguaglianza 2.4, possiede tutte le proprietà di una legge di probabilità su $\mathbb{R}$. La tripla $(\mathbb{R}, \mathcal{E}, P_X)$ costituisce un nuovo spazio di probabilità e P_X è chiamata la legge di probabilità indotta o distribuzione indotta dalla variabile aleatoria X su $\mathbb{R}$. Per semplicità si è soliti chiamare P_X legge di probabilità di X.

Esempio 16. Lanciamo 3 volte una moneta. In un singolo lancio la probabilità di ottenere croce è p con $0 \leq p \leq 1$. Per un singolo lancio utilizzeremo la convenzione di indicare con t l'esito testa e con c l'esito croce. Modelliamo questo esperimento con lo spazio campionario Ω^3 contenente gli 8 possibili esiti

$$\begin{array}{llll} \omega_1 = (t,t,t) & \omega_2 = (c,t,t) & \omega_3 = (t,c,t) & \omega_4 = (t,t,c) \\ \omega_5 = (t,c,c) & \omega_6 = (c,t,c) & \omega_7 = (t,c,c) & \omega_8 = (c,c,c) \end{array}.$$

È immediato verificare che sotto l'ipotesi di indipendenza tra un lancio e l'altro la probabilità di ogni singolo esito è data da

$$Q(\{\omega_i\}) = p^{\#c}(1-p)^{\#t}$$

dove "$\#c$" indica il numero di "c" in ω_i. La tripla $(\Omega^3, \bigotimes_{i=1}^{n} \mathcal{E}_i, Q)$ è un esperimento statistico e forma uno spazio di probabilità.

Su questo spazio definiamo ora la *V.A.* reale $X : \Omega^3 \to \mathbb{R}$, $\omega \mapsto$ "il numero di croci ottenute" che in pratica consiste nel contare il numero di "c" presenti in ω. Vale dunque

$$\begin{array}{llll} X(\omega_1) = 0 & , X(\omega_2) = 1 & , X(\omega_3) = 1 & , X(\omega_4) = 1 \\ X(\omega_5) = 2 & , X(\omega_6) = 2 & , X(\omega_7) = 2 & , X(\omega_8) = 3 \end{array}.$$

La $V.A.$ X assumerà così i valori $0,1,2,3$. Con che probabilità la variabile aleatoria X sarà uguale a 0? Guardando la tabella dei possibili valori deduciamo che $X=0$ se e solo se $\omega=\omega_1$, quindi

$$P_X(\{0\}) = P(X=0) = P\left(\{\omega_1\}\right).$$

In maniera analoga, dato l'evento $E'=[1,2]\subset\mathbb{R}$ diremo che E' si è realizzato se $X(\omega)\in E'$, dove ω denota la realizzazione su Ω^3. La probabilità che E' si realizzi è uguale a

$$P_X(E') = P(X^{-1}(E')) = P(\{\omega_2,\omega_3,\omega_4,\omega_5,\omega_6,\omega_7\}).$$

La probabilità dell'evento $E'=\{0,3\}$ è uguale a $P(\{\omega_1,\omega_8\})$ e la probabilità di $P_X\left([0,2)\right)$ è semplicemente uguale a $P(\{\omega_1,\omega_2,\omega_3,\omega_4\})$.
La variabile aleatoria X induce una legge di probabilità su $(\mathbb{R},\mathcal{E})$ che abbiamo chiamato P_X.
La tripla $(\mathbb{R},\mathcal{E},P_X)$ è uno spazio di probabilità. In particolare, P_X è la distribuzione Binomiale $Bin(3,p)$ la cui famiglia parametrica è data da

$$\mathcal{P} = \left\{ f(x;p) = \left(\begin{array}{c} 3 \\ x \end{array} \right) p^x(1-p)^{3-x} : \ x\in\{0,1,2,3\} \text{ e } p\in[0,1] \right\}.$$

2.5 Esperimento statistico e variabili aleatorie

La medesima situazione studiata nella Sezione 2.3 la ritroviamo nell'ambito delle variabili aleatorie. Prendiamo una variabile aleatoria X a valori in $\mathbb{R}$ definita su uno spazio di probabilità $(\Omega, \mathcal{E}, P)$. Come appena studiato, la *V.A.* X genera un nuovo esperimento aleatorio su $\mathbb{R}$, notato $(\mathbb{R}, \mathcal{E}, P_X)$, dove P_X è la legge di probabilità indotta. Consideriamo ora un campione di n variabili aleatorie $X_1, X_2, ..., X_n$ mutualmente indipendenti e distribuite identicamente a X. Questo n-campione estratto da X induce uno spazio prodotto notato

$$\left(\mathbb{R}^n, \bigotimes_{i=1}^{n} \mathcal{E}_i, Q_{X_1,...,X_n} = \bigotimes_{i=1}^{n} P_{X,i}\right). \tag{2.5}$$

Se la legge di probabilità P_X appartiene ad una famiglia parametrica di distribuzioni con funzione di densità $f(\cdot;\theta)$, varrà per lo stesso discorso fatto in precedenza che la distribuzione $Q_{X_1,...,X_n}$ apparterrà alla famiglia

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

Come in precedenza la funzione di densità (o probabilità nel caso discreto) di $Q_{X_1,...,X_n}$ è il prodotto delle n funzioni di densità $f(x_i;\theta)$. Poiché le *V.A.* $X_1, X_2, ..., X_n$ sono identicamente distribuite, la forma delle funzioni di densità $f(x_i;\theta)$ è la stessa per ogni variabile i.

Esempio 17. (Continuazione Esempio 16) Disponiamo di osservazioni indipendenti di n giocate, notate rispettivamente $X_1, X_2, \ldots, X_n$. La loro distribuzione congiunta avrà quale funzione di probabilità (ci troviamo di fronte a variabili aleatorie discrete)

$$\mathcal{P} = \left\{ f(x_1, ..., x_n; p) = \prod_{i=1}^{n} \binom{3}{x_i} p^{x_i}(1-p)^{3-x_i} : \ x_i \in \{0,1,2,3\} \text{ e } p \in [0,1] \right\}.$$

Utilizzando le proprietà delle potenze la funzione f può essere riscritta come

$$f(x_1, ..., x_n; p) = p^{\sum_{j=1}^{n} x_j}(1-p)^{3n-\sum_{j=1}^{n} x_j} \prod_{i=1}^{n} \binom{3}{x_i}.$$

Definizione 13. Dato un esperimento statistico $(\mathbb{R}, \mathcal{E}, P_X)$ diremo che $\mathcal{P}$ è correttamente specificata se esiste almeno un valore $\theta_0 \in \Theta$ tale per cui $f(\cdot;\theta_0)$ è la vera funzione di densità/probabilità di X.

2.6 Campione

Definizione 14. Chiamiamo *campione* un particolare sottoinsieme di numerosità n della popolazione obiettivo utilizzato ai fini dell'inferenza. Nel caso di un esperimento statistico con *campione* intendiamo l'osservazione di un n-campione (inteso come la realizzazione di n variabili aleatorie[5] *i.i.d*).

Ci sono alcune osservazioni su cui chinarsi.

Osservazione 4. La prima osservazione riguarda il caso di un esperimento statistico in cui si desidera stimare la distribuzione di una variabile aleatoria X tramite un n-campione. È fondamentale ai fini della comprensione di quanto seguirà rendersi conto che l'n-campione $X_1, ..., X_n$ è una collezione di variabili aleatorie che generano su $\mathbb{R}^n$ uno spazio di probabilità (confronta Sezione 2.5). Così come una variabile aleatoria X induce una legge di probabilità P_X su $(\mathbb{R}, \mathcal{E})$, anche $X_1, X_2, ..., X_n$ inducono una legge di probabilità su $\left(\mathbb{R}^n, \bigotimes_{i=1}^{n} \mathcal{E}_i\right)$. La probabilità indotta è stata indicata con $Q_{X_1,...,X_n}$ (confronta Sezione 2.3).

Osservazione 5. Osservato l'esperimento, il ricercatore avrà a disposizione le realizzazioni di $X_1, ..., X_n$ che sono dei numeri. Le realizzazioni saranno disposte in modo ordinato in una lista notata $(x_1, ..., x_n)$ che chiameremo vettore. La notazione maiuscola denota la variabile aleatoria (funzione!) mentre la notazione minuscola ne indica il particolare valore realizzato. Ad esempio $x_1 = X_1(\omega)$.

Osservazione 6. Consideriamo la stima della distribuzione statistica di una caratteristica della popolazione obiettivo. In tal caso il campione è definito come un sottoinsieme della popolazione obiettivo. Questa definizione non considera il modo come si è ottenuto questo sottoinsieme.

- Che regola di selezione è stata adottata?
- Come sono stati selezionati gli n individui o le n unità dalla popolazione?

Vedremo che applicando dei criteri aleatori è possibile ricondurre il problema del campionamento da una popolazione campione a quello relativo alla

[5] D'ora innanzi utilizzeremo semplicemente l'abbreviazione *i.i.d.* per indicare variabili aleatorie *indipendenti* ed *identicamente distribuite*.

costruzione di un n-campione in un esperimento statistico. Affronteremo però questo tema nel corso del prossimo capitolo dove analizzeremo le varie tecniche di campionamento. Per il momento basterà ricordare che anche nel caso di una popolazione obiettivo la cui distribuzione statistica è sconosciuta il *campione* osservato sarà il risultato di un procedimento aleatorio e quindi è da considerarsi anch'esso come la realizzazione di n variabili aleatorie.

Ma perché, nell'analisi di una popolazione obiettivo, effettuare un'indagine campionaria e non un'indagine esaustiva? I motivi sono diversi; fra i principali annoveriamo:

- una rilevazione completa è impossibile (costi troppo elevati, impossibilita di raggiungere tutti gli individui);
- determinare le caratteristiche delle unità campionate ne determina la distruzione (ad esempio, misurare la durata di vita di una lampadina);
- La popolazione non è definibile in termini della sua numerosità e quindi non ha senso parlare di indigine esaustiva (ad esempio una data tecnologia permette di produrre lampadine la cui durata di vita è aleatoria. Quanto vale N in questo caso?).
- è necessaria una particolare accuratezza nella rilevazione e quindi a causa delle risorse limitate solo un numero ristretto di unità può essere osservato;
- tempi d'esecuzione elevati se confrontati alla necessità d'informazione (sondaggi elettorali).

2.7 Distribuzione campionaria, statistica, stimatore corretto

Il concetto di distribuzione di una *V.A.* (caso univariato) o di un insieme di *V.A.* (caso multivariato) ci accompagnerà per tutta la durata di questo corso. Abbiamo visto nei paragrafi precedenti il concetto di esperimento statistico e di n-campione ad esso associato. Affrontiamo ora la definizione di distribuzione campionaria che altro non è che la distribuzione congiunta $Q_{X_1,...,X_n}$ delle n variabili aleatorie $X_1, ..., X_n$ vista in precedenza.

Definizione 15. Distribuzione campionaria. Sia $X_1, X_2, ..., X_n$ un campione aleatorio di numerosità n. Si dice distribuzione campionaria di $X_1, X_2, ..., X_n$ la distribuzione o legge di probabilità congiunta di $X_1, X_2, ..., X_n$.

Perché la distribuzione campionaria è così importante? La sua importanza risiede nell'uso che se ne farà ai fini della stima del o dei parametri sconosciuti. Infatti,

- definito il modello di probabilità e l'esperimento statistico ad esso associato
- ricavata la distribuzione campionaria dei dati in nostro possesso

si passa alla fase di stima. Lo scopo della teoria della stima è quello di costruire delle funzioni $T(X_1, ..., X_n)$ delle variabili aleatorie osservate con cui produrre delle stime puntuali $T(x_1, ..., x_n)$ del vettore di parametri[6] sconosciuti θ. Ancora una volta è fondamentale fare una netta distinzione fra il campione $X_1, ..., X_n$ quale insieme ordinato di *V.A.* ed il campione osservato $x_1, ..., x_n$. Come precedentemente osservato, la sequenza $x_1, ..., x_n$ contiene dei *numeri* e raccoglie le rispettive realizzazioni delle *V.A.* $X_1, ..., X_n$, cioè

$$x_1 = X_1(\omega),\ x_2 = X_2(\omega),\ ..., x_n = X_n(\omega).$$

Definizione 16. Statistica. Una statistica T è una funzione di variabili aleatorie osservabili (e quindi a sua volta variabile aleatoria osservabile) che non dipende da alcun parametro incognito.

Esempio 18. Alcuni esempi di statistiche sono

1. La media $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.
2. $X_{\max} = \max\{X_1, X_2, ..., X_n\}$.
3. Lo stimatore corretto della varianza $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$.
4. Siano $X_1, ..., X_n \sim i.i.d.\ N(3, \sigma^2)$, σ^2 sconosciuto. La variabile aleatoria

$$Z = \frac{\overline{X} - 3}{\sigma}$$

non è una statistica in quanto la varianza σ^2 (e quindi anche σ) non è nota.

[6] Come già accennato in precedenza, il nome assegnato ai vari parametri sconosciuti cambia a seconda delle circostanze. Ad esempio, il valore atteso di una *V.A.* è generalmente indicato con il simbolo μ. Se il parametro sconosciuto da stimare fosse la varianza si utilizzerebbe σ^2.

Osservazione 7. Ci preme sottolineare che una statistica è una variabile aleatoria. La distribuzione di $\overline{X}$ ad esempio, dipende dalla distribuzione congiunta di $X_1, ..., X_n$. Lo stesso vale per $X_{\max}$, S^2 e qualsiasi altra funzione misurabile di $X_1, ..., X_n$.

Definizione 17. Stimatore corretto. Sia θ un parametro sconosciuto. Diremo che $T(X_1, ..., X_n)$ è uno stimatore corretto di θ se

$$E\left(T(X_1, ..., X_n)\right) = \theta.$$

Esempio 19. Sia $X_1, ..., X_n$ un campione di variabili aleatorie *i.i.d.* (μ, σ^2). La media campionaria è uno stimatore corretto di μ in quanto

$$E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}n\mu = \mu.$$

Anche

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

è uno stimatore corretto di σ^2. Per contro, ammettiamo che μ sia conosciuto,

$$\widetilde{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \mu\right)^2$$

non è uno stimatore corretto di σ^2.

Esercizio 3. Si dimostri che lo stimatore $\widetilde{S}^2$ non è uno stimatore corretto di σ^2

Terminiamo il capitolo con due ulteriori esempi di modelli parametrici.

Esempio 20. Rendimenti azionari. Sia P_t il prezzo di una certa azione il giorno[7] t. Il rendimento logaritmico giornaliero è definito come

$$R_t = \ln\frac{P_t}{P_{t-1}}.$$

Assumiamo che

$$\mathcal{P} = \left\{f(R;\theta) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}\left(R-\mu\right)^2\right) :\ \theta = (\mu, \sigma^2) \in \mathbb{R}\times\mathbb{R}^+\right\}.$$

[7] In questo caso viene naturale utilizzare l'indice t anziché l'indice i per indicare la successione di osservazioni di R.

Supponiamo di avere un campione comprendente gli ultimi 100 rendimenti logaritmici giornalieri $R_t, R_{t-1}, ..., R_{t-99}$ e che i rendimenti a date diverse siano fra loro indipendenti. Vogliamo calcolare

$$P\left(R_{t+1} \leq -5\%\right).$$

Esempio 21. Funzione di produzione Cobb-Douglas. Definiamo la seguente relazione

$$Y := \ln Q_t = \beta_1 + \beta_2 \ln L_t + \beta_3 \ln K_t + U_t \tag{2.6}$$

dove

1. Q_t rappresenta la quantità prodotta l'anno t di un certo bene.

2. L_t rappresenta le ore di lavoro utilizzate per la produzione di Q_t.

3. K_t rappresenta la quantità di capitale utilizzato.

4. U_t è una variabile aleatoria, non osservabile che racchiude tutti quei fattori non sistematici che non sono stati inclusi nel modello. Assumiamo: $U_t \sim i.i.d.\ N(0, \sigma^2)$.

La famiglia parametrica $\mathcal{P}$ è uguale a

$$\mathcal{P} = \left\{ \begin{array}{l} f(y;\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\left(\ln q_t - \beta_1 - \beta_2 \ln l_t - \beta_3 \ln k_t\right)^2\right) : \\ \theta = (\beta_1, \beta_2, \beta_3, \sigma^2) \in \mathbb{R}^3 \times \mathbb{R}^+ \end{array} \right\}.$$

In quest'ultimo esempio la funzione di produzione Cobb-Douglas

$$Q = \exp(\beta_1) L^{\beta_2} K^{\beta_3} \tag{2.7}$$

ci è stata suggerita dalla teoria economica. Ne segue che i risultati empirici ottenuti sono interpretabili nel contesto dei modelli (micro-) economici ben conosciuti. Ad esempio è possibile interpretare i coefficienti β_2 e β_3 come le elasticità della produzione:

$$\frac{\partial Q}{\partial L} / \frac{Q}{L} = ...$$

L'equazione (2.7) è deterministica. In essa la produzione Q è una funzione deterministica delle quantità di lavoro L e capitale K. Nel modello (2.6) invece, Q è una variabile aleatoria in quanto funzione del termine d'errore[8] U.

[8] È bene non confondere il concetto di osservabilità di una variabile con il concetto di aleatorietà. Una variabile può essere deterministica ma non osservabile, deterministica osservabile, aleatoria non osservabile o aleatoria ed osservabile. Nel modello (2.6) gli U_t sono aleatori e non osservabili.

È bene fare presente quanto segue. Affinché il modello sia utilizzabile sono state fatte numerose ipotesi semplificatrici:

(a) forma funzionale della relazione tra le variabili Q, L e K: lineare nei logaritmi,

(b) costanza delle elasticità β_2 e β_3 nel tempo (progresso tecnologico?),

(c) assenza di errori di misurazione in Q, L e K.

Raramente la teoria economica ci permette di specificare completamente un modello. Generalmente altre ipotesi, non motivabili economicamente, sono necessarie.

- La normalità del termine d'errore U_t nell'Esempio (21) non è sostenibile in modo naturale usando esclusivamente la teoria economica.
- Questo implica la necessità di verificare tramite opportuni test statistici le ipotesi assunte.
- Talvolta l'analisi esplorativa dei dati ci aiuta nella comprensione del fenomeno in esame e ci permette di ottenere dei modelli più prossimi alla realtà.

Concludendo: le informazioni utilizzate nella specificazione dei modelli fanno capo alla

- teoria della probabilità,
- teoria economica,
- analisi esplorativa (si veda ad esempio la motivazione al premio Nobel 2003 per l'economia Robert F. Engle).

Capitolo 3

Campionamento

In questo capitolo studieremo come ottenere un campione aleatorio estratto da una popolazione obiettivo finita di numerosità N. Con piano o disegno di campionamento si intende il procedimento utilizzato per costruire il campione partendo da una popolazione finita o infinita. Del disegno di campionamento possiamo evidenziare

- la struttura del campione (liste, sottoliste, attributi, strati);
- le regole seguite per identificare gli insiemi di unità da inserire nel campione;
- la probabilità di inclusione delle singole unità;
- il modo con cui si determina la numerosità ottima del campione e la relativa frazione di campionamento. "... la numerosità ottima di un campione è quella che permette di ottenere gli obiettivi dell'indagine al minimo costo, e sarà il più piccolo numero in base al quale le stime ragiungono il livello di attendibilità atteso dal ricercatore" (Fabbris, p. 26).

La conoscenza della tecnica di campionamento utilizzata è essenziale ai fini della determinazione delle proprietà probabilistiche (distribuzione congiunta) del campione nonché della validità e correttezza dei risultati. Un piano di campionamento può infatti essere

- probabilistico, in tal caso le unità vengono estratte secondo un meccanismo aleatorio;
- deterministico o non probabilistico, le unità sono scelte dalla popolazione tramite una regola *deterministica*.

Un esempio di piano di campionamento deterministico è il seguente: dall'elenco telefonico estraggo, per ciascuna lettera dell'alfabeto, i primi 10 individui. Se da un lato risulta evidente la sua semplicità di implementazione, il metodo deterministico di campionamento presenta alcuni svantaggi non indifferenti. Esso non permette di calcolare il grado di precisione con cui viene eseguita la stima. Inoltre, la validità dei risultati ottenuti è fortemente legata alle informazioni utilizzate a priori per la scelta del campione. Se tali informazioni sono sbagliate oppure obsolete e non corrispondono più alla realtà, i risultati della ricerca saranno distorti. Per tale motivo in questo corso tratteremo esclusivamente piani di campionamento basati sulla selezione casuale delle unità (individui) della popolazione obiettivo. In particolare vedremo le tecniche di campionamento casuale semplice, di campionamento stratificato, di campionamento sistematico. Tutti questi tipi di campionamento si basano su delle tecniche di selezione.

3.1 Campionamento tramite selezione con reinserimento

Supponiamo di avere quale popolazione obiettivo l'insieme delle automobili immatricolate in Svizzera ad una certa data. La numerosità della popolazione è indicata con la lettera N. Una casa automobilistica potrebbe essere interessata alla distribuzione del colore dell'automobile così da poi prevedere il consumo di vernice. Per semplicità supponiamo che i colori siano solo tre: nero, bianco e rosso e che la distribuzione sia tale per cui il 45% delle auto immatricolate è nero, il 25% è bianco ed il 30% è rosso. Definiamo ora il seguente esperimento aleatorio definito sull'insieme $\Omega = \{1, 2, ...N\}$ che consiste nel scegliere a caso in maniera *equiprobabile* un numero compreso tra 1 e N. Lo spazio di probabilità è dunque composto dalla solita tripla $(\Omega, \mathcal{E}, P)$ dove in questo caso P è la distribuzione uniforme su Ω

$$P(\{i\}) = \frac{1}{N} \text{ per ogni } i \in \Omega.$$

Supponiamo di avere una scheda per ogni automobile immatricolata. Nella scheda sono riportati il propritario ed i dati tecnici del veicolo: colore, cilindrata, anno d'immatricolazione, ecc.. Numeriamo le schede da 1 a N. Ad ogni automobile immatricolata in Svizzera sarà così assegnato un numero $i \in \Omega$. Definiamo ora la variabile aleatoria (funzione!) X da Ω verso l'insieme dei colori $\Omega' = \{$"nero","bianco","rosso"$\}$ in modo tale che per ogni $i \in \Omega$ la *V.A.* X restituisca il colore della i-esima automobile

$$i \mapsto X(i) = \text{"il colore della } i\text{-esima automobile"} .$$

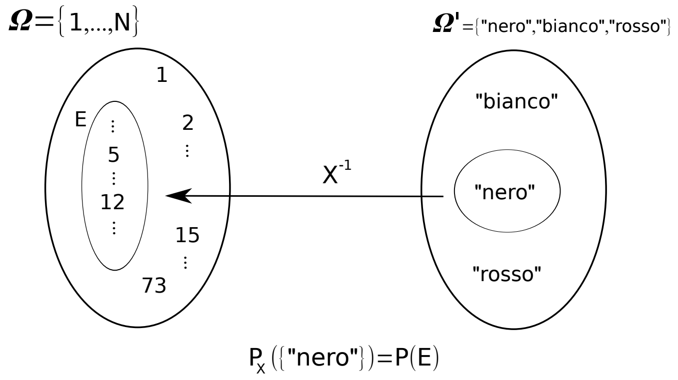


Figura 3.1: Estrazione colore automobile

Domanda 11. Qual è la legge di probabilità indotta dalla variabile aleatoria X, notata P_X, sull'insieme dei tre colori?

Risposta: la legge di probabilità P_X non è altro che la distribuzione statistica P_p della caratteristica in esame della popolazione obiettivo. Infatti, sia N_1 il numero di automobili di colore nero. Sappiamo che $\frac{N_1}{N} = 0.45$ è la frequenza relativa di automobili nere nella popolazione. Qual è la probabilità di selezionare una macchina nera? Questa probabilità corrisponde a

$$P_X(\{\text{“nero”}\}) = P(X = \text{“nero”}) = P(E)$$

dove E corrisponde agli indici delle schede delle automobili di colore nero. Il numero di esiti in E è evidentemente N_1. Come calcolato precedentemente la probabilità di ciascun esito è $\frac{1}{N}$ e quindi $P(E) = \frac{N_1}{N} = 0.45$. La situazione è rappresentata graficamente nella Figura 3.1.

La legge di probabilità P_X corrisponde esattamente alla distribuzione statistica P_p del colore della popolazione obiettivo.

Se la caratteristica in esame fosse stata la cilindrata in cm^3 dell'automobile anziché il suo colore, la variabile aleatoria X avrebbe una funzione di ripartizione F_X. Ancora una volta la situazione è riportata graficamente nella Figura 3.2.

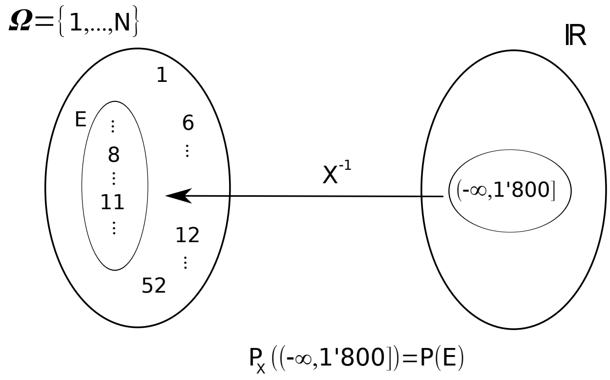


Figura 3.2: Estrazione cilindrata automobile

$$\begin{aligned}
F_X(x) &= P_X\left((-\infty, x]\right) \\
&= P(X \in (-\infty, x])) \\
&= P(E) \\
&= (\#\text{ unità con caratteristica } \leq x)\frac{1}{N} \\
&= \frac{\#\text{ unità con caratteristica } \leq x}{\#\text{ unità della popolazione}} \qquad (3.1)
\end{aligned}$$

Come si evince dalla (3.1) la funzione di ripartizione della variabile aleatoria X è identica alla funzione di ripartizione della cilindrata della popolazione obiettivo.

È dunque possibile, partendo da un semplice spazio di probabilità discreto la cui legge di probabilità è uniforme, costruire una variabile aleatoria X la cui distribuzione P_X è uguale alla distribuzione dell'intera popolazione obiettivo.

$$(\Omega, \mathcal{E}, P) \underset{X}{\longrightarrow} (\mathbb{R}, \mathcal{E}, P_X)$$

La problematica della non conoscenza della distribuzione statistica P_p della popolazione obiettivo risulta quindi identica alla problematica della stima della legge di probabilità di una variabile aleatoria!

Ripetiamo l'esperimento n volte in maniera indipendente l'una dall'altra, reinserendo nell'urna il numero scelto dopo ogni estrazione e mescolando nuovamente molto bene. Il nuovo spazio campionario è $\Omega^n = \{1, 2, ...N\}^n$, le liste di lunghezza n i cui elementi appartengono all'insieme $\Omega = \{1, 2, ...N\}$. Il numero di elementi di Ω^n è N^n. Poiché il numero selezionato ad ogni estrazione è reinserito nell'urna, è possibile osservare vettori in cui un numero si ripete più volte. Ad esempio, se la popolazione contasse $N = 5$ elementi e $n = 3$ (sorteggiamo 3 volte) un possibile esito potrebbe essere $\omega = (2, 5, 2)$. Qual è la probabilità di osservare un qualsiasi esito[1] $\omega \in \Omega^n$? Poiché le estrazioni sono indipendenti l'una dall'altra e ad ogni estrazione la probabilità di estrarre un qualsiasi numero è pari a $1/N$ vale

$$Q(\omega) = \frac{1}{N^n} \text{ per ogni } \omega \in \Omega^n \ . \tag{3.2}$$

Ecco così definito lo spazio di probabilità $\left(\Omega^n, \overset{n}{\underset{i=1}{\otimes}} \mathcal{E}_i, Q = \overset{n}{\underset{i=1}{\otimes}} P_i\right)$ relativo al nostro nuovo esperimento. Possiamo ora definire le n variabili aleatorie che notiamo con X_i, $i = 1, ..., n$ tramite la procedura

$$X_i(\omega) := \begin{cases} \text{1. Estraggo l' } i-\text{esimo elemento da } \omega\text{, notato } \omega(i) \in \{1, 2, ...N\}. \\ \text{2. Associo la cilindrata della } \omega(i)-\text{esima automobile.} \end{cases}$$

Proseguendo con l'esempio in cui $N = 5$ e $n = 3$, data la realizzazione $\omega = (2, 5, 2)$ abbiamo $\omega(1) = 2$, $\omega(2) = 5$, $\omega(3) = 2$ (praticamente abbiamo selezionato due volte la seconda automobile ed una volta la quinta). Il valore che la *V.A.* X_i assume dipende *esclusivamente* dall'esito parziale della i-esima estrazione. Poiché le n estrazioni sono per costruzione indipendenti anche le variabili aleatorie X_i lo saranno e la legge di probabilità $Q_{X_1,...,X_n}$ indotta da $X_1, ..., X_n$ su $\mathbb{R}^n$ è semplicemente il prodotto delle singole leggi di probabilità P_{X_i}. L'esperimento aleatorio

$$(\Omega, \mathcal{E}, P) \underset{X}{\longrightarrow} (\mathbb{R}, \mathcal{E}, P_X)$$

viene così generalizzato in un esperimento statistico

$$(\Omega^n, \overset{n}{\underset{i=1}{\otimes}} \mathcal{E}_i, Q) \underset{X_1,...,X_n}{\longrightarrow} (\mathbb{R}^n, \overset{n}{\underset{i=1}{\otimes}} \mathcal{E}_i, Q_{X_1,...,X_n}) \ .$$

---

[1] Ora ω denota un qualsiasi elemento di Ω^n ed avrà quindi la forma

$$\omega = \underbrace{(5, 12, ..., 4)}_{n \text{ numeri tra 1 e } N}$$

di una lista a n componenti.

Quando la distribuzione statistica P_p della popolazione obiettivo è sconosciuta possiamo pensare di stimare P_X andando a selezionare da una particolare famiglia parametrica di distribuzioni $\mathcal{P}$ la distribuzione che riteniamo migliore rispetto all'evidenza empirica (le n osservazioni) in nostro possesso. È quindi possibile ricondurre il problema della stima della distribuzione P_p della popolazione obiettivo alla definizione di esperimento statistico e di n-campione ad esso associato. In altre parole, le n variabili aleatorie *i.i.d.* $X_1, ..., X_n$ costruite estraendo a caso con reimmissione n unità dalla popolazione obiettivo costituiscono un n-campione di un esperimento aleatorio $(\mathbb{R}, \mathcal{E}, P_X)$. Poiché la legge di probabilità P_X è sconosciuta, scriveremo $(\mathbb{R}, \mathcal{E}, \mathcal{P})$ dove $\mathcal{P}$ rappresenta la famiglia parametrica a cui P_X appartiene.

La tecnica di selezione casuale con reinserimento utilizzata precedentemente per la costruzione di un n-campione a partire da una data popolazione obiettivo di numerosità N è riassunta nel modo seguente.

1. A ciascuna delle N unità della popolazione viene assegnato univocamente un numero da 1 a N.

2. Si estrae a caso una sola pallina da un'urna contenente N palline numerate da 1 a N. Il numero indicato sulla pallina corrisponde all'unità del campione da selezionare per l'osservazione della caratteristica o attributo in esame.

3. Si reintroduce la pallina estratta e si mescola nuovamente.

4. Si ripetono i punti 2. e 3. per un totale di n volte.

Le n osservazioni, raccolte in una lista $(x_1, ..., x_n)$, costituiscono una realizzazione delle n variabili aleatorie *i.i.d.* $X_1, ..., X_n$ distribuite secondo la legge di probabilità definita dalla distribuzione della popolazione obiettivo.

3.2 Campionamento tramite selezione senza reinserimento

Questa tecnica è identica alla precedente con l'unica differenza che le palline estratte non vengono reinserite nell'urna. Cosa cambia rispetto al paragrafo precedente? Lo spazio campionario è lo stesso? Possiamo ancora parlare di esperimento statistico? Quali sono le proprietà della nuova legge di probabilità Q? Per rispondere a queste domande partiamo innanzi tutto dalla definizione del nuovo spazio campionario. Supponiamo per un momento di utilizzare ancora Ω^n. Poiché la tecnica di estrazione è senza reinserimento lo spazio Ω^n conterrà esiti che non osserveremo mai. Ad esempio, prendiamo $N = 5$, $n = 2$ e l'esito $\omega = (2, 2)$. Ovviamente un simile ω non sarà mai osservabile in quanto la selezione ora è senza reinserimento. In teoria potremmo escluderlo dallo spazio campionario. Tuttavia per semplicità nella definizione del nostro spazio campionario lo lasciamo. L'esito $\omega = (2, 2)$ avrà semplicemente probabilità zero. In generale, dato un numero n qualsiasi di estrazioni, tutti gli esiti ω in cui un numero si ripete due o più volte avranno probabilità zero.

Dobbiamo ora calcolare la probabilità di un esito ω le cui componenti sono tutte diverse fra loro. Quanti esiti del genere ci sono? Per rispondere a questa domanda calcoliamo quante liste di lunghezza n si possono costruire partendo da un numero N di elementi diversi. Tenendo conto dell'ordine (disposizioni), ci sono $D_N^n = N!/(N-n)!$ modi diversi per estrarre n elementi da un insieme di N elementi. Se l'esperimento è condotto correttamente gli esiti con probabilità positiva saranno tutti equiprobabili. La probabilità di osservare un esito ω a componenti tutte diverse è dunque

$$Q(\omega) = \frac{1}{\#\text{ esiti a componenti diverse}} = \frac{(N-n)!}{N!} \ .$$

La nuova legge di probabilità Q è data da

$$Q(\omega) = \begin{cases} 0 & \text{se } \omega \text{ contiene almeno due valori uguali,} \\ \frac{(N-n)!}{N!} & \text{se i valori in } \omega \text{ sono tutti diversi fra loro.} \end{cases}$$

ed è dunque diversa dalla legge di probabilità vista nel caso di un'estrazione con reinserimento (confronta la 3.2).

Applichiamo la definizione delle n variabili aleatorie X_i, $i = 1, ..., n$ al nuovo spazio di probabilità:

$$X_i(\omega) := \begin{cases} \text{1. Estraggo l' } i-\text{esimo numero da } \omega, \text{ notato } \omega(i) \in \{1, 2, ...N\}. \\ \text{2. Associo la cilindrata dell'} \omega(i)-\text{esima automobile.} \end{cases}$$

Il nome delle variabili aleatorie è lo stesso ma la loro distribuzione è cambiata. In particolare, $X_1, ..., X_n$ non sono più indipendenti. Questo perché l'esito della variabile aleatoria X_i dipende dall'esito delle precedenti $V.A.$ $X_1, ..., X_{i-1}$. Per mostrare la dipendenza facciamo un esempio prendendo quale popolazione obiettivo il colore delle automobili immatricolate. Supponiamo che nella popolazione obiettivo di numerosità N ci siano k automobili rosse. Calcoliamo

$$P(X_2 = \text{“}rosso\text{”} \mid X_1 = \text{“}rosso\text{”}) = \frac{k-1}{N-1}$$

mentre

$$P(X_2 = \text{“}rosso\text{”} \mid X_1 \neq \text{“}rosso\text{”}) = \frac{k}{N-1}.$$

Se X_1 e X_2 fossero indipendenti le due probabilità condizionate appena calcolate dovrebbero essere uguali fra loro e uguali a $P(X_2 =$“$rosso$”$)$. Le due probabilità condizionate sono fra loro diverse e da ciò concludiamo che X_1 e X_2 non sono indipendenti. Questa conclusione è generalizzabile alle n variabili aleatorie $X_1, ..., X_n$ costruite tramite selezione senza reinserimento. La tripla $(\mathbb{R}^n, \bigotimes_{i=1}^{n} \mathcal{E}_i, Q_{X_1,...,X_n})$ è uno spazio di probabilità ma, sebbene utile ai fini dell'inferenza statistica, non costituisce un esperimento statistico. Le n *V.A.* X_i non sono *i.i.d.* e pertanto non costituiscono un n-campione. Per quanto riguarda questo corso utilizzeremo quindi il termine *n-campione di un esperimento statistico* solo in referimento a osservazioni di variabili aleatorie *i.i.d.*.

3.3 Campionamento tramite selezione sistematica

La selezione sistematica si effettua nella maniera seguente.

1. A ciascuna delle N unità della popolazione assegnamo univocamente un numero da 1 a N.

2. Calcoliamo il passo di campionamento, notato k, definito come il rapporto tra N (la numerosità della popolazione) e n (la numerosità del campione):
$$k = \frac{N}{n}.$$

3. Selezioniamo un numero a caso r compreso tra 1 e k:
$$1 \leq r \leq k. \tag{3.3}$$

4. Includiamo nel campione le n unità aventi posizioni
$$r,\ r+k,\ r+2k, \ldots, r+(n-1)k.$$

Osservazione 8. Il salto che si effettua tra due unità consecutive selezionate si chiama *passo di campionamento*. In pratica la selezione sistematica consiste nel partizionare[2] la popolazione in k sottoinsiemi di numerosità n per poi selezionarne uno a caso. All'interno di ogni sottoinsieme la distanza tra le unità è costante ed uguale k.

Esercizio 4. Costruite lo spazio di probabilità che utilizzereste per eseguire un campionamento sistematico. Come sono definite le variabili aleatorie $X_1, ..., X_n$? Sono indipendenti? Quanto vale la probabilità di estrarre un qualsiasi elemento della popolazione?

Quando N/n non è un numero intero, utilizzeremo la tecnica della *lista circolare*. In pratica, la procedura sarà modificata nel modo seguente.

1. Calcoliamo k^*, il numero intero più vicino a $\frac{N}{n}$.

2. Estraiamo un numero casuale compreso tra 1 e N (si noti la differenza con (3.3)):
$$1 \leq r \leq N.$$

3. Includiamo nel campione le n unità aventi posizioni
$$r,\ r + k^*,\ r + 2k^*, \ldots,\ r + (n-1)k^*$$
dove, una volta esaurita la lista all'unità N-esima senza aver estratto tutte le n unità campionarie, si continuerà a contare dalla prima unità (da qui il termine lista circolare).

3.4 Varianza di - e covarianza fra - somme pesate di variabili aleatorie

Molti stimatori in statistica ed econometria si possono rappresentare come particolari *somme pesate* di variabili aleatorie. Al fine di studiare alcune delle loro caratteristiche quali ad esempio la correttezza (Definizione 17), la varianza o la covarianza con altre statistiche sarà necessario utilizzare le proprietà del valore atteso, della varianza e della covarianza di somme di variabili aleatorie.

Esempio 22. Rendimenti di scala costanti. Nell'Esempio 21 abbiamo visto come si potrebbe stimare una funzione di produzione di tipo Cobb-Douglas. I parametri β_2 e β_3 corrispondono rispettivamente alle elasticità della produzione rispetto al lavoro e al capitale. Potremmo chiederci se una simile

[2]Una partizione di un insieme Ω si può definire come una collezione di sottoinsiemi di Ω non vuoti, mutuamente disgiunti e tali che la loro unione è l'insieme Ω stesso. (Definizione tratta da Wikipedia, Partizione (teoria degli insiemi)).

tecnologia possiede rendimenti di scala costanti. In termine dei coefficienti β_2 e β_3 tale ipotesi implica

$$\beta_2 + \beta_3 = 1. \tag{3.4}$$

Tuttavia β_2 e β_3 sono dei parametri sconosciuti che andranno stimati tramite opportune statistiche $\widehat{\beta}_i = T_i(Q_1, L_1, K_1, ..., Q_n, L_n, K_n)$ $i = 1, 2, 3$. $\widehat{\beta}_2$ e $\widehat{\beta}_3$ sono quindi delle variabili aleatorie, generalmente correlate fra loro in quanto funzioni delle stesse variabili aleatorie. Per verificare tramite un opportuno test d'ipotesi la validità dell'ipotesi di rendimenti di scala costanti procederemo come segue:

1. sostituiremo i parametri sconosciuti β_2 e β_3 nella (3.4) con i rispettivi valori stimati, $\widehat{\beta}_2$ e $\widehat{\beta}_3$.

2. calcoleremo la differenza (o scarto) tra il valore osservato $\widehat{\beta}_2 + \widehat{\beta}_3$ ed il valore teorico che in questo caso è 1.

3. Se la differenza in valore assoluto risulterà essere "grande" rifiuteremo l'ipotesi di rendimenti di scala costanti altrimenti non la rifiuteremo.

Per stabilire se la differenza così calcolata è significativa ("grande") o non significativa ("piccola") sarà necessario calcolare la varianza di $\widehat{\beta}_2 + \widehat{\beta}_3$ che come già sapete è pari a

$$V\left(\widehat{\beta}_2 + \widehat{\beta}_3\right) = V\left(\widehat{\beta}_2\right) + V\left(\widehat{\beta}_3\right) + 2Cov(\widehat{\beta}_2, \widehat{\beta}_3).$$

Poiché $\widehat{\beta}_2$ e $\widehat{\beta}_3$ risulteranno essere delle particolari somme pesate[3] delle $V.A.$ $Q_1, ..., Q_n$ e cioè

$$\widehat{\beta}_2 = \sum_{i=1}^{n} a_i Q_i \quad \text{e} \quad \widehat{\beta}_3 = \sum_{j=1}^{n} b_j Q_j$$

dovremo essere in grado di calcolare espressioni del tipo

$$Cov(\widehat{\beta}_2, \widehat{\beta}_3) = Cov(\sum_{i=1}^{n} a_i Q_i, \sum_{j=1}^{n} b_j Q_j).$$

3.4.1 Definizione della varianza e della covarianza

Rivediamo brevemente alcune definizioni. Consideriamo le due variabili aleatorie reali X e Y la cui funzione di densità[4] congiunta è notata $f_{x,y}(x, y)$ mentre

[3] I termini a_i e b_i sono dei pesi deterministici (dei numeri).

[4] Consideriamo il caso di variabili aleatorie continue. Il caso discreto è del tutto identico.

le funzioni di densità marginali sono notate rispettivamente $f_x(x)$ e $f_y(y)$. La varianza di X e definita come il valore atteso ... del quadrato ... degli scarti di X dal proprio valore atteso μ_X:

$$V(X) := E\left((X-\mu_X)^2\right) = \int_{\mathbb{R}} (x-\mu_X)^2 f_x(x)dx \ . \tag{3.5}$$

Quale formula alternativa vale anche

$$V(X) = E\left(X^2\right) - (\mu_X)^2 = \int_{\mathbb{R}} x^2 f_x(x)dx - (\mu_X)^2 \ . \tag{3.6}$$

La covarianza fra X e Y è definita come il valore atteso ... del prodotto ... degli scarti di X e Y dai rispettivi valori attesi:

$$Cov(X,Y) := E\left((X-\mu_X)(Y-\mu_Y)\right) = \int_{\mathbb{R}}\int_{\mathbb{R}} (x-\mu_X)(y-\mu_Y) f_{x,y}(x,y)dx\,dy \ . \tag{3.7}$$

Come per la varianza abbiamo una formula equivalente

$$Cov(X,Y) = E\left(XY\right) - \mu_X\mu_Y = \int_{\mathbb{R}}\int_{\mathbb{R}} xy\ f_{x,y}(x,y)dx\ dy - \mu_X\mu_Y \ .$$

Guardando alla definizione (3.7) di covarianza notiamo che abbiamo il prodotto di due termini $x-\mu_X$ il primo e $y-\mu_Y$ il secondo. Questo prodotto è ulteriormente moltiplicato per la rispettiva[5] "probabilità". Poiché la funzione di densità è sempre maggiore o uguale a zero ma mai negativa il prodotto

$$(x-\mu_X)(y-\mu_Y)f_{x,y}(x,y)$$

sarà negativo quando $(x-\mu_X)$ e $(y-\mu_Y)$ possiedono segno diverso ($+\ -$ oppure $-\ +$) e positivo quando $(x-\mu_X)$ e $(y-\mu_Y)$ hanno lo stesso segno ($+\ +$ o $-\ -$). La covarianza è la "somma" su tutti i possibili esiti pesati per la rispettiva "probabilità". Essa sarà dunque positiva se, mediamente, quando X è sopra o sotto il suo valore atteso lo è anche Y. La covarianza sarà invece negativa se, mediamente, quando X è sopra (sotto) il proprio valore atteso Y sarà sotto (sopra) μ_Y. Intuitivamente, sapendo ad esempio che X e Y sono correlate negativamente e che il valore osservato di X è maggiore al suo valore atteso mi aspetto che la realizzazione di Y sia inferiore a μ_Y.

[5]Ricordiamo che nel caso di variabili aleatorie continue la funzione di densità non indica la probabilità di un esito.

3.4.2 Tecnica di calcolo (difficoltà pari alla battaglia navale)

Iniziamo questo paragrafo prendendo due *V.A.* che chiameremo X e Y per le quali vale la seguente relazione

$$X = \sum_{i=1}^{n} a_i X_i \text{ e } Y = \sum_{j=1}^{m} b_j Y_j, \tag{3.8}$$

dove $X_1, ..., X_n$ e $Y_1, ..., Y_m$ sono delle *V.A.* tali per cui $Cov(X_i, Y_j) = c_{ij}$. c_{ij} rappresenta la covarianza (un numero quindi) fra le variabili aleatorie X_i e Y_j. Sottolineamo nuovamente che la variabile aleatoria X è semplicemente una somma pesata (si dice anche combinazione lineare) delle variabili aleatorie $X_1, ...X_n$. I pesi di tale somma sono i coefficienti a_i e sono dei numeri.

Esempio 23. Prendiamo $n = 3$, $m = 2$ ed i seguenti valori degli a_i e b_j:

$$\begin{array}{llll} \text{Indice } i\text{:} & a_1 = 2 & a_2 = -3 & a_3 = 1 \\ \text{Indice } j\text{:} & b_1 = 1 & b_2 = -2 & \end{array}$$

Abbiamo dunque

$$\begin{aligned} X &= 2X_1 - 3X_2 + X_3 \ , \\ Y &= Y_1 - 2Y_2 \ . \end{aligned}$$

Supponiamo che le covarianze c_{ij} siano conosciute. Come possiamo calcolare $Cov(X, Y)$ in funzione delle covarianze c_{ij}? Per spiegare come fare (la dimostrazione è data in appendice) facciamo un passo indietro e rivediamo la proprietà di linearità del valore atteso. Dal corso di analisi matematica sapete che una funzione $f : \mathbb{R} \to \mathbb{R}$ è lineare se soddisfa le seguenti due condizioni:

1. Per qualsiasi $x_1, x_2 \in \mathbb{R}$ vale che

$$f(x_1 + x_2) = f(x_1) + f(x_2), \tag{3.9}$$

2. Per qualsiasi $\lambda, x \in \mathbb{R}$ vale che

$$f(\lambda x) = \lambda f(x). \tag{3.10}$$

L'operatore valore atteso, notato E, che associa ad una *V.A.* X il suo valore atteso

$$E(X) = \sum_{k=1}^{N} x_k p(x_k) \text{ oppure } E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

è anch'esso lineare. Infatti, come visto nel corso di Statistica I, le proprietà (3.9) (3.10) sono soddisfatte per qualsiasi *V.A.* X, X_1 e X_2 nonché scalare $\lambda \in \mathbb{R}$:

$$E(X_1 + X_2) = E(X_1) + E(X_2) \text{ e } E(\lambda X) = \lambda E(X) \ . \tag{3.11}$$

Dalla proprietà di linearità del valore atteso segue in maniera immediata che

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E\left(X_i\right). \tag{3.12}$$

A parole possiamo dire che "il valore atteso di una somma pesata è uguale alla somma pesata dei valori attesi".

Esempio 24. (continuato) Supponiamo che i valori attesi di X_1, X_2 e X_3 siano rispettivamente uguali a 1, -2, 3. Il valore atteso di X sarà pertanto uguale a

$$\begin{aligned} E\left(X\right) &= E(2X_1 - 3X_2 + X_3) \\ &= 2E(X_1) - 3E(X_2) + E(X_3) \\ &= 2 + 6 + 3 = 11. \end{aligned}$$

Torniamo ora al calcolo della covarianza fra (e alla varianza di) somme pesate di *V.A.*. Questo calcolo risulta estremamente semplice se si considera che la covarianza è una funzione (operatore) *bilineare*. Cosa significa bilineare? La spiegazione è molto semplice.

Innanzi tutto osserviamo che la funzione di covarianza ha due argomenti: quando calcoliamo la covarianza lo facciamo fra due *V.A.*

$$Cov\left(\underset{1^\circ \text{ argomento}}{X}, \underset{2^\circ \text{ argomento}}{Y}\right) \ .$$

Una funzione a due variabili $f(x, y)$ è detta bilineare se, fissato il secondo (primo) argomento risulta essere lineare nel primo (secondo). In pratica deve valere che

1. Per qualsiasi $x_1, x_2 \in \mathbb{R}$ e $y \in \mathbb{R}$ vale che

$$f(x_1 + x_2, y) = f(x_1, y) + f(x_2, y), \tag{3.13}$$

   Per qualsiasi $\lambda, x \in \mathbb{R}$ e $y \in \mathbb{R}$ vale che

$$f(\lambda x, y) = \lambda f(x, y). \tag{3.14}$$

2. Per qualsiasi $y_1, y_2 \in \mathbb{R}$ e $x \in \mathbb{R}$ vale che

$$f(x, y_1 + y_2) = f(x, y_1) + f(x, y_2), \tag{3.15}$$

Per qualsiasi $\lambda, y \in \mathbb{R}$ e $x \in \mathbb{R}$ vale che

$$f(x, \lambda y) = \lambda f(x, y). \tag{3.16}$$

Esempio 25. La funzione $f(x, y) = xy$ è bilineare. La dimostrazione è lasciata come esercizio.

La covarianza è un operatore bilineare. Questo significa che essa si comporta in maniera simile al valore atteso quando noi "fissiamo" uno dei due argomenti.

Ad esempio, supponiamo di voler calcolare $Cov\,(2X_1 - 3X_2 + X_3, Y_1 - 2Y_2)$. Partiamo considerando "fisso" il secondo argomento che in questo caso è dato da $Y_1 - 2Y_2$. Per semplicità chiamiamo $Z = Y_1 - 2Y_2$ l'argomento fisso.

$$\begin{aligned} Cov\,(2X_1 - 3X_2 + X_3, Z) &= Cov\,(2X_1, Z) + Cov\,(-3X_2, Z) + Cov\,(X_3, Z) \\ &= 2Cov\,(X_1, Z) - 3Cov\,(X_2, Z) + Cov\,(X_3, Z) \end{aligned}$$

Grazie alla bilinearità della covarianza possiamo ora sviluppare ciascuno dei tre termini rispetto al secondo argomento, Z, che riscriviamo nuovamente in funzione di Y_1 e Y_2:

$$\begin{aligned} Cov\,(X_1, Y_1 - 2Y_2) &= Cov\,(X_1, Y_1) + Cov\,(X_1, -2Y_2) \\ &= Cov\,(X_1, Y_1) - 2Cov\,(X_1, Y_2)\ , \end{aligned}$$

$$\begin{aligned} Cov\,(X_2, Y_1 - 2Y_2) &= Cov\,(X_2, Y_1) + Cov\,(X_2, -2Y_2) \\ &= Cov\,(X_2, Y_1) - 2Cov\,(X_2, Y_2)\ , \end{aligned}$$

$$\begin{aligned} Cov\,(X_3, Y_1 - 2Y_2) &= Cov\,(X_3, Y_1) + Cov\,(X_3, -2Y_2) \\ &= Cov\,(X_3, Y_1) - 2Cov\,(X_3, Y_2) \end{aligned}$$

Quale risultato finale otteniamo la seguente espressione

$$\begin{aligned} Cov\,(2X_1 - 3X_2 + X_3, Y_1 - 2Y_2) &= 2Cov\,(X_1, Y_1) - 4Cov\,(X_1, Y_2) \\ &\quad -3Cov\,(X_2, Y_1) + 6Cov\,(X_2, Y_2) \\ &\quad +Cov\,(X_3, Y_1) - 2Cov\,(X_3, Y_2) \\ &= 2c_{11} - 4c_{12} - 3c_{21} + 6c_{22} + c_{31} - 2c_{32}. \end{aligned}$$

Come potete vedere il calcolo della covarianza fra somme di variabili aleatorie è laborioso ma fondamentalmente semplice una volta capite le proprietà della covarianza.

È possibile aiutarsi utilizzando una tabella che ricorda la battaglia navale. La tabella consiste in n righe e m colonne, dove n è il numero di $V.A.$ che figurano nella sommatoria della X (primo argomento) e m in numero di quelle della Y (secondo argomento). All'esempio precedente possiamo associare una tabella 3×2 come questa:

	Y_1	Y_2
X_1	c_{11}	c_{12}
X_2	c_{21}	c_{22}
X_3	c_{31}	c_{32}

Questa tabella è anche chiamata tabella (o matrice) delle covarianze fra X_1, X_2, X_3 e Y_1, Y_2. Per il calcolo della covarianza aggiungiamo dapprima i coefficienti della somma pesata (combinazione lineare) davanti ad ogni variabile

	$1Y_1$	$-2Y_2$
$2X_1$	c_{11}	c_{12}
$-3X_2$	c_{21}	c_{22}
$1X_3$	c_{31}	c_{32}

ed in seguito all'interno della tabella:

	$1Y_1$	$-2Y_2$
$2X_1$	$2 \cdot 1 \cdot c_{11}$	$2 \cdot (-2) \cdot c_{12}$
$-3X_2$	$(-3) \cdot 1 \cdot c_{21}$	$(-3) \cdot (-2) \cdot c_{22}$
$1X_3$	$1 \cdot 1 \cdot c_{31}$	$1 \cdot (-2) \cdot c_{32}$

Per terminare sommiamo tutti i termini della tabella ottenendo così la covarianza fra le due somme di $V.A.$:

$$Cov\,(2X_1 - 3X_2 + X_3, Y_1 - 2Y_2) = 2c_{11} - 4c_{12} - 3c_{21} + 6c_{22} + c_{31} - 2c_{32}.$$

Per completezza di informazione diamo la formula generale in termini di sommatoria per il calcolo della covarianza fra due combinazioni lineari qualsiasi di $V.A.$:

$$Cov(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j Cov(X_i, Y_j)\ .$$

3.4.3 Varianza di una somma di variabili aleatorie

La varianza di una somma di variabili aleatorie, ad esempio

$$V(2X_1 - 3X_2 + X_3) \ ,$$

si calcola in maniera del tutto simile alla covarianza fra somme di $V.A.$. È infatti immediato verificare che per una qualsiasi $V.A.$ X

$$V(X) = E\left[(X - E(X))^2\right] = E\left[(X - E(X))(X - E(X))\right] = Cov(X, X).$$

La varianza di una somma pesata di $V.A.$ altro non è che la somma debitamente pesata della varianze di e delle covarianze fra le $V.A.$ che costituiscono la trasformazione lineare.

Esempio 26. Varianza di $2X_1 - 3X_2 + X_3$
Costruiamo la tabella delle covarianze fra X_1, X_2, X_3 e ... X_1, X_2, X_3:

	X_1	X_2	X_3
X_1	c_{11}	c_{12}	c_{13}
X_2	c_{21}	c_{22}	c_{23}
X_3	c_{31}	c_{32}	c_{33}

Notiamo dapprima che la tabella delle covarianze ha in questo caso lo stesso numero di righe e di colonne. Inoltre, sulla diagonale principale (cioè in posizione $(1,1)$, $(2,2)$ e $(3,3)$) abbiamo i termini c_{ii} che corrispondono a $Cov(X_i, X_i) = V(X_i)$. Queste covarianze non sono altro che le varianze degli X_i! Per tale motivo questa tabella è chiamata la tabella (o matrice) delle *varianze-covarianze* delle $V.A.$ X_1, X_2, X_3.

Fuori dalla diagonale troviamo le covarianze fra X_i e X_j, con $i \neq j$. Poiché (verificatelo dalla definizione e dalla proprietà di commutatività della moltiplicazione) $Cov(X_i, X_j) = Cov(X_j, X_i)$ la tabella delle varianze-covarianze è *simmetrica*, cioè vale

$$c_{ij} = c_{ji}.$$

Per il calcolo della varianza procediamo come per il calcolo della covarianza fra somme di $V.A.$, andando ad aggiungere alla tabella (matrice) delle varianze-covarianze i coefficienti della trasformazione lineare (somma pesata)

	$2X_1$	$-3X_2$	$1X_3$
$2X_1$	$2^2 \cdot c_{11}$	$2 \cdot (-3) \cdot c_{12}$	$2 \cdot 1 \cdot c_{13}$
$-3X_2$	$(-3) \cdot 2 \cdot c_{21}$	$(-3)^2 \cdot c_{22}$	$(-3) \cdot 1 \cdot c_{23}$
$1X_3$	$1 \cdot 2 \cdot c_{31}$	$1 \cdot (-3) \cdot c_{32}$	$1^2 \cdot c_{33}$

Le varianze sono moltiplicate per il quadrato del coefficiente della trasformazione lineare. Fuori dalla diagonale troviamo le covarianze fra le rispettive $V.A.$ moltiplicate per i pesi corrispondenti. La varianza di questa trasformazione lineare è uguale alla somma di tutti gli elementi della tabella, ovvero

$$\begin{aligned} V(2X_1 - 3X_2 + X_3) &= 4c_{11} + 9c_{22} + c_{33} + \\ &\quad -6c_{12} + 2c_{13} + \\ &\quad -6c_{21} - 3c_{23} + \\ &\quad +2c_{13} - 3c_{32} \end{aligned} \tag{3.17}$$

Infine, grazie alla simmetria della matrice delle varianze-covarianze ($c_{ij} = c_{ji}$) la (3.17) può essere riscritta come

$$\begin{aligned} V(2X_1 - 3X_2 + X_3) &= 4c_{11} + 9c_{22} + c_{33} + \\ &\quad -12c_{12} + 4c_{13} - 6c_{23} \end{aligned} \tag{3.18}$$

In generale, data una qualsiasi trasformazione lineare delle n $V.A.$ $X_1, \ldots, X_n$ la varianza è calcolata tramite la seguente formula

$$\begin{aligned} V\left(\sum_{i=1}^{n} a_i X_i\right) &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j Cov(X_i, X_j) \\ &= \sum_{i=1}^{n} a_i^2 V(X_i) + \sum_{\substack{i=1 \\ i\neq j}}^{n}\sum_{j=1}^{n} a_i a_j Cov(X_i, X_j) \\ &= \sum_{i=1}^{n} a_i^2 V(X_i) + 2\sum_{\substack{i=1 \\ i>j}}^{n}\sum_{j=1}^{n} a_i a_j Cov(X_i, X_j)\ . \end{aligned} \tag{3.19}$$

Caso particolare: variabili non correlate. Quando le $V.A.$ $X_1, \ldots, X_n$ sono non correlate, cioè

$$Cov(X_i, X_j) = 0 \text{ per } i \neq j$$

la formula (3.19) per il calcolo della varianza di una combinazione lineare di $V.A.$ si semplifica notevolmente in

$$V\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 V\left(X_i\right). \tag{3.20}$$

In pratica, nel caso di variabili non correlate, la varianza di una somma è la somma delle varianze pesate per il *quadrato* dei coefficienti della combinazione lineare. Ricordate che la varianza è il valore atteso "di un quadrato": per tale motivo i coefficienti sono al quadrato!

3.5 Il campionamento casuale semplice

Consideriamo una popolazione di numerosità N.

Definizione 18. Chiameremo campione casuale semplice un campione di numerosità n dove ciascuna unità possiede, *ad ogni passo*, una probabilità pari a $1/N$ di essere estratta.

Esercizio 5. Mostrate che entrambe le tecniche di campionamento con e senza reinserimento soddisfano la condizione imposta dalla definizione di campione casuale semplice. Suggerimento: per quanto riguarda la selezione senza reinserimento, aiutatevi utilizzando la proprietà condizionata.

È dunque possibile ottenere un campione casuale semplice applicando una delle due tecniche di selezione casuale con o senza reinserimento viste in precedenza.

> Il campionamento casuale semplice è raramente applicato nelle indagini statistiche, sia perché la selezione è completamente affidata al caso e non incorpora le informazione note a priori sulla popolazione o sulle caratteristiche distributive delle variabili, sia perché nelle indagini su vasta scala è pesante quanto a costi di relevazione dei dati e a organizzazione del lavoro "sul campo" ...
> Il campionamento casuale semplice è invece quello che si assume nella teoria dell'inferenza statistica quando non è precisato il disegno adottato. Per questo tipo di campionamento sono, infatti proposte *metodiche* di *stima* e di *verifica* della significatività statistica di ipotesi sui più disparati parametri delle distribuzioni, anche multivariate, di indici di relazione tra variabili, di dati organizzati in serie temporali ecc. (Fabbris, pp. 53-54).

Il campionamento casuale semplice è lo standard verso il quale confrontare gli altri tipi di campionamento. Per tale motivo sarà quello da noi maggiormente trattato.

3.5.1 Valore atteso e varianza della media campionaria

In questo e nei prossimi paragrafi desideriamo studiare le proprietà di correttezza e di precisione della media campionaria $\overline{X}_n$ quale stimatore del valore medio μ di una particolare caratteristica numerica X della popolazione (reddito da lavoro, patrimonio, spesa in beni alimentari, ecc.). In questo caso scalziamo dal trono la distribuzione statistica P_p della popolazione obiettivo quale principale oggetto di studio. Con modestia ci accontentiamo di stimare il valore medio μ della caratteristica X della popolazione

$$\mu = \sum_{k=1}^{K} x_k p_k$$

dove K rappresenta il numero dei diversi valori osservabili della caratteristica in esame e p_k rappresenta la frazione di unità della popolazione (la frequenza relativa dunque) aventi un valore della caratteristica uguale a x_k. Sappiamo che la popolazione è costituita da N unità. Denotiamo con a_j il valore della caratteristica della j-esima unità. Ad esempio, se decidessimo di studiare la distribuzione del reddito della popolazione svizzera, a_j rappresenterebbe il reddito dello j-esimo individuo. Un possibile parametro sconosciuto al quale potremmo essere interessati potrebbe essere il valore medio della popolazione

$$\mu = \sum_{k=1}^{K} x_k p_k = \sum_{j=1}^{N} a_j \frac{1}{N}$$

oppure la varianza della distribuzione del reddito

$$\sigma^2 = \sum_{k=1}^{K} (x_k - \mu)^2 p_k = \sum_{j=1}^{N} (a_j - \mu)^2 \frac{1}{N} .$$

μ e σ^2 sono due quantità che potrebbero essere calcolate se avessimo a disposizione i dati sull'intera popolazione. Assumiamo di non essere interessati all'intera distribuzione della popolazione ma unicamente al primo momento ed alla varianza: per questa volta eviteremo di specificare una famiglia parametrica di distribuzioni.

Definizione 19. Frazione di campionamento. Chiameremo la quantità

$$f := \frac{n}{N} \tag{3.21}$$

frazione di campionamento. Essa indica semplicemente la % di unità estratte rispetto al totale della popolazione.

Ci interessiamo ora alla media del campione quale stimatore di μ.

3.5.1.1 Campionamento con reinserimento

Supponiamo di estrarre un campione di numerosità n utilizzando la tecnica di selezione con reinserimento. Indichiamo con $X_1, X_2, \dots X_n$ le n variabili aleatorie indipendenti ed identicamente distribuite che descrivono il risultato della prima, seconda, ... n-esima estrazione. È immediato verificare[6] che per un qualsiasi $i \in (1, \ldots, n)$

$$E(X_i) = \sum_{j=1}^{N} a_j \frac{1}{N} = \mu$$

e

$$V(X_i) = E\left((X_i - \mu)^2\right) = \sum_{j=1}^{N} (a_j - \mu)^2 \frac{1}{N} = \sigma^2.$$

Per quanto riguarda la media campionaria $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ avremo quindi

$$\begin{aligned} E(\overline{X}_n) &= E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E\left(X_i\right) = \mu \\ V(\overline{X}_n) &= V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} V\left(\sum_{i=1}^{n} X_i\right) \\ &= \frac{1}{n^2}\sum_{i=1}^{n} V\left(X_i\right) = \frac{\sigma^2}{n}. \end{aligned} \tag{3.22}$$

Conclusioni: la media campionaria è uno stimatore corretto del valore atteso della popolazione. La sua precisione, valutata utilizzando la varianza quale misura di dispersione, dipende da due fattori.

1. La numerosità del campione: la varianza della media $\overline{X}_n$ decresce in maniera inversamente proporzionale al numero di osservazioni nel campione.

2. La varianza della popolazione: il recercatore non ha alcun influsso su questo parametro.

[6] Potete verificare applicando le proprietà della sommatoria che $V(X_i)$ è anche uguale a

$$\sum_{j=1}^{N} a_j^2 \frac{1}{N} - \mu^2$$

che corrisponde alla formula alternativa (3.6) della varianza data da

$$V(X_i) = E(X_i^2) - \left(E(X_i)\right)^2.$$

Da ultimo sottolineamo nuovamente che $\overline{X}_n$ è una variabile aleatoria la cui distribuzione dipende dalla distribuzione della popolazione campione. Purtroppo vale la seguente osservazione. Anche qualora la distribuzione della popolazione obiettivo fosse conosciuta, la distribuzione della media $\overline{X}_n$ non sarebbe, in generale, derivabile analiticamente. Ci sono però delle eccezioni. Una fra queste è la distribuzione Normale. Infatti, quando le variabili aleatorie $X_1, ..., X_n$ sono *i.i.d.* e distribuite secondo la legge Normale $N(\mu, \sigma^2)$, allora anche la loro media è una variabile aleatoria Normale.

Esercizio 6. Utilizzando il file di excel sulla distribuzione del numero di scarpe eseguite una simulazione ed una analisi dei risultati.

3.5.1.2 Campionamento senza reinserimento

In questo caso l'estrazione degli n individui dalla popolazione avviene utilizzando la tecnica di selezione senza reinserimento. Da un punto di vista teorico il campionamento senza reinserimento appare giustificato dal fatto che l'osservazione ripetuta dello stesso individuo (unità) non aumenta l'informazione disponibile sull'intera popolazione. Appare quindi sensato togliere dall'urna gli individui già osservati e procedere solo con la parte restante della popolazione. Vedremo che tale intuizione è corretta: la media di un campione estratto con la tecnica di selezione senza reinserimento possiede una varianza inferiore alla media di un campione estratto con reinserimento[7]. Come in precedenza abbiamo che

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i. \tag{3.23}$$

Le variabili X_i sono ora definite nel seguente modo:

$$X_i = \sum_{j=1}^{N} a_j U_{ij} \tag{3.24}$$

dove la variabile aleatoria

$$U_{ij} = \begin{cases} 1 & \text{se lo } j\text{-esimo individuo è estratto alla } i\text{-esima estrazione,} \\ 0 & \text{altrimenti.} \end{cases}$$

Esercizio 7. Consideriamo una popolazione obiettivo di $N = 5$ unità ed un numero $n = 3$ di estrazioni senza reinserimento. Lo spazio di probabilità è

[7] Il prezzo da pagare consiste in una maggiore difficoltà nel calcolo di $V(\overline{X}_n)$.

quello studiato nella Sezione 3.2

$$\left(\Omega^3, \bigotimes_{i=1}^{3} \mathcal{E}_i, Q\right).$$

Prendiamo le variabili aleatorie U_{13}, U_{21} e U_{35}.

1. Definite a parole le variabili aleatorie X_1, X_2 e X_3.
2. Definite a parole le tre variabili aleatorie U_{13}, U_{21} e U_{35}.
3. Considerate gli esiti $\omega_a = (3, 2, 5)$, $\omega_b = (4, 2, 1)$ e calcolate per ciascuna delle tre variabili U il valore che esse assumono in ω_i, $i = a, b$.
4. Mostrate che in ω_a vale $X_1 = a_1U_{11} + a_2U_{12} + a_3U_{13} + a_5U_{15}$.
5. Mostrate che in ω_b non vale $X_1 = a_1U_{11} + a_2U_{12} + a_3U_{13} + a_5U_{15}$.

Possiamo riassumere il tutto nella seguente tabella

$$\begin{array}{ccccccccc}
 & & \text{Indiv. 1} & & \text{Indiv. 2} & & & & \text{Indiv. } N \\
X_1 & = & a_1U_{11} & + & a_2U_{12} & + & \dots & + & a_NU_{1N} \\
X_2 & = & a_1U_{21} & + & a_2U_{22} & + & \dots & + & a_NU_{2N} \\
\vdots & & \vdots & & \vdots & & & & \vdots \\
X_n & = & a_1U_{n1} & + & a_2U_{n2} & + & \dots & + & a_NU_{nN} \\
\hline
\sum_{i=1}^n X_i & = & \underbrace{a_1\sum_{i=1}^n U_{i1}}_{T_1} & + & \underbrace{a_2\sum_{i=1}^n U_{i2}}_{T_2} & + & \dots & + & \underbrace{a_N\sum_{i=1}^n U_{iN}}_{T_N}
\end{array}$$

Questa tabella possiede più colonne che righe: $n \leq N$ (non posso estrare più di N individui). Le N variabili aleatorie T_j sono la somma rispetto al numero di estrazioni (indice i) delle variabili U_{ij}. In pratica T_j assumerà il valore 1 o 0 a seconda che lo j-esimo individuo sia stato sorteggiato o meno.
Inserendo la relazione (3.24) nella (3.23) possiamo riscrivere la media come una somma pesata delle $V.A.$ T_j:

$$\begin{aligned}
\overline{X}_n &= \frac{1}{n}\sum_{i=1}^n X_i = \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^N a_jU_{ij} = \frac{1}{n}\sum_{j=1}^N\sum_{i=1}^n a_jU_{ij} \\
&= \frac{1}{n}\sum_{j=1}^N a_j\sum_{i=1}^n U_{ij} = \frac{1}{n}\sum_{j=1}^N a_jT_j. \qquad (3.25)
\end{aligned}$$

A questo punto è necessario calcolare valore atteso, varianza di - e covarianza fra - variabili aleatorie T_j.
Per quanto riguarda il valore atteso di T_j abbiamo

$$E(T_j) = 0\ P(T_j = 0) + 1\ P(T_j = 1) = P(T_j = 1)$$

$P(T_j = 1)$ è la probabilità che sull'arco delle n estrazioni lo j-esimo individuo venga sorteggiato. Poiché gli esiti con probabilità non nulla sono tutti equiprobabili, è possibile calcolare $P(T_j = 1)$ semplicemente come

$$P(T_j = 1) = \frac{\#\text{ esiti favorevoli}}{\#\text{ esiti possibili}}.$$

Il numero di esiti possibili è già stato calcolato nella Sezione 3.2 ed è

$$D_N^n = \frac{N!}{(N-n)!}.$$

Per l'individuo j, il numero degli esiti favorevoli è uguale al numero di esiti ω che hanno probabilità positiva[8] e che contengono il numero j. Tale numero è la risposta alla seguente domanda: in quanti modo posso disporre gli N individui $1, ..., N$ in n scatole, sapendo che una scatola sarà occupata dall'individuo j? Abbiamo n possibilità per disporre il numero j nelle n scatole e successivamente resteranno $n-1$ scatole libere che dovranno essere riempite scegliendo da $N-1$ elementi. Il risultato è dunque

$$n\, D_{N-1}^{n-1} = \frac{n\ (N-1)!}{(N-1-(n-1))!} = \frac{n\ (N-1)!}{(N-n)!}.$$

Otteniamo così

$$P(T_j = 1) = \frac{\frac{n\ (N-1)!}{(N-1-(n-1))!}}{\frac{N!}{(N-n)!}} = \frac{n}{N}. \tag{3.26}$$

Per quanto riguarda la varianza di T_j, essa è uguale a

$$V(T_j) = E(T_j^2) - (E(T_j))^2 = \frac{n}{N} - \left(\frac{n}{N}\right)^2 = \frac{n(N-n)}{N^2}.$$

La covarianza fra T_j e T_k è invece uguale a

$$\begin{aligned} Cov(T_j, T_k) &= E(T_jT_k) - E(T_j)E(T_k) \\ &= E(T_jT_k) - \left(\frac{n}{N}\right)^2. \end{aligned} \tag{3.27}$$

[8]Ricordiamo che gli elementi ω di Ω^n in cui un numero appare due o più volte non vanno tenuti in considerazione in quanto non si realizzeranno mai (questa è un'estrazione senza reinserimento!).

Quanto vale $E(T_jT_k)$? Il prodotto T_jT_k è sempre uguale a 0 salvo quando entrambi gli individui j e k figurano nel campione. In tal caso $T_jT_k = 1$. Quindi

$$\begin{aligned} E(T_jT_k) &= 0P(T_j = 0, T_k = 0) + 0P(T_j = 1, T_k = 0) \\ &+ 0P(T_j = 0, T_k = 1) + 1P(T_j = 1, T_k = 1) \\ &= P(T_j = 1, T_k = 1) \end{aligned}$$

Per calcolare $P(T_j = 1, T_k = 1)$ utilizziamo nuovamente la combinatoria:

1. Esiti favorevoli

   Degli n posti 2 sono occupati: uno dall' individuo j e l'altro dall'individuo k. Abbiamo $n(n-1)$ modi per disporre i nostri due individui nelle n scatole. Rimangono $n-2$ posti liberi che possono essere riempiti dagli altri $N-2$ individui. Abbiamo quindi

   $$\#\text{ esiti favorevoli} = n(n-1)D_{N-2}^{n-2} \ .$$

2. Esiti possibili (calcolati in precedenza) uguali a $\frac{N!}{(N-n)!}$.

Abbiamo quindi

$$P(T_j = 1, T_k = 1) = \frac{n(n-1)\frac{(N-2)!}{(N-2-(n-2))!}}{\frac{N!}{(N-n)!}} = \frac{n(n-1)}{N(N-1)}.$$

Inserendo quest'ultimo risultato nella (3.27) otteniamo (si noti il segno dell'espressione finale)

$$Cov(T_j, T_k) = \frac{n(n-1)}{N(N-1)} - \left(\frac{n}{N}\right)^2 = -\frac{n(N-n)}{N^2(N-1)}.$$

Siamo ora pronti a calcolare il valore atteso e la varianza di $\overline{X}_n$.

Il valore atteso di $\overline{X}_n$

$$\begin{aligned} E(\overline{X}_n) &= E\left(\frac{1}{n}\sum_{j=1}^{N} a_jT_j\right) = \frac{1}{n}\sum_{j=1}^{N} a_jE\left(T_j\right) \\ &= \frac{1}{n}\sum_{j=1}^{N} a_j\frac{n}{N} = \frac{1}{N}\sum_{j=1}^{N} a_j = \mu. \end{aligned}$$

Quando il campionamento è effettuato senza reinserimento la media campionaria è uno stimatore corretto del valore atteso della popolazione. Questo risultato è identico a quello ottenuto nel caso di campionamento con reinserimento. Sarà dunque interessante confrontare le varianze della media campionaria nei due casi per vedere quale delle due tecniche di campionamento fornisce i risultati migliori in termini di precisione.

La varianza di $\overline{X}_n$ Dall'equazione (3.25) otteniamo che

$$\begin{aligned} V(\overline{X}_n) &= V(\frac{1}{n}\sum_{j=1}^{N} a_j T_j) = \frac{1}{n^2} V(\sum_{j=1}^{N} a_j T_j) \\ &= \frac{1}{n^2}\sum_{j=1}^{N}\sum_{s=1}^{N} a_j a_s Cov(T_j, T_s) \\ &= \frac{1}{n^2}\left(\sum_{j=1}^{N} a_j^2 V(T_j) + \sum_{\substack{j=1 \\ j\neq s}}^{N}\sum_{s=1}^{N} a_j a_s Cov(T_j, T_s)\right). \end{aligned}$$

Utilizzando i risultati sulla varianza e covarianza delle $V.A.$ T_j

$$\begin{aligned} V(\overline{X}_n) &= \frac{1}{n^2}\left(\frac{n(N-n)}{N^2}\sum_{j=1}^{N} a_j^2 - \frac{n(N-n)}{N^2(N-1)}\sum_{\substack{j=1 \\ j\neq s}}^{N}\sum_{s=1}^{N} a_j a_s\right) \\ &= \frac{1}{n^2}\frac{n(N-n)}{N^2(N-1)}\left((N-1)\sum_{j=1}^{N} a_j^2 - \sum_{\substack{j=1 \\ j\neq s}}^{N}\sum_{s=1}^{N} a_j a_s\right) \end{aligned}$$

$$\begin{aligned} &= \frac{1}{n}\frac{N-n}{N^2(N-1)}\left(N\sum_{j=1}^{N} a_j^2 - \sum_{j=1}^{N}\sum_{s=1}^{N} a_j a_s\right) \\ &= \frac{1}{n}\frac{N-n}{N^2(N-1)}\left(N\sum_{j=1}^{N} a_j^2 - \left(\sum_{j=1}^{N} a_i\right)\left(\sum_{s=1}^{N} a_s\right)\right) \\ &= \frac{1}{n}\frac{N-n}{N-1}\left(\frac{1}{N}\sum_{j=1}^{N} a_j^2 - \mu^2\right) \end{aligned} \tag{3.28}$$

Ma $\sum_{j=1}^{N} a_j^2 \frac{1}{N} - \mu^2$ è la formula alternativa della varianza della popolazione (confronta la Sezione 3.5.1.1). Per tale motivo otteniamo quale risultato finale

$$V(\overline{X}_n) = \frac{N-n}{N-1}\frac{\sigma^2}{n}. \tag{3.29}$$

Confrontando le due formule (3.22) e (3.29) ottenute per la varianza dello stimatore $\overline{X}_n$ notiamo che il campionamento senza reinserimento è più efficiente. Infatti la quantità $\frac{N-n}{N-1}$ è sempre minore di uno salvo nel caso particolare (in pratica non rilevante) in cui $n = 1$. Questo risultato conferma la nostra intuizione riguardo all'inutilità di reinserire l'unità osservata al fine di acquisire informazioni sulla popolazione. In questo caso la tecnica di reinserimento si rivela persino dannosa in termini della varianza di $\overline{X}_n$.

3.5.1.3 Campionamento sistematico

Supponiamo che il campione di numerosità n sia stato scelto utilizzando la tecnica di selezione sistematica (si veda il paragrafo 3.3) e per semplicità che $k = \frac{N}{n}$ sia un numero intero. Ricordiamo che k corrisponde al numero di gruppi in cui l'intera popolazione è partizionata. Ciascun gruppo conterrà dunque n unità della popolazione. La probabilità di includere un particolare individuo nel campione è uguale alla probabilità di selezionare il suo gruppo d'appartenenza: poiché in totale ci sono k gruppi ed ogni individuo appartiene ad esattamente un solo gruppo tale probabilità è $\frac{1}{k}$. Tutti gli individui hanno la medesima probabilità di essere estratti. Tale probabilità è però diversa da $\frac{1}{n}$ a meno che la numerosità N della popolazione non sia tale per cui $N = n^2$. La media campionaria $\overline{X}_n$ è come al solito uguale alla somma delle n variabili aleatorie X_i divisa per n. Tuttavia, mentre nei casi della tecnica di selezione con e senza reinserimento avevamo potuto definire le variabili aleatorie X_i come il reddito dell'individuo estratto all'i-esima estrazione, ora ci troviamo di fronte ad un'unica estrazione in cui tutti gli n individui sono estratti simultaneamente. In pratica è come se estraessimo in un unico colpo la somma $\sum_{i=1}^{n} X_i$ che, nel caso del campionamento sistematico, corrisponde alla somma della caratteristica degli n individui appartenenti al gruppo selezionato. Per tale motivo in questa Sezione utilizzeremo la seguente convenzione: l'indice j denoterà il gruppo mentre l'indice i verrà utilizzato per indicare l'individuo all'interno del gruppo. Definiamo per ciascun gruppo $j \in (1, ..., k)$ la quantità

$$g_j := \sum_{i=1}^{n} a_{ji}$$

dove a_{ji} corrisponde per definizione alla caratteristica della i-esima persona dello j-esimo gruppo. Le quantità g_j non sono aleatorie ma *sono dei numeri* in quanto il partizionamento della popolazione in k gruppi è avvenuto in maniera deterministica. La casualità viene introdotta definendo le variabili aleatorie T_j nel modo seguente

$$T_j = \begin{cases} 1 & \text{lo } j\text{-esimo gruppo è selezionato} \\ 0 & \text{altrimenti} \end{cases}.$$

Come calcolato in precedenza è facile dimostrare che

$$\begin{aligned} E(T_j) &= \frac{1}{k}, \\ V(T_j) &= \left(\frac{1}{k} - \frac{1}{k^2}\right), \\ Cov(T_j, T_s) &= -\frac{1}{k^2}. \end{aligned}$$

A questo punto è semplice convincersi che la media campionaria di un campionamento sistematico altro non è che

$$\overline{X}_n = \frac{1}{n} \sum_{j=1}^{k} g_j T_j.$$

Il valore atteso di $\overline{X}_n$ Utilizziamo le proprietà ormai note del valore atteso:

$$\begin{aligned} E\left(\overline{X}_n\right) &= E\left(\frac{1}{n} \sum_{j=1}^{k} g_j T_j\right) = \frac{1}{n} \sum_{j=1}^{k} g_j E\left(T_j\right) = \frac{1}{n} \frac{1}{k} \sum_{j=1}^{k} g_j \\ &= \frac{1}{N} \underbrace{\sum_{j=1}^{k} \sum_{i=1}^{n} a_{ji}}_{\text{somma sull'intera pop.}} = \mu. \end{aligned}$$

La media campionaria è dunque uno stimatore corretto del valore atteso della popolazione anche quando la tecnica di campionamento è quella del campionamento sistematico. Veniamo ora alla calcolo del suo grado di precisione.

La varianza di $\overline{X}_n$

$$\begin{aligned}
V(\overline{X}_n) &= \frac{1}{n^2}V(\sum_{j=1}^{k} g_j T_j) = \frac{1}{n^2}\sum_{j=1}^{k}\sum_{s=1}^{k} g_j g_s Cov(T_j, T_s) \\
&= \frac{1}{n^2}\left(\sum_{j=1}^{k} g_j^2 V(T_j) + \sum_{\substack{j=1 \\ j\neq s}}^{k}\sum_{s=1}^{k} g_j g_s Cov(T_j, T_s)\right) \\
&= \frac{1}{n^2}\left(\left(\frac{1}{k}-\frac{1}{k^2}\right)\sum_{j=1}^{k} g_j^2 - \frac{1}{k^2}\sum_{\substack{j=1 \\ j\neq s}}^{k}\sum_{s=1}^{k} g_j g_s\right) \\
&= \frac{1}{n^2}\left(\frac{1}{k}\sum_{j=1}^{k} g_j^2 - \frac{1}{k^2}\sum_{j=1}^{k}\sum_{s=1}^{k} g_j g_s\right).
\end{aligned}$$

Continuando esattamente come visto nella (3.28) si ottiene

$$V(\overline{X}_n) = \frac{1}{n^2}\frac{1}{k}\sum_{j=1}^{k}(g_j - \overline{g})^2 .$$

Interpretazione: la varianza di $\overline{X}_n$ dipenderà quindi dall'eterogeneità dei k gruppi rispetto alla caratteristica in esame. Se ad esempio g_1 contiene prevalentemente redditi bassi, g_2 prevalentemente redditi medi e g_3 prevalentemente redditi alti avremo una varianza elevata. Se invece g_1, g_2 e g_3 contengono tutti redditi bassi/medi/alti avremo che $g_1 \approx g_2 \approx g_3$ e la varianza dello stimatore sarà quindi più bassa.

3.6 Campionamento stratificato

Stratificare una popolazione consiste nel creare una partizione della stessa sulla base di determinati criteri. Ogni sottoinsieme è detto *strato*. I criteri utilizzati per definire gli strati dipendono dalle caratteristiche della popolazione nonché dall'obiettivo dello studio. Possibili criteri sono il sesso, l'età, il reddito, l'appartenenza ad una regione linguistica, la dimensione dell'azienda in termini di numero di dipendenti o cifra d'affari, il settore in cui opera l'azienda, ecc. Ovviamente gli strati non avranno quasi mai lo stesso numero di unità. Fabbris (p. 71) identifica i seguenti obiettivi che possono indurre ad effettuare una stratificazione della popolazione.

1. Evidenziare insiemi di unità significative per la ricerca.
2. Separare dalle altre le sottopopolazioni fisicamente isolate e con caratteristiche speciali.
3. Individuare certe unità che si vogliono osservare con tecniche particolari.
4. Introdurre sulla selezione il massimo controllo, pur mantenendola casuale.
5. Individuare sottopopolazioni al massimo omogenee rispetto alla variabile o alle variabili da rilevare e ricavare così stime più efficienti di quelle ottenibili con un campione casuale semplice.

Supponiamo per un istante di osservare l'intera popolazione. Potremmo suddividere gli uomini dalle donne e calcolare per ciascun strato i rispettivi valori attesi[9] μ_d e μ_u. Qual è la relazione tra il valore atteso μ dell'intera popolazione e quello dei due strati? Indichiamo con N_d e N_u il numero di donne e uomini contenuti nelle rispettive sottopopolazioni. Ovviamente varrà che $N = N_d + N_u$. Avremo dunque

$$
\begin{aligned}
\mu &= \frac{1}{N}\left(\sum_{i=1}^{N} a_i\right) = \frac{1}{N}\left(\sum_{i=1}^{N_d} a_{d,i} + \sum_{j=1}^{N_u} a_{u,j}\right) \\
&= \frac{1}{N}\sum_{i=1}^{N_d} a_{d,i} + \frac{1}{N}\sum_{j=1}^{N_u} a_{u,j} = \frac{N_d}{N}\frac{1}{N_d}\sum_{i=1}^{N_d} a_{d,i} + \frac{N_u}{N}\frac{1}{N_u}\sum_{j=1}^{N_u} a_{u,j} \\
&= \frac{N_d}{N}\mu_d + \frac{N_u}{N}\mu_u = \pi_d\mu_d + \pi_u\mu_u.
\end{aligned}
$$

Il reddito medio della popolazione è dunque la *somma pesata* dei redditi medi dei due strati. Generalizzando al caso con H strati, varrà la seguente formula

$$\mu = \sum_{h=1}^{H} \pi_h \mu_h \tag{3.30}$$

dove

- π_h corrisponde al *peso* assegnato all'h-esimo strato e sarà uguale alla frazione di individui (unità) appartenenti allo strato h rispetto al numero totale di individui (unità) della popolazione

$$\pi_h = \frac{N_h}{N}. \tag{3.31}$$

[9] Per la precisione le due quantità μ_d e μ_u corrispondono ai valori attesi condizionati rispetto all'attributo uomo/donna.

- μ_h è il valore medio dell' h-esimo strato, ovvero

$$\mu_h = \frac{1}{N_h} \sum_{j=1}^{N_h} a_{h,j}.$$

L'idea fondamentale del campionamento stratificato è quella di sfruttare l'informazione disponibile e riguardante

- i pesi π_h sulla struttura della popolazione;
- l'omogeneità degli strati al loro interno

per migliorare la qualità della stima. Anziché stimare direttamente il valore atteso dell'intera popolazione si utilizzerà un approccio indiretto costituito da due fasi.

- Fase 1: si stimano i valori attesi μ_h delle sottopopolazioni (strati) selezionando da ciascuna sottopopolazione un campione casuale di numerosità n_h.
- Fase 2: utilizzando le stime dei singoli μ_h e grazie alla conoscenza dei pesi π_h si stima il valore atteso μ della popolazione con la formula (3.30), sostituendo ai valori attesi sconosciuti μ_h le rispettive medie campionarie, notate $\overline{X}_{n_h}$:

$$\overline{X}_{str,n} = \sum_{h=1}^{H} \pi_h \overline{X}_{n_h}.$$

Esempio 27. In Svizzera il numero di persone occupate (dati del secondo trimestre 2007, in milioni) sono 4,369 di cui 2,415 uomini e 1,954 donne. Avremo quindi $\pi_u = \frac{2{,}415}{4{,}369}$ e $\pi_d = \frac{1{,}954}{4{,}369}$. Supponiamo che $n_u = 200$ e $n_d = 100$. Sulla base dei due campioni abbiamo calcolato $\overline{x}_{n_u} = 60,000$ e rispettivamente, $\overline{x}_{n_d} = 50,000$. La stima del reddito medio dei lavoratori svizzeri sarebbe quindi

$$\begin{aligned} \overline{x}_{str,300} &= \frac{2,415}{4,369}\, 60,000 + \frac{1,954}{4,369}\, 50,000 \\ &= 33,165 + 22,362 = 55,527. \end{aligned}$$

La numerosità del campione dell' h-esimo strato è indicato con n_h. I sottocampioni potranno essere di numerosità diversa. La numerosità del campione è semplicemente la somma delle H numerosità dei sottocampioni

$$n = n_1 + n_2 + ... + n_H = \sum_{h=1}^{H} n_h.$$

Il campione è l'unione degli H sottocampioni.

Definizione 20. Frazione di campionamento. La quantità

$$f_h = \frac{n_h}{N_h} \tag{3.32}$$

è chiamata frazione di campionamento dell'h-esimo strato.

Quando si estrae la stessa frazione di unità da ogni strato, o in altre parole quando la frazione di campionamento è uguale per tutti gli strati

$$f_h = c \quad \forall h \tag{3.33}$$

allora il campione si dice *stratificato proporzionale*. Quando invece f_h non è costante si parlerà di campione *stratificato non proporzionale* o *a probabilità variabili*. Si noti che nel caso di campione stratificato proporzionale, c è uguale alla frazione di campionamento f definita dalla (3.21) e pari a $\frac{n}{N}$. Infatti, utilizzando le (3.32) e (3.33) otteniamo

$$\begin{aligned} n_h &= cN_h \quad \forall h \\ \sum_{h=1}^{H} n_h &= c\sum_{h=1}^{H} N_h \\ n &= cN \\ c &= \frac{n}{N}. \end{aligned}$$

Per un campione stratificato proporzionale vale dunque la relazione

$$n_h = \frac{N_h}{N}n \text{ o equivalentemente } \frac{n_h}{n} = \frac{N_h}{N} \ \forall h. \tag{3.34}$$

Fissata la numerosità del campione sulla base di vincoli economici dettati dalla limitatezza delle risorse a disposizione, il numero di unità da estrarre dall'h-esimo strato è dato dalla formula (3.34). Il campione selezionato avrà in questo caso le stesse caratteristiche della popolazione in termini di rappresentatività di ogni strato al suo interno (formula (3.33)).

Osservazione 9. Un campione stratificato proporzionale non è condizione né necessaria né sufficiente per garantire la correttezza dello stimatore $\overline{X}_{str,n}$.

Osservazione 10. Un campione di numerosità n verrà generalmente costruito eseguendo H campionamenti casuali semplici di numerosità n_h ciascuno, $h = 1, ..., H$. Se il campione è stratificato proporzionale, avremo che individui appartenenti a strati diversi avranno la medesima probabilità di figurare nel campione. Infatti, la probabilità di figurare nel campione di un qualsiasi individuo (prendiamo ad esempio il primo) dello strato h è uguale (cf. formula (3.26))

$$P(\text{individuo 1 dello strato } h \text{ è estratto}) = \frac{n_h}{N_h} = c = \frac{n}{N}.$$

Se, al contrario, il campione non è stratificato proporzionale, il rapporto $\frac{n_h}{N_h}$ varierà con h e quindi le probabilità di inclusione dei singoli individui varieranno anch'esse da strato a strato.

Osservazione 11. Se il campione è stratificato proporzionale, la formula per il calcolo di $\overline{X}_{str,n}$ è semplificabile. Infatti

$$\begin{aligned}\overline{X}_{str,n} &= \sum_{h=1}^{H} \pi_h \overline{X}_{n_h} = \sum_{h=1}^{H} \frac{N_h}{N} \overline{X}_{n_h} = \sum_{h=1}^{H} \frac{N_h}{N} \frac{1}{n_h} \sum_{j=1}^{n_h} X_{h,j} \\ &\underset{n_h=\frac{N_h}{N}n}{=} \sum_{h=1}^{H} \frac{N_h}{N} \frac{N}{N_h} \frac{1}{n} \sum_{j=1}^{n_h} X_{h,j} = \frac{1}{n} \underbrace{\sum_{h=1}^{H} \sum_{j=1}^{n_h} X_{h,j}}_{\text{somma sull'intero campione}} .\end{aligned}$$

Tabella 3.1: Tabella riassuntiva

Strato	h	$1,$	...,	$h,$	...,	H	Popolazione
# Unità componenti	(N_h)	$N_1,$	...,	$N_h,$	...,	N_H	N
Peso	$\left(\pi_h = \frac{N_h}{N}\right)$	$\pi_1,$	...,	$\pi_h,$	...,	π_H	1
Varianza	(σ_h^2)	$\sigma_1^2,$	...,	$\sigma_h^2,$	...,	σ_H^2	σ^2
Unità campionarie	(n_h)	$n_1,$	...,	$n_h,$	...,	n_H	n
Frazione di campionamento	(f_h)	$f_1,$	...,	$f_h,$	...,	f_H	f

3.6.1 La correttezza di $\overline{X}_{str,n}$

Se gli stimatori $\overline{X}_{n_h}$ di μ_h sono tutti degli stimatori corretti, $\overline{X}_{str,n}$ sarà automaticamente uno stimatore corretto di μ:

$$E\left(\overline{X}_{str,n}\right) = E\left(\sum_{h=1}^{H} \pi_h \overline{X}_{n_h}\right) = \sum_{h=1}^{H} \pi_h E\left(\overline{X}_{n_h}\right) = \sum_{h=1}^{H} \pi_h \mu_h = \mu.$$

La stima di μ richiede dunque la stima di H valori attesi, uno per strato. Per ciascun strato siamo liberi di scegliere la tecnica di selezione nonché il tipo di stimatore più appropriato alle sue caratteristiche.

3.6.2 La varianza di $\overline{X}_{str,n}$

Poiché i campionamenti eseguiti sui vari strati sono fra loro indipendenti, la varianza di $\overline{X}_{str,n}$ è semplicemente uguale a

$$V\left(\overline{X}_{str,n}\right) = V\left(\sum_{h=1}^{H} \pi_h \overline{X}_{n_h}\right) = \sum_{h=1}^{H} \pi_h^2 V\left(\overline{X}_{n_h}\right)$$

Quest'ultima formula è valida indipendentemente dalle tecniche di campionamento applicate ai vari strati (purché sia mantenuta l'indipendenza fra un campionamento e l'altro). Ammettiamo ora che ad ogni strato venga applicata la tecnica di campionamento senza reimmissione. Abbiamo visto nei paragrafi precedenti che in tal caso la varianza[10] di $\overline{X}_{n_h}$ (cf. formula (3.29)) è uguale a

$$V(\overline{X}_{n_h}) = \frac{N_h - n_h}{N_h - 1} \frac{\sigma_h^2}{n_h} = \frac{1 - f_h}{1 - 1/N_h} \frac{\sigma_h^2}{n_h}$$

dove ora σ_h^2 rappresenta la varianza dello strato h, ovvero

$$\sigma_h^2 = \frac{1}{N_h} \sum_{j=1}^{N_h} (a_{h,j} - \mu_h)^2 .$$

Nel caso di un campionamento senza reinserimento avremo quindi

$$V\left(\overline{X}_{str,n}\right) = \sum_{h=1}^{H} \pi_h^2 \frac{1 - f_h}{1 - 1/N_h} \frac{\sigma_h^2}{n_h}. \tag{3.35}$$

Osservazione 12. La varianza σ^2 della popolazione non compare più nella formula (3.35) della varianza dello stimatore $\overline{X}_{str,n}$.

Osservazione 13. $V\left(\overline{X}_{str,n}\right)$ risulterà essere tanto più piccola quanto più omogenei saranno i diversi strati al loro interno. Ricordiamo che se uno strato h è perfettamente omogeneo (tutte le unità dello strato presentano lo stesso valore della caratteristica in esame) allora la sua varianza è evidentemente nulla. Nasce da qui il vantaggio di stratificare la popolazione quando gli strati possiedono una certa uniformità rispetto alla caratteristica studiata.

Esempio 28. Caso estremo di stratificazione. Supponiamo che

- la popolazione conti 100 individui;
- i possibili valori dell'attributo in esame siano solo 4;

[10] Rispetto alla formula (3.29) occorre evidentemente sostituire N con N_h e n con n_h.

- i 4 valori siano sconosciuti;
- si sappia quali e quanti siano gli individui col medesimo attributo.

Si potrebbe dunque stratificare la popolazione in quattro strati perfettamente omogenei al loro interno. La varianza delle sottopopolazioni è nulla! Sarebbe sufficiente campionare $n = 4$ individui - un individuo da ogni strato - per riuscire in questo caso estremo a stimare con precisione assoluta il valore atteso della popolazione!

Osservazione 14. La formula (3.35) esprime la varianza di $\overline{X}_{str,n}$ in funzione delle numerosità n_h dei sottocampioni. Sapendo che per un campione stratificato proporzionale vale (cf. la (3.34))

$$n_h = \frac{N_h}{N} n$$

è possibile riscrivere la varianza di $\overline{X}_{str,n}$ come

$$V(\overline{X}_{str,n}) = \frac{1-f}{n} \sum_{h=1}^{H} \pi_h \frac{N_h}{N_h - 1} \sigma_h^2. \tag{3.36}$$

3.6.3 Effetto della stratificazione sulla precisione di stima

Abbiamo visto in precedenza che quando il campione è ottenuto tramite campionamento casuale semplice senza reinserimento la varianza di $\overline{X}_n$ è uguale a (cf. (3.29))

$$V(\overline{X}_n) = \frac{N-n}{N-1} \frac{\sigma^2}{n} = \frac{1-f}{n} \frac{N}{N-1} \sigma^2. \tag{3.37}$$

Possiamo confrontare questo resultato con la formula (3.36) della varianza di $\overline{X}_{str,n}$ di un campione stratificato proporzionale. Tuttavia, prima di procedere al confronto vero e proprio, occorre chiarire che relazione sussiste tra la varianza dell'intera popolazione e la varianza di ogni singolo strato.

Teorema 1. Per qualsiasi popolazione stratificata vale la seguente relazione

$$\sigma^2 = \sum_{h=1}^{H} \pi_h \sigma_h^2 + \sum_{h=1}^{H} \pi_h (\mu_h - \mu)^2 \tag{3.38}$$

Interpretazione: La varianza della popolazione può essere decomposta come la somma pesata delle varianze di ogni singolo strato (misura della eterogeneità interna agli strati) più la varianza dei valori medi degli strati (misura di eterogeneità fra strati).

Utilizzando l'enunciato del Teorema 1 per il confronto fra le varianze $V(\overline{X}_n)$ e $V(\overline{X}_{str,n})$ dei due stimatori notiamo che

$$\begin{aligned} V(\overline{X}_n) - V(\overline{X}_{str,n}) &= \frac{1-f}{n}\frac{N}{N-1}\sigma^2 - \frac{1-f}{n}\sum_{h=1}^{H}\pi_h \frac{N_h}{N_h-1}\sigma_h^2 \\ &= \frac{1-f}{n}\left(\frac{N}{N-1}\sigma^2 - \sum_{h=1}^{H}\pi_h \frac{N_h}{N_h-1}\sigma_h^2\right) \end{aligned}$$

Per N e N_h sufficientemente grandi $\frac{N}{N-1} \simeq 1$ e $\frac{N_h}{N_h-1} \simeq 1$ dimodoché

$$\begin{aligned} V(\overline{X}_n) - V(\overline{X}_{str,n}) &\simeq \frac{1-f}{n}\left(\sigma^2 - \sum_{h=1}^{H}\pi_h\sigma_h^2\right) \\ &\underset{(3.38)}{=} \frac{1-f}{n}\sum_{h=1}^{H}\pi_h\left(\mu_h-\mu\right)^2 \geq 0. \end{aligned}$$

Da quest'ultima disuguaglianza deduciamo che la varianza della media calcolata utilizzando un campione stratificato proporzionale è in genere inferiore a quella di un campione casuale semplice di uguale numerosità. Il guadagno risultante dal processo di stratificazione della popolazione, rappresentato dalla quantità

$$\frac{1-f}{n}\sum_{h=1}^{H}\pi_h\left(\mu_h-\mu\right)^2 ,$$

è proporzionale alla varianza dei valori medi di strato. Esso sarà nullo se tutti i valori medi sono uguali fra loro.

Dimostrazione del Teorema 1. Notiamo con SS (SS_h) la somma del quadrato degli scarti dal valore atteso, ovvero

$$SS = \sum_{i=1}^{N}\left(a_i-\mu\right)^2 \text{ e } SS_h = \sum_{j=1}^{N_h}\left(a_{h,i}-\mu_h\right)^2 .$$

$$
\begin{aligned}
SS &= \sum_{i=1}^{N}(a_i-\mu)^2 = \sum_{i=1}^{N} a_i^2 - N\mu^2 = \sum_{h=1}^{H}\sum_{i=1}^{N_h} a_{h,i}^2 - N\mu^2 \\
&= \sum_{h=1}^{H}\sum_{i=1}^{N_h}\left(a_{h,i}^2-\mu_h^2+\mu_h^2\right) - N\mu^2 \\
&= \sum_{h=1}^{H}\left(\left(\sum_{i=1}^{N_h}\left(a_{h,i}^2-\mu_h^2\right)\right)+N_h\mu_h^2\right) - N\mu^2 \\
&= \sum_{h=1}^{H}\left(\underbrace{\left(\sum_{i=1}^{N_h} a_{h,i}^2 - N_h\mu_h^2\right)}_{SS_h} + N_h\mu_h^2\right) - \underbrace{N}_{\sum_{h=1}^{H} N_h}\mu^2 \\
&= \sum_{h=1}^{H} SS_h + \sum_{h=1}^{H} N_h\mu_h^2 - \sum_{h=1}^{H} N_h\mu^2 \\
&= \sum_{h=1}^{H} SS_h + \sum_{h=1}^{H} N_h\left(\mu_h^2-\mu^2\right)
\end{aligned}
$$

Il secondo termine, $\sum_{h=1}^{H} N_h\left(\mu_h^2-\mu^2\right)$ è uguale a $\sum_{h=1}^{H} N_h\left(\mu_h-\mu\right)^2$. Infatti

$$
\begin{aligned}
\sum_{h=1}^{H} N_h\left(\mu_h-\mu\right)^2 &= \sum_{h=1}^{H} N_h\left(\mu_h^2-2\mu_h\mu+\mu^2\right) = \\
&= \sum_{h=1}^{H} N_h\mu_h^2 - 2\mu\sum_{h=1}^{H} N_h\mu_h + \mu^2\sum_{h=1}^{H} N_h \\
&= \sum_{h=1}^{H} N_h\mu_h^2 - 2N\mu^2 + N\mu^2 = \sum_{h=1}^{H} N_h\left(\mu_h^2-\mu^2\right)
\end{aligned}
$$

Per il termine SS vale dunque

$$
SS = \sum_{h=1}^{H} SS_h + \sum_{h=1}^{H} N_h\left(\mu_h-\mu\right)^2 .
$$

Poiché $\sigma^2 = \frac{SS}{N}$ abbiamo che

$$\begin{aligned}
\sigma^2 &= \frac{1}{N}\sum_{h=1}^{H} SS_h + \frac{1}{N}\sum_{h=1}^{H} N_h\left(\mu_h - \mu\right)^2 \\
&= \frac{1}{N}\sum_{h=1}^{H} \frac{N_h}{N_h} SS_h + 1\sum_{h=1}^{H} \frac{N_h}{N}\left(\mu_h - \mu\right)^2 \\
&= \sum_{h=1}^{H} \frac{N_h}{N}\frac{SS_h}{N_h} + 1\sum_{h=1}^{H} \frac{N_h}{N}\left(\mu_h - \mu\right)^2
\end{aligned}$$

e quindi

$$\sigma^2 = \sum_{h=1}^{H} \pi_h \sigma_h^2 + \sum_{h=1}^{H} \pi_h \left(\mu_h - \mu\right)^2.$$

3.7 La stima di σ^2

Nei paragrafi precedenti è stata analizzata la proprietà di correttezza della media campionaria sotto diverse tipologie di campionamento: con e senza reinserimento, sistematico e stratificato. Oltre alla proprietà di correttezza dei vari stimatori ci siamo interessati anche alla loro varianza, ovvero alla precisione con la quale il valore atteso della popolazione è approssimato. Nelle diverse formule (3.22), (3.29), (3.35) e (3.36) inerenti alla varianza della media campionaria, figura sempre σ^2 (σ_h^2), la varianza della popolazione (dell'h-esimo strato). Ad esempio, per un campione casuale semplice estratto senza reimmissione la varianza della media campionaria è pari a

$$V(\overline{X}_n) = \frac{N-n}{N-1}\frac{\sigma^2}{n}.$$

N e n sono noti mentre σ^2 (σ_h^2) può o non può essere conosciuta.

1. Quando σ^2 è noto non sussiste alcun problema: $V(\overline{X}_n)$ è calcolabile.

2. Quando σ^2 (σ_h^2) non è noto.

   Quando la varianza σ^2 (σ_h^2) non è conosciuta la varianza degli stimatori del valore atteso μ studiati in precedenza non è calcolabile. Affinché le suddette formule siano utilizzabili, occorre stimare σ^2 (σ_h^2). Ai fini di questo corso - a meno che non venga specificato diversamente - utilizzeremo sempre S^2 (S_h^2) quale stimatore di σ^2 (σ_h^2)

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

o rispettivamente

$$S_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} \left(X_{h,i} - \overline{X}_h\right)^2 ,$$

dove

- n_h rappresenta la numerosità del sottocampione estratto dall'h-esimo strato,
- $X_{h,i}$ la variabile aleatoria relativa all'i-esima estrazione dallo strato h
- $\overline{X}_h$ la media calcolata sull'h-esimo sottocampione.

$V(\overline{X}_n)$ rappresenta la vera varianza della *media campionaria* $\overline{X}_n$, mentre S^2 è unicamento lo stimatore della varianza σ^2 di un singolo X_i. Inoltre, quando σ^2 (σ_h^2) non è nota ed è stimata tramite lo stimatore S^2 (S_h^2) aggiungeremo l'accento circonflesso alla V della varianza di $\overline{X}_n$. Ad esempio, scriveremo

$$\widehat{V}(\overline{X}_n) = \frac{N-n}{N-1}\frac{S^2}{n}$$

per indicare che si tratta della stima della varianza di $\overline{X}_n$ e non della sua vera varianza $V(\overline{X}_n)$.

3.8 L'intervallo di confidenza per μ

Quando si presenta il risultato di una stima si è soliti fornire tale risultato sotto forma di un valore (stima puntuale) più o meno una certa quantità. Ad esempio, se foste interessati ad organizzare il ballo dell'università potreste stimare il numero di partecipanti a $1'000 \pm 100$ persone, intendendo in tal modo che il numero di persone sarà compreso con "molta probabilità" fra le 900 e le $1'100$ persone. Un secondo esempio potrebbe essere quello di una manifestazione sportiva dove il numero esatto di partecipanti non è noto. Gli organizzatori dovranno stimare un intervallo plausibile di partecipanti. Sulla base di tale informazione verranno poi dimensionare le infrastrutture necessarie allo svolgimento della manifestazione. Capita dunque spesso di vedere delle stime in forma di intervallo. Vedremo nei successivi paragrafi come costruire un intervallo di confidenza per μ, ovvero un intervallo dentro il quale, con molta probabilità, è contenuto il valore μ. La varianza della popolazione o della variabile aleatoria X di cui si desidera stimare il valore atteso μ è, se non indicato diversamente, da considerarsi conosciuta. Prima di affrontare tale discussione è necessario presentare due importanti argomenti della teoria asintotica: la legge dei grandi numeri ed il teorema del limite centrale.

3.8.1 La legge (debole) dei grandi numeri

La legge dei grandi numeri è utile al fine di comprendere il comportamento *asintotico* di medie di variabili aleatorie. Studiare il comportamento asintotico di uno stimatore significa descrivere il suo comportamento quando la numerosità del campione diventa molto grande tendendo all'infinito. Supponiamo infatti di avere a disposizione una successione infinita $X_1, X_2, X_3, ...$ di variabili aleatorie

- *indipendenti*
- di valore atteso μ
- di varianza σ^2.

In questo paragrafo non imponiamo la condizione che le variabili aleatorie X_i abbiano la stessa distribuzione. Per mezzo di questa successione infinita costruiamo una seconda successione, la successione delle medie dei primi n elementi

$$\begin{array}{rcl} \overline{X}_1 & = & X_1 \\ \overline{X}_2 & = & (X_1 + X_2)/2 \\ \overline{X}_3 & = & (X_1 + X_2 + X_3)/3 \\ & \vdots & \\ \overline{X}_n & = & \frac{1}{n}\sum_{i=1}^n X_i \\ & \vdots & \end{array}$$

La successione delle medie $\overline{X}_1, \overline{X}_2, \overline{X}_3, ...$ è ancora una successione di variabili aleatorie. È immediato calcolare, per fisso n, il valore atteso e la varianza di $\overline{X}_n$:

$$\begin{array}{rcl} E(\overline{X}_n) & = & \mu \ , \\ V(\overline{X}_n) & = & \dfrac{\sigma^2}{n} \ . \end{array}$$

Come per la successione $X_1, X_2, X_3, ...$, le *V.A.* $\overline{X}_1, \overline{X}_2, \overline{X}_3, ...$ possiedono tutte il medesimo valore atteso. Tuttavia questo non è vero per la loro varianza, che tende a diminuire al crescere di n. Asintoticamente abbiamo che

$$\lim_{n\to\infty} V(\overline{X}_n) = 0 \ .$$

Ma questo significa che quando n è grande le realizzazioni di $\overline{X}_n$ saranno molto concentrate attorno al valore atteso μ.

Esempio 29. Per meglio comprendere quanto accade, aggiungiamo alle tre ipotesi precedenti l'ipotesi di normalità, vale a dire

- $X_1, X_2, X_3, ...$ sono distribuite secondo la legge Normale.

Questa ulteriore ipotesi non è necessaria per la conclusione a cui arriveremo, ma ci permette di eseguire il grafico della situazione. Infatti, poiché la somma pesata di $V.A.$ normali indipendenti è una $V.A.$ Normale, possiamo eseguire il grafico della funzione di densità degli $\overline{X}_n$, ad esempio per $n = 1, 2$ e 16. Scegliamo, senza perdita di generalità, $\mu = 1$ e $\sigma^2 = 4$.

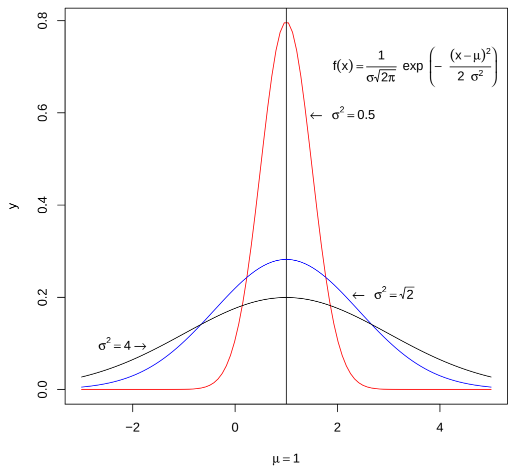


Figura 3.3: Funzione di densità di $\overline{X}_1$, $\overline{X}_2$ (blu) e $\overline{X}_{16}$(rossa)

Al crescere di n, la funzione di densità (e quindi la probabilità) tende a concentrarsi attorno al valore atteso μ (in questo esempio $\mu = 1$). Il prossimo teorema formalizza quanto visibile nella figura precedente al caso non Normale.

Teorema 2. Legge (debole) dei grandi numeri. Sia $X_1, X_2, X_3, ...$ una successione di $V.A.$ indipendenti, di valore atteso μ e di varianza $\sigma^2 < \infty$. Allora per qualsiasi $\varepsilon > 0$ piccolo a piacere si avrà che

$$\lim_{n\to\infty} P(|\overline{X}_n - \mu| > \varepsilon) = 0. \tag{3.39}$$

Per fisso $\varepsilon > 0$, la probabilità $P(|\overline{X}_n - \mu| \leq \varepsilon)$ corrisponde alla probabilità che il valore realizzato di $\overline{X}_n$ cada in un intervallo di centro μ e raggio ε. La (3.39) ci permette di affermare che questa probabilità tende a 1 al crescere di n all'infinito.

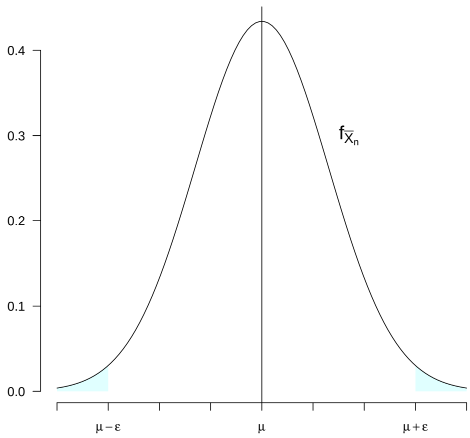


Figura 3.4: Legge dei grandi numeri

3.8.2 Uguaglianza e convergenza in distribuzione

Nel corso di Statistica I avete studiato che ad ogni variabile aleatoria a volori reali X è associata una funzione di ripartizione $F_X : \mathbb{R} \to [0, 1]$ definita per qualsiasi numero $c \in \mathbb{R}$ da

$$F_X(c) := P(X \leq c).$$

La funzione di ripartizione F_X caratterizza la variabile aleatoria X in termini della sua *probabilità*. Ora, date due variabili aleatorie X e Y è lecito chiedersi se esse abbiano la medesima distribuzione. Diremo che le due variabili aleatorie X e Y sono uguali in distribuzione, e in tal caso scriveremo $X \sim Y$, se

$$F_X(c) = F_Y(c) \quad \forall c \in \mathbb{R}. \tag{3.40}$$

Se le due variabili aleatorie X e Y sono uguali in distribuzione avremo dunque che

$$P(X \leq c) = P(Y \leq c) \quad \forall c \in \mathbb{R}. \tag{3.41}$$

A scanso di equivoci è bene precisare quanto segue. Affermare che $X \sim Y$ non significa che la realizzazione di X sarà uguale a quella di Y. Ad esempio, supponiamo di lanciare due volte un dado. Siano X e Y le $V.A.$ che descrivono il risultato del primo e, rispettivamente, del secondo lancio: entrambe le $V.A.$ hanno una distribuzione discreta uniforme $U(1, 6)$. X e Y sono uguali in distribuzione: $X \sim Y$. Ciò non toglie che nel primo lancio potrò osservare un cinque mentre nel secondo un tre.

Definizione 21. Convergenza in distribuzione. Diremo che una *qualsiasi* successione $X_1, X_2, X_3, \dots$ di $V.A.$ converge in distribuzione verso Y se per ogni $c \in \mathbb{R}$ con F_Y continua in c

$$\lim_{n\to\infty} F_{X_n}(c) = F_Y(c) \text{ o equivalentemente } \lim_{n\to\infty} P(X_n \leq c) = P(Y \leq c).$$

3.8.3 Il Teorema del Limite Centrale

Il Teorema del Limite Centrale (TLC) è di estrema importanza in probabilità e statistica in quanto è il fondamento sul quale sono costruiti moltissimi test d'ipotesi (rimandiamo al capitolo sull'inferenza statistica per la definizione formale di un test d'ipotesi statistica). Nelle pagine precedenti abbiamo visto che, sotto certe condizioni, una media di $V.A.$ converge verso il suo valore atteso. Il valore limite non è più una variabile aleatoria ma un numero ben preciso.

L'idea del teorema del limite centrale è quella di evitare, tramite un'opportuna trasformazione, che la successione di $V.A.$ $\overline{X}_1, \overline{X}_2, \overline{X}_3, \dots$ converga verso un numero (in questo caso il valore atteso μ). In altre parole, si desidera mantenere l'aleatorietà del limite della successione e, se possibile, ottenere come valore limite una $V.A.$ *la cui distribuzione sia nota*. Come fare dunque? L'idea è quella di standardizzare le $V.A.$ $\overline{X}_1, \overline{X}_2, \overline{X}_3, \dots$ così da ottenere una nuova successione di $V.A.$ che chiameremo $Y_1, Y_2, Y_3, \dots$. In pratica dovremo eseguire i due passi necessari alla standardizzazione, ovvero

1. sottrarre a $\overline{X}_n$ il suo valore atteso che in questo caso è μ ;
2. dividere per la sua deviazione standard $\sqrt{\frac{\sigma^2}{n}}$,

così da ottere una nuova successione $Y_1, Y_2, Y_3, \dots$ di $V.A. \sim (0, 1)$

$$Y_n = \frac{\overline{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} \ . \tag{3.42}$$

Esempio 30. Riprendiamo l'Esempio (29) in cui $\overline{X}_n \sim N(\mu, \frac{\sigma^2}{n})$. Ancora una volta, grazie alla proprietà della distribuzione Normale, dopo la standardizzazione le variabili aleatorie Y_n saranno ancora tutte normali $N(0, 1)$ per qualsiasi valore di $n = 1, 2, \dots$.

La proprietà di normalità delle $V.A.$ Y_n nell'Esempio 30 non è vera in generale. Se la successione di $V.A.$ $X_1, X_2, \ldots$ non è costituita da variabili normali, la media $\overline{X}_n$ non sarà più distribuita secondo la legge Normale e, di riflesso, pure Y_n. La conseguenza della non normalità delle $V.A.$ $X_1, \ldots, X_n$ è la non normalità di Y_n, sebbene che $Y_n \sim (0,1)$ per ogni $n = 1, 2, \ldots$. Cosa si potrà affermare sulla distribuzione di Y_n in questo caso? Il TLC è la soluzione a questa domanda.

Prima di presentare la versione classica del TLC facciamo notare una particolarità relativa alla standardizzazione di $\overline{X}_n$ (formula (3.42)). Tramite una serie di semplici trasformazioni algebriche è possibile riscrivere Y_n come una somma pesata di $V.A.$ $\sim (0,1)$.

Esercizio 8. Dimostrate dapprima che

$$\overline{X}_n - \mu = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)$$

e successivamente che

$$Y_n = \frac{\overline{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i \ , \tag{3.43}$$

dove $\varepsilon_i = \frac{X_i - \mu}{\sigma}$.

In pratica Y_n, che corrisponde alla media $\overline{X}_n$ standardizzata, puo' essere interpretato come la somma debitamente pesata delle $V.A.$ $\varepsilon_1, \ldots, \varepsilon_n$ dove ogni ε_i altro non è che $\frac{X_i - \mu}{\sigma} \sim (0,1)$. La somma $\sum_{i=1}^{n} \varepsilon_i$ delle n variabili aleatorie $i.i.d. \sim (0,1)$ ha valore atteso zero e varianza n. La standardizzazione richiede che il peso applicato alla somma $\sum_{i=1}^{n} \varepsilon_i$ sia $\frac{1}{\sqrt{n}}$, come appare nell'uguaglianza (3.43).

Teorema 3. Teorema del Limite Centrale (TLC). Sia $\varepsilon_1, \varepsilon_2, \ldots$ una successione di $V.A.$ indipendenti, di valore atteso nullo e varianza unitaria. Vale allora

$$Y_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i \underset{n \to \infty}{\sim} N(0,1). \tag{3.44}$$

Osservazione 15. Dal Teorema del Limite Centrale e la relazione (3.43) si deriva che, sotto le condizioni sopraelencate (quali sono?) sulla sequenza di $V.A.$ $X_1, X_2, \ldots$, per valori di n sufficientemente grandi la media standardizzata è distribuita come una variabile aleatoria Normale standard. È importante notare e *ricordarsi* che il denominatore nella parte sinistra della (3.42) non è altro che la deviazione standard di $\overline{X}_n$.

Una variabile aleatoria Normale standard è generalmente indicata con la lettera Z:

$$Z \sim N(0,1).$$

Il vantaggio di lavorare con una variabile aleatoria Normale standard risiede nel fatto che per essa è possibile calcolare probabilità del tipo

$$P(Z \leq c)$$

utilizzando le tavole che già conoscete. Ora se $\overline{X}_n$ soddisfa il TLC avremo che per n sufficientemente grande[11]

$$\frac{\overline{X}_n - \mu}{\sqrt{V(\overline{X}_n)}} \underset{\text{circa}}{\sim} Z.$$

Poiché Z e $\frac{\overline{X}_n - \mu}{\sqrt{V(\overline{X}_n)}}$ sono uguali in distribuzione avremo (confronta (3.40) e (3.41)) che

$$P\left(\frac{\overline{X}_n - \mu}{\sqrt{V(\overline{X}_n)}} \leq c\right) = P\left(\frac{\overline{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} \leq c\right) \underset{TLC}{\simeq} P\left(Z \leq c\right) \quad \forall c \in \mathbb{R}.$$

Sarà questo il punto di partenza per la costruzione dell'intervallo di confidenza di μ.

3.8.4 L'intervallo di confidenza

Iniziamo la costruzione dell'intervallo di confidenza ragionando sulla variabile aleatoria Z. Siamo interessati a calcolare la probabilità

$$P(|Z| \leq 1.96) = P(-1.96 \leq Z \leq 1.96).$$

Detto a parole stiamo semplicemente calcolando la probabilità che la realizzazione di Z sia compresa nell'intervallo $[-1.96, 1.96]$. Poiché Z è una variabile aleatoria Normale standard è immediato verificare che

$$P(|Z| \leq 1.96) = 0.95 \ . \tag{3.45}$$

[11] L'uguaglianza non sarà esatta ma al crescere di n all'infinito eventuali differenze saranno in pratica trascurabili.

Ora, poiché per n sufficientemente grande Z e $\frac{\overline{X}_n-\mu}{\sqrt{V(\overline{X}_n)}}$ sono approssimativamente uguali in distribuzione[12], possiamo scrivere

$$P(|Z| \leq 1.96) = P(|\frac{\overline{X}_n - \mu}{\sqrt{V(\overline{X}_n)}}| \leq 1.96)$$

da cui ricaviamo

$$P(\mu - 1.96\sqrt{V(\overline{X}_n)} \leq \overline{X}_n \leq \mu + 1.96\sqrt{V(\overline{X}_n)}) = 0.95. \tag{3.46}$$

All'uguaglianza (3.46) si dà la seguente interpretazione: 95 volte su 100 lo stimatore $\overline{X}_n$ assumerà dei valori compresi nell'intervallo di centro μ e raggio $1.96\sqrt{V(\overline{X}_n)}$, ovvero

$$P(\overline{X}_n \in \left[\mu - 1.96\sqrt{V(\overline{X}_n)}\ , \mu + 1.96\sqrt{V(\overline{X}_n)}\right]) = 0.95. \tag{3.47}$$

Tuttavia, essendo μ sconosciuto, questa informazione è poco utile. Infatti dove sarà posizionato sulla retta dei numeri reali l'intervallo

$$\left[\mu - 1.96\sqrt{V(\overline{X}_n)}\ , \mu + 1.96\sqrt{V(\overline{X}_n)}\right]\ ?$$

Per mezzo di semplici trasformazioni algebriche possiamo però riscrivere la (3.46) nel seguente modo

$$P(\overline{X}_n - 1.96\sqrt{V(\overline{X}_n)} \leq \mu \leq \overline{X}_n + 1.96\sqrt{V(\overline{X}_n)}) = 0.95. \tag{3.48}$$

Quale interpretazione possiamo ora assegnare alla nuova rappresentazione (3.48) della precedente uguaglianza (3.46)? Per dare un significato a questa espressione osserviamo i due estremi dell'espressione

$$\overline{X}_n - 1.96\sqrt{V(\overline{X}_n)} \leq \mu \leq \overline{X}_n + 1.96\sqrt{V(\overline{X}_n)}$$

che definiscono l'intervallo

$$\left[\overline{X}_n - 1.96\sqrt{V(\overline{X}_n)}\ , \overline{X}_n + 1.96\sqrt{V(\overline{X}_n)}\right]. \tag{3.49}$$

L'estremo sinistro di questo intervallo è dato da $\overline{X}_n - 1.96\sqrt{V(\overline{X}_n)}$. Il termine $1.96\sqrt{V(\overline{X}_n)}$ non è aleatorio ma è un numero. $\overline{X}_n$ per contro è

[12]Ricordiamo che l'uguaglianza è vera solo al limite.

una variabile aleatoria. Lo stesso ragionamento si applica all'estremo destro dell'intervallo. Per tale motivo l'intervallo definito dalla (3.49) è chiamato *intervallo aleatorio* o più comunemente *intervallo di confidenza al 95%.* La (3.48) può essere riscritta in maniera simile alla (3.47) utilizzanzo la notazione insiemistica "$\in$" ovvero

$$P(\mu \in \left[\overline{X}_n - 1.96\sqrt{V(\overline{X}_n)}\ , \overline{X}_n + 1.96\sqrt{V(\overline{X}_n)}\right]) = 0.95. \tag{3.50}$$

Siamo ora pronti a dare un'interpretazione alla (3.48): con una probabilità del 95% l'intervallo aleatorio (3.49) conterrà il valore μ. Oppure: 95 volte su 100 l'intervallo $\left[\overline{X}_n - 1.96\sqrt{V(\overline{X}_n)}\ , \overline{X}_n + 1.96\sqrt{V(\overline{X}_n)}\right]$ conterrà μ.

Esiste una sottile ma importante differenza fra la (3.47) e la (3.50). Nella (3.47) l'intervallo è deterministico ma sconosciuto e quindi inutilizzabile. Nella (3.50) per contro l'intervallo è osservabile ma aleatorio ed il punto μ è deterministico ma sconosciuto. È importante sottolineare l'equivalenza delle due espressioni: è possibile passare dall'una all'altra tramite semplici operazioni algebriche. È tuttavia la seconda espressione in termini dell'intervallo aleatorio

$$\left[\overline{X}_n - 1.96\sqrt{V(\overline{X}_n)}\ , \overline{X}_n + 1.96\sqrt{V(\overline{X}_n)}\right]$$

quella che ci fornisce uno strumento pratico per la stima del valore atteso μ. Infatti sulla base dei valori realizzati di $X_1, ..., X_n$, notati $x_1, ..., x_n$, calcoleremo dapprima la realizzazione di $\overline{X}_n$, notata $\overline{x}_n$, ed in seguito la realizzazione dell'intervallo aleatorio ovvero l'intervallo[13]

$$\left[\overline{x}_n - 1.96\frac{\sigma}{\sqrt{n}}\ ,\ \overline{x}_n + 1.96\frac{\sigma}{\sqrt{n}}\right]. \tag{3.51}$$

Se μ fosse conosciuto, potremmo verificare se l'intervallo così calcolato lo contiene oppure no. Tuttavia questa verifica è impossibile. Quello che sappiamo è che nel 95% dei casi questa procedura genera un intervallo contenente μ.

3.8.5 L'ampiezza dell'intervallo di confidenza

Vogliamo studiare l'ampiezza dell'intervallo di confidenza e ci chiediamo quali siano i fattori che la determinano. Osservando la formula (3.49) di intervallo aleatorio notiamo che l'ampiezza dell'intervallo è data dal termine

[13]Sappiamo che quando le osservazioni sono i.i.d. $V(\overline{X}_n) = \frac{\sigma 2}{n}$. Ricordiamo che in questo paragrafo la varianza σ^2 è da considerarsi come nota.

$1.96\sqrt{V(\overline{X}_n)}$, ovvero dalla combinazione dei due termini 1.96 e $\sqrt{V(\overline{X}_n)}$. Come già sappiamo la varianza di $\overline{X}_n$ dipende dalla numerosità n del campione, dalla varianza della popolazione e dalla tecnica di campionamento utilizzata. La costante 1.96 dipende invece dalla livello di probabilità che era stato scelto all'inizio di questo paragrafo e che ammonta al 95%. Ricordiamo infatti che la costante 1.96 è stata calcolata come la soluzione in c della seguente equazione (confronta la (3.45))

$$P(|Z| \leq c) = 0.95$$

e corrisponde al 0.975-quantile della distribuzione Normale.

La probabilità $\alpha = 0.95$ è chiamata il *livello di confidenza*. Essa rappresenta la confidenza che riponiamo nel fatto che la realizzazione dell'intervallo aleatorio (3.51) contenga μ. È il ricercatore a determinare il livello di confidenza. $\alpha = 90\%$, 95% e 99% sono i valori comunemente utilizzati. Fissato dunque il livello di confidenza α desiderato, occorre risolvere l'equazione

$$P(|Z| \leq c) = \alpha$$

in funzione di c. Le soluzioni per i tre livelli α di confidenza indicati in precedenza corrispondono agli $(\frac{1+\alpha}{2})$-quantili della distribuzione Normale e sono nell'ordine uguali a 1.64, 1.96 e 2.58. Essi rimpiazzerano il valore della costante 1.96 nella costruzione dell'intervallo di confidenza (3.51). Riassumendo: il valore che la costante c assume in funzione del livello α di confidenza corrisponde all'$(\frac{1+\alpha}{2})$-quantile della distribuzione Normale. Ad esempio, per $\alpha = 90\%$, lo 0.95-quantile è uguale a 1.64.

Esercizio 9. Descrivete come varia l'ampiezza dell'intervallo di confidenza al variare

- del livello di significatività α;
- della numerosità n del campione.

Capitolo 4

Teoria della stima

Nel capitolo introduttivo è stato detto che lo scopo dell'induzione statistica è quello di trarre delle conclusioni di validità generale rispetto ad una o più caratteristiche di una popolazione obiettivo partendo da un insieme limitato di osservazioni (campione). È stato inoltre osservato come la validità di tali conclusioni non sia assoluta e definitiva, ma soggetta a continua verifica. Con l'arrivo di ulteriori informazioni sarà necessario confermare alla luce della nuova evidenza empirica quanto inferito precedentemente. L'inferenza statistica si occupa quindi di due importanti problemi: la stima e la verifica d'ipotesi. In questo capitolo ci occuperemo del problema di stima ed in particolare della *stima parametrica puntuale*.

Come descritto nel capitolo 2, assumeremo che la caratteristica in esame della popolazione obiettivo possa essere rappresentata da una variabile aleatoria X a valori reali la cui distribuzione è conosciuta a meno di un parametro θ. In altre parole si ipotizza che la distribuzione della variabile aleatoria X appartenga ad una famiglia di distribuzioni (cf. paragrafo 2.2)

$$\mathcal{P} = \{f(x;\theta),\ \theta \in \Theta\}$$

dove $f(x;\theta)$ come al solito rappresenta la funzione di densità (caso continuo) o di probabilità (caso discreto) di X. Per fisso θ, la densità[1] $f(x;\theta)$ è dunque nota. Diremo che

1. il modello parametrico è correttamente specificato se esiste un $\theta \in \Theta$ tale per cui $f(x;\theta)$ corrisponde alla *vera funzione* di densità di X;

2. θ è identificato, se per qualsiasi altro valore $\theta' \in \Theta$ vale che $f(x;\theta) \neq f(x;\theta')$.

[1]Si parlerà semplicemente di densità intendendo qualora X fosse una *V.A.* discreta la funzione di probabilità di X.

Se il giusto valore del parametro θ fosse conosciuto, la funzione di densità (e quindi la distribuzione di X) sarebbe completamente specificata ed in tal caso non sarebbe necessario ricorrere all'inferenza. Tuttavia, non essendo questo il caso, dovremo, sulla base di un n-campione di X, stimare il valore θ. Avremo a questo punto due possibilità: eseguire una stima puntuale di θ oppure costruire un intervallo di confidenza per θ.

Esempio 31. Nel caso della stima del valore medio μ della popolazione abbiamo stimato il valore atteso della *V.A.* X sulla base del valore assunto dalla statistica (stimatore) $\overline{X}_n$. La realizzazione di $\overline{X}_n$, notata $\overline{x}_n$, è dunque una stima puntuale di μ.

Osservazione 16. In generale indicheremo con $T(X_1, X_2, ..., X_n)$ la statistica utilizzata quale stimatore puntuale di θ. Nell'esempio 31 abbiamo dunque $T(X_1, X_2, ..., X_n) = \overline{X}_n$.

La seconda possibilità per stimare θ è quella di costruire un intervallo di confidenza. In tal caso avremo bisogno di due statistiche, che indicheremo con $I_1(X_1, ..., X_n)$ e $I_2(X_1, ..., X_n)$, tali per cui $I_1 < I_2$ per qualsiasi realizzazione di $X_1, ..., X_n$. I_1 e I_2 sono gli estremi dell'intervallo aleatorio $I = [I_1(X_1, ..., X_n)\ , I_2(X_1, ..., X_n)]$.

Esempio 32. Nel paragrafo 3.8.4, le due statistiche I_1 e I_2 ad un livello di confidenza del 95% erano date da (ricordiamo che che la varianza σ^2 era considerata nota)

$$I_1 = \overline{X}_n - 1.96\sqrt{\sigma^2/n} \text{ e } I_2 = \overline{X}_n + 1.96\sqrt{\sigma^2/n}.$$

Questo secondo approccio per stimare il parametro sconosciuto è chiamato stima per intervalli. In questo capitolo ci soffermeremo sulla stima puntuale. Per tutto il presente capitolo, salvo indicato diversamente, assumeremo che

- X è una variabile aleatoria con funzione di densità in $\mathcal{P}$, cioè esiste un $\theta \in \Theta$ tale per cui $f(x; \theta)$ è la densità di X;
- θ è identificato;
- θ è sconosciuto;
- $X_1, ..., X_n$ è un n-campione di X.

Le variabili aleatorie $X_1, ..., X_n$ saranno quindi indipendenti ed identicamente distribuite. Il nostro obiettivo consiste nel trovare delle statistiche (funzioni di $X_1, ..., X_n$ e dunque a tutti gli effetti delle *V.A.*) da usare quali stimatori del parametro sconosciuto θ.

4.1 Il metodo dei momenti

In questa sezione del capitolo studieremo un metodo di stima chiamato *metodo dei momenti*. Questo suo nome trae origine dal fatto che l'idea chiave su cui esso è costruito consiste nel concetto di momento di una variabile aleatoria. Prima di affrontare questa classe di stimatori diamo alcune definizioni.

Definizione 22. Momento r-esimo. Sia X una $V.A.$ con funzione di densità f. Il momento r-esimo di X, notato M_r, è definito come

$$M_r = E(X^r)$$

ammesso che il valore atteso esista.

Quando $r = 1$, M_1 corrisponde al valore atteso di X che indicheremo semplicemente come d'abitudine con μ. M_2 è il secondo momento di X che ritroviamo ad esempio nella formula alternativa della varianza di X

$$V(X) = E(X^2) - \mu^2.$$

Utilizzando questa nuova notazione, $V(X)$ può dunque essere scritta come $M_2 - \mu^2$.

Osservazione 17. Per una variabile aleatoria X distribuita simmetricamente[2] attorno allo 0 e tale per cui $E\left(\mid X^r \mid\right) < \infty$, gli r-esimi momenti dispari sono tutti uguali a 0.

L'r-esimo momento di una variabile aleatoria X non è altro che il valore atteso della variabile aleatoria Y, dove $Y := g(X)$ con

$$g : \mathbb{R} \to \mathbb{R},\ x \mapsto y = x^r.$$

Dal corso di Statistica I sappiamo che il valore atteso di Y può essere calcolato come

$$E(Y) = E\left(X^r\right) = \int_{-\infty}^{\infty} x^r f(x;\theta)dx. \tag{4.1}$$

Come intuibile dalla (4.1) (verificate nell'Esempio 34), il valore di

$$M_r = \int_{-\infty}^{\infty} x^r f(x;\theta)dx$$

dipenderà dal particolare valore del parametro sconosciuto θ. Quando necessario espliciteremo la dipendenza di M_r dal parametro sconosciuto θ scrivendo semplicemente $M_r(\theta)$.

[2] X possiede una distribuzione simmetrica rispetto allo 0 se $f(x) = f(-x)$.

Esempio 33. Supponiamo che la variabile aleatoria X sia distribuita secondo la legge Normale $N(0, \sigma^2)$ dove la varianza σ^2 della distribuzione è sconosciuta. In questo caso il parametro sconosciuto θ corrisponde alla varianza σ^2. Consideriamo il quarto momento

$$M_4 = E\left(X^4\right) = \int_{-\infty}^{\infty} x^4 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{x^2}{\sigma^2}\right) dx \ . \tag{4.2}$$

Anziché calcolare direttamente l'integrale (4.2) definiamo la *V.A.* Z

$$Z = \frac{X}{\sigma} \sim N(0,1).$$

Utilizzando l'integrazione per parti (si veda il suggerimento), è semplice calcolare il valore atteso di

$$\begin{aligned} E\left(Z^4\right) &= \int_{-\infty}^{\infty} z^4 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) dz \\ &= \int_{-\infty}^{\infty} \underbrace{z^3}_{v} \underbrace{z\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right)}_{u'} dz. \\ &= \ldots = 3 \end{aligned}$$

Da cui otteniamo

$$3 = E\left(Z^4\right) = E\left(\left(\frac{X}{\sigma}\right)^4\right) = \frac{1}{\sigma^4} E\left(X^4\right)$$

e quindi

$$M_4 = E\left(X^4\right) = 3\sigma^4 = 3\left(\sigma^2\right)^2.$$

Come potete vedere, in questo esempio il quarto momento dipende dal valore del parametro sconosciuto, in questo caso la varianza di X.

Definizione 23. Momento r-esimo centrato. Il momento r-esimo *centrato* di X, notato M_r^c è definito come

$$M_r^c = E\left((X - \mu)^r\right).$$

Il primo momento centrato di X è dunque uguale a 0. Il secondo momento centrato è la varianza di X, da cui segue come appena visto l'uguaglianza

$$M_2^c = M_2 - \mu^2.$$

Se X è simmetrica rispetto al suo valore atteso e tale per cui $E(\mid X^r \mid) < \infty$ allora tutti i momenti centrati di ordine dispari saranno uguali a 0.

Osservazione 18. (Curiosità). Esiste la seguente relazione tra i momenti M_r ed i momenti centrati M_r^c

$$M_r^c = \sum_{i=0}^{r} (-1)^i \ C_r^i \ \mu^i \ M_{r-i}$$

dove $C_r^i = \frac{r!}{i!(r-i)!}$ è il coefficiente binomiale. La dimostrazione della precedente uguaglianza segue direttamente dalla linearità del valore atteso e dal fatto che

$$(a+b)^n = \sum_{i=0}^{n} C_r^i \ a^i \ b^{n-i}.$$

Esempio 34. Sia $X_1, ..., X_n$ un n-campione estratto da una distribuzione di Poisson. La famiglia parametrica $\mathcal{P}$ di X è data da

$$\mathcal{P} = \left\{ f(x;\theta) = \frac{\theta^x e^{-\theta}}{x!}, \ x \in \{0,1,...\}, \ \theta \in \mathbb{R}^+ \right\}.$$

Il momento M_1 è semplicemente μ, il valore atteso di X:

$$\mu = \sum_{x=0}^{\infty} x \frac{\theta^x e^{-\theta}}{x!} = \theta. \tag{4.3}$$

Il secondo momento M_2 è invece

$$M_2 = \sum_{x=0}^{\infty} x^2 \frac{\theta^x e^{-\theta}}{x!} = \theta + \theta^2. \tag{4.4}$$

Infine, per il terzo momento M_3 vale

$$M_3 = \sum_{x=0}^{\infty} x^3 \frac{\theta^x e^{-\theta}}{x!} = \theta + 3\theta^2 + \theta^3. \tag{4.5}$$

Come potete ancora una volta constatare i momenti M_1, M_2 e M_3 sono particolari funzioni del parametro θ (confronta Figura 4.1).

Le Definizioni 22 e 23 considerano i momenti teorici della popolazione (o di una variabile aleatoria X). È tuttavia possibile estendere la definizione di r-esimo momento ad un n-campione $X_1, ..., X_n$.

Definizione 24. Momenti campionari. Sia $X_1, ..., X_n$ un n-campione estratto da una popolazione X. Il momento campionario r-esimo, notato $M_{r,n}$, è definito come

$$M_{r,n} = \frac{1}{n} \sum_{i=1}^{n} X_i^r.$$

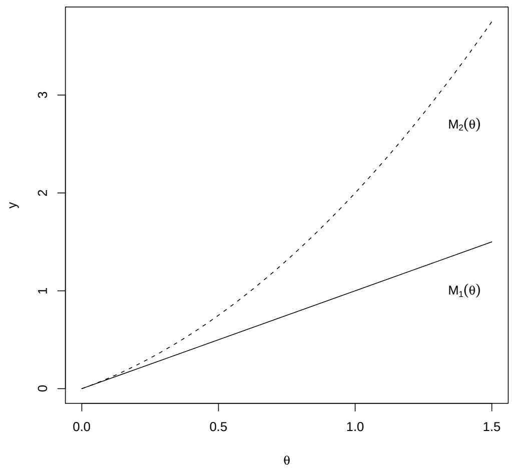


Figura 4.1: Primo e secondo momento della distribuzione di Poisson in funzione del parametro θ.

Il primo momento campionario corrisponde a $\overline{X}_n$, la media del campione. Analogamente alla Definizione 23 il momento campionario r-esimo centrato è definito come

$$M_{r,n}^c = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^r.$$

Il secondo momento campionario centrato $M_{2,n}^c = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$ è l'equivalente *empirico* di σ^2, la varianza di X. È interessante osservare che lo stimatore S^2 della varianza è derivabile da $M_{2,n}^c$ tramite la semplice formula

$$S^2 = \frac{n}{n-1}M_{2,n}^c.$$

Il fattore di proporzionalità $\frac{n}{n-1}$ tende a 1 al crescere della numerosità n del campione.

Osservazione 19. Sia $M_{r,n}$ che $M_{r,n}^c$ sono delle statistiche (delle funzioni di $X_1, ..., X_n$ che possono essere calcolate in quanto non dipendono da nessun parametro sconosciuto) e quindi sono esse stesse delle variabili aleatorie. In generale non è possibile ricavare analiticamente la distribuzione di queste statistiche. Tuttavia, per alcuni valori di r, è possibile calcolarne il valore atteso e/o la varianza.

4.1.1 Convergenza dei momenti campionari

La proprietà fondamentale che caratterizza i momenti campionari è la loro convergenza verso il rispettivo momento teorico. Più precisamente vale il seguente teorema.

Teorema 4. Sia $X_1, ..., X_n$ un n-campione estratto da una popolazione X avente i primi r-momenti finiti, ovvero: $E\left(X^j\right) = M_j$, $j = 1, \ldots, r$. Allora per qualsiasi $\varepsilon > 0$ si avrà che

$$\lim_{n\to\infty} P(|M_{r,n} - M_r| > \varepsilon) = 0. \tag{4.6}$$

Per fisso $\varepsilon > 0$ con ε piccolo a piacere la probabilità $P(|M_{r,n} - M_r| \leq \varepsilon)$ corrisponde alla probabilità che il valore[3] realizzato di $M_{r,n}$ cada in un intervallo di centro M_r e raggio ε. Il Teorema 4 afferma semplicemente che per quanto piccolo possa essere l'intervallo attorno al vero valore del momento teorico M_r, al crescere di n all'infinito il valore calcolato del momento campionario (utilizzando l' n-campione) sarà contenuto nell'intervallo stesso con probabilità prossima ad 1.

4.1.2 Stimatore dei momenti

Abbiamo ora tutti gli elementi necessari per definire lo stimatore dei momenti e comprendere l'idea che lo caratterizza. Ricapitolando abbiamo visto che

- La distribuzione della variabile aleatoria X è sconosciuta. Tuttavia sappiamo che essa appartenere alla famiglia di distribuzioni

  $$\mathcal{P} = \left\{f(x;\theta),\ \theta \in \Theta\right\}.$$

  In particolare esiste un unico $\theta \in \Theta$, che indichiamo con θ_0, tale per cui $f(x;\theta_0)$ corrisponde alla funzione di densità f di X.

- Per fisso $\theta \in \Theta$ possiamo calcolare l'r-esimo momento (momento centrato) di X che sarà notato $M_r(\theta)$ o rispettivamente $M_r^c(\theta)$.

- Osservando un n-campione $X_1, ..., X_n$ estratto dalla popolazione X di densità f possiamo calcolare l'r-esimo momento (centrato) campionario $M_{r,n}$ ($M_{r,n}^c$). Il limite per $n \to \infty$ di $M_{r,n}$ è uguale al corrispettivo momento della popolazione $M_r(\theta_0)$. Avremo dunque che per n "sufficientemente" grande

  $$M_{r,n} \simeq M_r(\theta_0). \tag{4.7}$$

[3] Il valore che noi calcoleremo una volta osservate $X_1, ..., X_n$.

Esempio 35. Supponiamo che la variabile aleatoria X possegga una distribuzione di Poisson. Abbiamo visto che per $r = 3$ vale la seguente relazione

$$M_3(\theta) = \theta + 3\theta^2 + \theta^3.$$

Questo significa che per n sufficientemente grande avremo

$$M_{3,n} \simeq M_3(\theta_0) = \theta_0 + 3\theta_0^2 + \theta_0^3.$$

L'approssimazione (4.7) è il punto di partenza dal quale costruire lo stimatore. Infatti, il lato sinistro della (4.7) è calcolabile una volta che il campione $X_1, ..., X_n$ è osservato. Il lato destro invece contiene il parametro incognito θ_0. Possiamo però tentare di risolvere "l'equazione" (4.7) rispetto a θ_0 ottenendo così un'approssimazione (stima) di θ_0 che indicheremo $\widehat{\theta}$. Nel caso dell'Esempio 35 supponiamo che la realizzazione di $M_{3,n}$, notata $m_{3,n}$ sia uguale 1. Scriveremo

$$m_{3,n} = 1 \simeq M_3(\theta_0) = \theta_0 + 3\theta_0^2 + \theta_0^3. \tag{4.8}$$

Risolvendo rispetto a θ_0 otteniamo le tre soluzioni $\sqrt{2}-1, -\sqrt{2}-1, -1$ di cui solo la prima è ammissibile in quanto, nel caso in esame della distribuzione di Poisson, il valore del parametro θ deve essere compreso nell'intervallo $(0, \infty)$. Poiché la soluzione è stata ricavata da un'approssimazione (stima) del momento teorico, scriveremo

$$\widehat{\theta} = \sqrt{2} - 1$$

e non $\theta_0 = \sqrt{2} - 1$.

Cambiando l'ordine r del momento cambia sia la relazione che lega il valore teorico del momento al valore del parametro sia il valore stimato del momento empirico corrispondente. Ad esempio, se nell'Esempio 35 anziché utilizzare il terzo momento utilizzassimo il secondo momento, otterremmo la seguente relazione

$$M_{2,n} \simeq M_2(\theta_0) = \theta_0 + \theta_0^2. \tag{4.9}$$

Con lo stesso n-campione utilizzato in precedenza per il calcolo di $M_{3,n}$ potremmo ora calcolare $M_{2,n}$, utilizzare tale valore per risolvere la (4.9) ed ottenere in questo modo la nuova stima $\widehat{\theta}$ di θ_0. Tale stima verosimilmente sarà diversa dalla precedente in quanto i due stimatori utilizzano l'informazione dell'$n-$campione in maniera diversa. Per fare un esempio, se trovassimo che $m_{2,n} = 0.39$, avremmo quale unica soluzione ammissibile $\widehat{\theta} = 0.3$ (diversa quindi dalla soluzione precedente). In questo caso diremo che $\widehat{\theta} = 0.3$ è la stima di θ_0 basata sul secondo momento.

È possibile dare un'interpretazione grafica della situazione? Sì, poiché risolvere le equazioni (4.8) e (4.9) rispetto a θ_0 corrisponde, in pratica, a calcolare la preimmagine della funzione dei momenti

$$M_r(\cdot) : \Theta \to \mathbb{R} \ , \ \theta \mapsto M_r(\theta)$$

nel punto $m_{r,n}$. Infatti, applicando la funzione inversa M_r^{-1} ad ambo i lati della (4.7) otteniamo

$$\widehat{\theta} = M_r^{-1}(m_{r,n}) \simeq M_r^{-1}\left(M_r(\theta_0)\right) = \theta_0. \tag{4.10}$$

La quantità $M_r^{-1}(m_{r,n})$ è calcolabile a condizione che la funzione inversa M_r^{-1} esista (l'equazione abbia un'unica soluzione). $M_r^{-1}(m_{r,n})$ rappresenta un'approssimazione (stima) del vero parametro θ_0. Come già accennato precedentemente questo è il motivo per cui utilizzeremo la notazione $\widehat{\theta}$ anziché θ_0 nell'indicare il valore stimato.

Riassumendo, quando la funzione inversa M_r^{-1} è disponibile in forma analitica lo stimatore col metodo dei momenti di θ_0 è uguale a

$$\widehat{\theta} = M_r^{-1}(M_{r,n}), \tag{4.11}$$

altrimenti $\widehat{\theta}$ verrà calcolato quale soluzione rispetto a θ dell'equazione

$$M_{r,n} = M_r(\theta)$$

come visto nei due esempi sulla distribuzione di Poisson con $r = 2, 3$. La Figura 4.2 mostra graficamente quanto appena descritto.

Esempio 36. Supponiamo che $X_1, ..., X_n$ sia un campione casuale estratto da una distribuzione Normale $N\left(\mu_0, \sigma_0^2\right)$ il cui valore atteso μ_0 e varianza σ_0^2 sono entrambi sconosciuti. La famiglia parametrica di distribuzioni $\mathcal{P}$ è uguale

$$\mathcal{P} = \left\{ f(x;\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\left(x-\mu\right)^2\right), \ x \in \mathbb{R} \ , \ \theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+ \right\}$$

In questo esempio il parametro sconosciuto θ è costituito da due elementi: $\theta = (\theta_1, \theta_2) = (\mu, \sigma^2)$. Parleremo perciò di due parametri sconosciuti. Sarà necessario utilizzare due momenti distinti al fine di ottenere un sistema di due equazioni e due incognite. Prendiamo arbitrariamente i primi due momenti M_1 e M_2 e costruiamo il seguente sistema di equazioni[4]

$$\begin{cases} M_{1,n} \simeq M_1\left(\mu_0, \sigma_0^2\right) = \mu_0 \\ M_{2,n} \simeq M_2\left(\mu_0, \sigma_0^2\right) = \sigma_0^2 + \mu_0^2 \end{cases} \tag{4.12}$$

[4]Ricordiamo che

$$\begin{aligned} \sigma^2 &= E(X^2) - \left(E(X)\right)^2 \\ &= M_2 - \mu^2 \end{aligned}$$

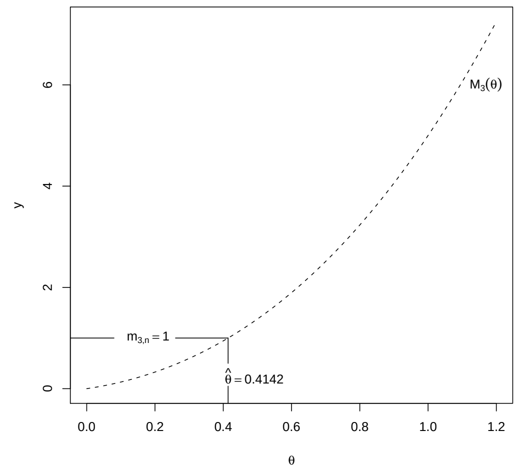


Figura 4.2: Terzo momento della distribuzione di Poisson in funzione di θ.

Il primo momento campionario $M_{1,n}$ corrisponde alla media campionaria $\overline{X}_n$ mentre il secondo momento campionario $M_{2,n}$ è uguale $\frac{1}{n}\sum_{i=1}^{n} X_i^2$. Trascurando l'errore d'approssimazione riscriviamo il sistema precedente utilizzando l'uguaglianza stretta così da ottenere il sistema a due equazioni e due incognite desiderato

$$\begin{cases} \overline{X}_n = \mu \\ \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \sigma^2 + \mu^2 \end{cases} .$$

Risolvendo rispetto ai due parametri sconosciuti μ e σ^2 otteniamo lo stimatore dei momenti di $\theta_0 = (\mu_0, \sigma_0^2)$

$$\begin{aligned} \widehat{\mu} &= \overline{X}_n \\ \widehat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\overline{X}_n\right)^2 = \frac{1}{n}\sum_{i=1}^{n} \left(X_i - \overline{X}_n\right)^2 . \end{aligned}$$

Lo stimatore dei momenti della varianza σ^2 di X è diverso da S^2 che sappiamo

da cui si ricava facilmente

$$M_2 = \sigma^2 + \mu^2.$$

essere uno stimatore corretto. Infatti per $\widehat{\sigma}^2$ vale che

$$\begin{aligned} E\left(\widehat{\sigma}^2\right) &= E\left(\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}_n\right)^2\right) = E\left(\frac{n-1}{n}\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}_n\right)^2\right) \\ &= \frac{n-1}{n}E\left(S^2\right) = \frac{n-1}{n}\sigma^2 = \left(1-\frac{1}{n}\right)\sigma^2. \end{aligned}$$

La distorsione (Bias in inglese) di $\widehat{\sigma}^2$ quale stimatore di σ^2 è uguale per definizione a

$$B(\widehat{\sigma}^2) := E\left(\widehat{\sigma}^2\right) - \sigma^2 = -\frac{1}{n}\sigma^2.$$

Tuttavia, per $n \to \infty$ tale distorsione tende a 0. Diremo allora che $\widehat{\sigma}^2$ è uno stimatore asintoticamente corretto di σ^2.

4.2 Il metodo di massima verosimiglianza

L'idea alla base del metodo di stima di massima verosimiglianza è molto semplice ed intuitiva. Il seguente esempio preso da A. Mood, F. Graybill e D. Boes (Introduzione alla statistica, McGraw-Hill, ISBN 88-386-0661-7) ne illustra l'idea.

Supponiamo che un'urna contenga un certo numero di palline nere e un certo numero di palline bianche, e supponiamo che si sappia che il rapporto fra i due numeri è di 3/1, ma che non si sappia se le palline più numerose siano le nere o le bianche. In altre parole, la probabilità di estrarre una pallina nera è 1/4 oppure 3/4. Se si estraggono dall'urna *con reimmissione* n palline, la distribuzione di X, la variabile aleatoria che denota il numero di palline nere presenti nel campione, è data dalla distribuzione binomiale

$$f(x;p) = C_{n,x}\, p^x q^{n-x} \qquad \text{per } x = 0, 1, 2, ..., n,$$

dove $q = 1 - p$ e p è la probabilità di estrarre una pallina nera. In questo caso $p = 1/4$ oppure $p = 3/4$. Estrarremo con reimmissione un campione di tre palline, cioè $n = 3$ e tenteremo di stimare il parametro incognito p della distribuzione. Il problema di stima è particolarmente semplice in questo caso perché dobbiamo solo scegliere fra $p = .25$ oppure $p = 0.75$. Anticipiamo i risultati dell'estrazione del campione. Si danno di seguito gli esiti possibili e le loro probabilità:

Esito: x	0	1	2	3
$f(x; \frac{3}{4})$	$\frac{1}{64}$	$\frac{9}{64}$	$\frac{27}{64}$	$\frac{27}{64}$
$f(x; \frac{1}{4})$	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$

In questo esempio, se trovassimo $x = 0$ su un campione di 3 estrazioni, sarebbe da preferire la stima $p = 0.25$ rispetto a $p = 0.75$, perché la probabilità 27/64 è maggiore di 1/64, o con altre parole, perché un campione con $x = 0$ è estratto più probabilmente (nel senso di avere una maggiore probabilità di essere osservato) da una popolazione con $p = 1/4$ piuttosto che da una con $p = 3/4$. Prendendo questo criterio di "massima probabilità" nella scelta di p, in generale dovremo stimare $p = 0.25$ quando $x = 0$ o $x = 1$ e $p = 0.75$ quando $x = 2$ o $x = 3$. Lo stimatore può quindi essere definito come

$$\widehat{p} = \widehat{p}(x) = \begin{cases} 0.25 & \text{quando osserviamo } x = 0, 1 \\ 0.75 & \text{quando osserviamo } x = 2, 3 \end{cases}.$$

Lo stimatore "decide" il valore di p, notato $\widehat{p}$, per ogni possibile x, in modo tale che

$$f(x; \widehat{p}) > f(x; p)\ \forall\, p \neq \widehat{p}.$$

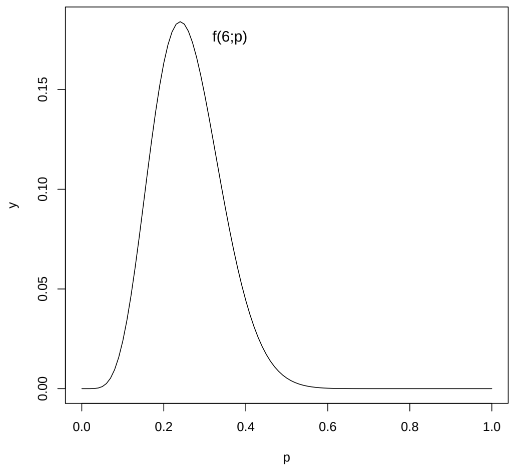


Figura 4.3: Funzione di verosimiglianza con $x = 6$ e $n = 25$.

Più in generale, se fossero possibili altri valori alternativi di p, potremmo ragionevolmente procedere nella stessa maniera. Così se noi trovassimo $x = 6$ in un campione di ampiezza 25 estratto da una popolazione binomiale, dovremmo sostituire tutti i possibili valori di p nell'espressione

$$f(6;p) = C_{25,6}\ p^6(1-p)^{19} \qquad \text{per } 0 \leq p \leq 1. \tag{4.13}$$

Possiamo calcolare agevolmente il punto di massimo della funzione (4.13) ponendo la sua derivata rispetto a p uguale a 0 e risolvendo l'equazione risultante. È immediato verificare che

$$\frac{\partial}{\partial p} f(6;p) = C_{25,6}\ p^5(1-p)^{18}\left[6(1-p) - 19p\right].$$

Ponendo $\frac{\partial}{\partial p} f(6;p) = 0$ e risolvendo rispetto a p troviamo le radici $p = 0, 1, 6/25$. Le prime due radici danno un minimo della funzione e quindi vanno scartate. Quindi la nostra stima sarà $\widehat{p} = 6/25$. Per costruzione dunque essa soddisfa la condizione

$$f(6;\widehat{p}) > f(6;p)$$

per qualsiasi altro valore $p \neq \widehat{p}$ nell'intervallo[5] $[0, 1]$.

[5] L'intervallo $[0, 1]$corrisponde all'insieme Θ.

L'esempio precedente (vedi Figura 4.3) mostra quindi che il valore stimato del parametro sconosciuto è stato ottenuto massimizzando, per il dato valore osservato x, la funzione di probabilità $f(x;p)$ rispetto al parametro p. La definizione generale di stimatore di massima verosimiglianza necessita dapprima la definizione di funzione di verosimiglianza.

Definizione 25. Funzione di verosimiglianza. Si dice funzione di verosimiglianza di n variabili casuali $X_1, ..., X_n$ la densità congiunta delle n variabili casuali, $f_{X_1,...,X_n}(x_1, ..., x_n; \theta)$, *considerata come funzione di* θ. In particolare, se $X_1, ..., X_n$ è un n-campione estratto dalla popolazione X di densità $f(x;\theta)$, la funzione di verosimiglianza è semplicemente il prodotto delle densità $f(x_1;\theta)\ f(x_2;\theta)\ ...\ f(x_n;\theta)$.

Osservazione 20. Per ricordarci che la funzione di verosimiglianza è una funzione di θ, utilizzeremo la notazione $L(\theta; x_1, ..., x_n)$.

La funzione di verosimiglianza $L(\theta; x_1, ..., x_n)$ fornisce la "verosimiglianza" che le variabili aleatorie $X_1, ..., X_n$ assumano il particolare valore osservato $(x_1, ..., x_n)$. La *verosimiglianza* è dunque il valore della funzione di densità congiunta valutata nell'esito $(x_1, ..., x_n)$; per variabili aleatorie discrete essa corrisponde a una probabilità. Osservati $(x_1, ..., x_n)$ la funzione di verosimiglianza dipende unicamente da θ, il parametro sconosciuto. Se la funzione $L(\theta; x_1, ..., x_n)$ possiede un massimo, tale massimo verrà indicato con $\widehat{\theta}$. Vale dunque che

$$\widehat{\theta} = \arg\max_{\theta\in\Theta} L(\theta; x_1, ..., x_n)\ . \tag{4.14}$$

$\widehat{\theta}$ è soluzione di un problema di massimizzazione! Il valore di $\widehat{\theta}$ dipende ovviamente dai valori osservati $(x_1, ..., x_n)$. Differenti valori di $(x_1, ..., x_n)$ daranno generalmente luogo a stime diverse. Esplicitiamo questa dipendenza scrivendo $\widehat{\theta} = T_{mv}(x_1, ..., x_n)$. $T_{mv}(x_1, ..., x_n)$ è dunque la funzione che associa alla realizzazione $(x_1, ..., x_n)$ di $X_1, ..., X_n$ il valore stimato del parametro θ. La funzione T_{mv} non dipende da alcun parametro sconosciuto e per questo motivo $T_{mv}(X_1, ..., X_n)$ definisce una statistica che prende il nome di *stimatore di massima verosimiglianza*.

Definizione 26. Stimatore di massima verosimiglianza. Sia

$$L(\theta) := L(\theta; x_1, ..., x_n) \tag{4.15}$$

la funzione di verosimiglianza delle variabili casuali $X_1, X_2, ..., X_n$. Se $\widehat{\theta}$ (dove $\widehat{\theta} = T_{mv}(x_1, ..., x_n)$ è una funzione delle osservazioni $(x_1, ..., x_n)$) è il valore di θ in Θ che massimizza $L(\theta)$, allora $T_{mv}(X_1, ..., X_n)$ è lo *stimatore di*

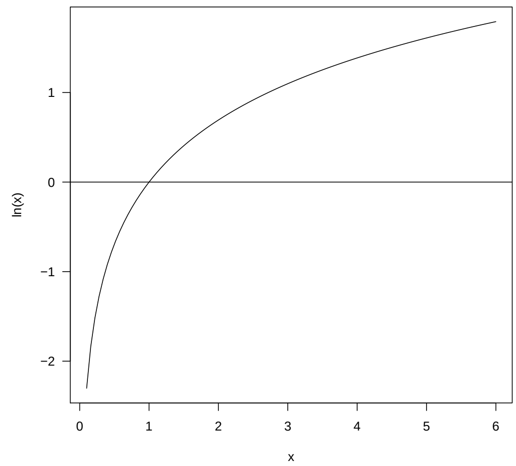


Figura 4.4: Grafico del logaritmo naturale.

massima verosimiglianza di θ. $\widehat{\theta} = T_{mv}(x_1, ..., x_n)$ è la *stima* di massima verosimiglianza di θ sulla base della realizzazione $(x_1, ..., x_n)$ dell'n-campione $X_1, X_2, ..., X_n$.

Come già detto nei capitoli precedenti in questo corso tratteremo esclusivamente il caso in cui le variabili casuali $X_1, ..., X_n$ costituiscono un n-campione, ovvero quando esse formano un campione casuale semplice estratto con reimmissione da una popolazione X avente funzione di densità $f(x; \theta)$. Le variabili casuali $X_1, ..., X_n$ saranno quindi indipendenti fra loro ed identicamente distribuite. Come già osservato nella definizione (25) di funzione di verosimiglianza, grazie all'indipendenza delle $V.A.$ $X_1, ..., X_n$ la funzione di verosimiglianza si riduce al seguente prodotto di densità

$$L(\theta) = f(x_1; \theta)\ f(x_2; \theta)\ ...\ f(x_n; \theta).$$

Se la funzione di verosimiglianza $L(\theta)$ soddisfa determinate *condizioni di regolarità*, lo stimatore di massima verosimiglianza sarà soluzione dell'equazione

$$\frac{\partial}{\partial \theta} L(\theta) = 0.$$

Ricordiamo che il logaritmo naturale $ln()$ è una funzione monotona crescente (cf. il suo grafico, Figura 4.4). Questo significa che gli estremi della funzione di verosimiglianza $L(\theta)$ e della cosiddetta *log-verosimiglianza* $ln(L(\theta))$ sono

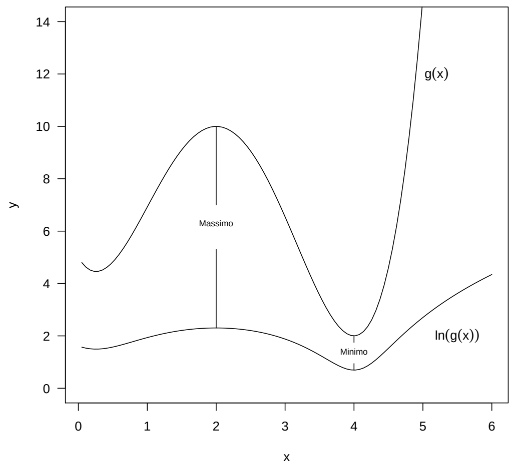


Figura 4.5: Generica funzione g e suo logaritmo.

gli stessi. Pertanto è indifferente ai fini della massimizzazione della funzione di verosimiglianza massimizzare $L(\theta)$ oppure il suo logaritmo $ln(L(\theta))$. I punti di massimo e minimo delle due funzioni sono identici (Figura 4.5).

Quando il numero di parametri sconosciuti è k, la funzione di verosimiglianza dipende da $\theta_1, \theta_2, ...\theta_k$ e scriveremo

$$L(\theta_1, \theta_2, ..., \theta_k) = \prod_{i=1}^{n} f(x_i; \theta_1, \theta_2, ..., \theta_k).$$

La funzione di verosimiglianza dovrà essere massimizzata rispetto ai k parametri. Come per la stima di un singolo parametro, il massimo della funzione di verosimiglianza sarà soluzione del sistema di equazioni

$$\begin{cases} \frac{\partial}{\partial\theta_1} L(\theta_1, \theta_2, ..., \theta_k) = 0 \\ \frac{\partial}{\partial\theta_2} L(\theta_1, \theta_2, ..., \theta_k) = 0 \\ \vdots \\ \frac{\partial}{\partial\theta_k} L(\theta_1, \theta_2, ..., \theta_k) = 0 \end{cases}$$

Qualora invece decidessimo di massimizzare il logaritmo naturale di $L(\theta_1, \theta_2, ..., \theta_k)$ avremo

$$\begin{cases} \frac{\partial}{\partial\theta_1} ln\left(L(\theta_1, \theta_2, ..., \theta_k)\right) = 0 \\ \frac{\partial}{\partial\theta_2} ln\left(L(\theta_1, \theta_2, ..., \theta_k)\right) = 0 \\ \vdots \\ \frac{\partial}{\partial\theta_k} ln\left(L(\theta_1, \theta_2, ..., \theta_k)\right) = 0 \end{cases}$$

In entrambi i casi indicheremo con $\widehat{\theta_i} = T_{mv,i}(x_1, ..., x_n)$, $i = 1, ..., k$ i valori che massimizzano la funzione di verosimiglianza $L(\theta_1, \theta_2, ..., \theta_k)$.

Esempio 37. Supponiamo di estrarre con reimmissione un campione casuale semplice di numerosità n da un'urna contenente una certa frazione p di palline rosse e $q = (1 - p)$ di palline bianche. Sia X definita da

$$X = \begin{cases} 1 & \text{se la pallina estratta è rossa,} \\ 0 & \text{altrimenti.} \end{cases}$$

X è dunque distribuita secondo la distribuzione di Bernoulli. La famiglia parametrica $\mathcal{P}$ è data da

$$\mathcal{P} = \left\{ f(x;p) = \begin{cases} p & \text{se } x = 1, \\ 1-p & \text{se } x = 0, \\ 0 & \text{altrimenti.} \end{cases}, \ p \in [0,1] \right\}.$$

Le realizzazioni campionarie $(x_1, ..., x_n)$ di $X_1, ..., X_n$ formeranno una successione di 0 e di 1. Poiché x_1 sarà uguale a 0 o 1, varrà la seguente uguaglianza

$$f(x_1; p) = p^{x_1} q^{1-x_1}. \tag{4.16}$$

Infatti, definendo $q = 1 - p$, se $x_1 = 1$, $f(1;p) = p^1 q^{1-1} = p$ mentre se $x_1 = 0$, $f(0;p) = p^0 q^{1-0} = q$. Poiché le n variabili aleatorie $X_1, ..., X_n$ sono indipendenti, la funzione di verosimiglianza è il prodotto delle n densità per cui grazie alla (4.16) otteniamo

$$\begin{aligned} L(p) &= \prod_{i=1}^{n} f(x_i; p) = \prod_{i=1}^{n} p^{x_i} q^{1-x_i} \\ &= p^{\sum x_i} q^{n - \sum x_i}. \end{aligned}$$

Ponendo $y = \sum_{i=1}^{n} x_i$ otteniamo[6] che la log-verosimiglianza è uguale a

$$ln\,(L(p)) = yln(p) + (n-y)ln(q).$$

Volendo massimizzare la funzione di log-verosimiglianza rispetto al parametro sconosciuto p poniamo la sua derivata prima uguale a zero:

$$\frac{\partial ln\,(L(p))}{\partial p} = \frac{y}{p} - \frac{n-y}{1-p} \stackrel{!}{=} 0.$$

[6] y corrisponde quindi al numero di palline rosse estratte.

Risolvendo quest'ultima equazione rispetto a p otteniamo lo stimatore di massima verosimiglianza

$$\widehat{p} = \frac{y}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}.$$

Verifichiamo che il punto estremo così calcolato sia effettivamente un punto di massimo:

$$\begin{aligned}\frac{\partial^2 ln\left(L(p)\right)}{\partial p^2} &= -\frac{y}{p^2} - \frac{n-y}{(1-p)^2}\\ &= ...\\ &= \frac{-p^2 + 2\overline{x}p - \overline{x}}{np^2\left(1-p\right)^2}\end{aligned}$$

È immediato verificare che la derivata seconda nel punto $p = \widehat{p} = \overline{x}$ è uguale a $\overline{x}^2 - \overline{x} < 0$. In $\widehat{p}$ la funzione di verosimiglianza ha dunque un massimo.

Osservazione 21. In questo caso lo stimatore di massima verosimiglianza di p coincide con la media campionaria che intuitivamente dovrebbe essere la stima di questo parametro.

Esempio 38. Due parametri sconosciuti. Supponiamo di estrarre un campione casuale semplice di numerosità $n = 4$ da una popolazione Normale $N(\mu_0, \sigma_0^2)$ con entrambi μ_0 e σ_0^2 sconosciuti. I valori osservati di $X_1, ..., X_4$ sono $x_1 = 3$, $x_2 = -3$, $x_3 = 5$, $x_4 = -2$. La famiglia parametrica è uguale a

$$\mathcal{P} = \left\{ f(x;\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\left(x-\mu\right)^2\right),\ \theta = (\mu, \sigma^2) \in \mathbb{R}\times\mathbb{R}^+ \right\}.$$

In questo esempio la funzione di verosimiglianza $L(\mu, \sigma^2)$ è data ancora una volta dal prodotto delle 4 densità

$$\begin{aligned}L(\mu,\sigma^2) &= \prod_{i=1}^{4} f(x_i;\mu,\sigma^2)\\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{4/2} \exp\left(-\frac{1}{2\sigma^2}\left[(3-\mu)^2 + (-3-\mu)^2 + (5-\mu)^2 + (-2-\mu)^2\right]\right)\end{aligned}$$

Come già visto nell'Esempio 37 passando alla log-verosimiglianza si ottiene un'espressione più accessibile

$$\begin{aligned}ln\left(L(\mu,\sigma^2)\right) &= -2\ln(2\pi) - 2\ln(\sigma^2) +\\ &\quad -\frac{1}{2\sigma^2}\left[(3-\mu)^2 + (-3-\mu)^2 + (5-\mu)^2 + (-2-\mu)^2\right].\end{aligned}$$

Quest'ultima espressione andrà massimizzata in funzione di μ e σ^2.

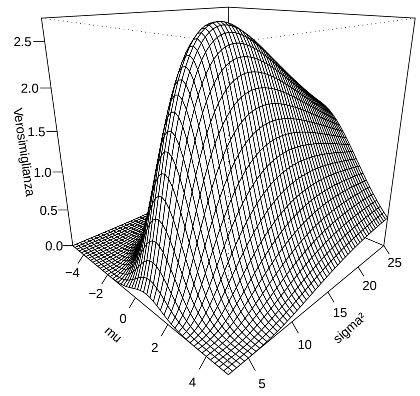


Figura 4.6: Funzione di verosimiglianza $L(\mu, \sigma^2)$ dell'Esempio 38.

Esercizio 10. Calcolate lo stimatore di massima verosimiglianza dei parametri μ_0 e σ_0^2 utilizzando i dati dell'Esempio 38.

4.3 Proprietà degli stimatori puntuali

Nei paragrafi precedenti abbiamo derivato due stimatori puntuali utilizzando il metodo dei momenti ed il metodo di massima verosimiglianza. Essi non sono questi gli unici metodi di stima. Lo stimatore dei minimi quadrati, la cui presentazione è rimandata al corso di Introduzione all'Econometria, potrebbe essere un ulteriore esempio. Vista l'abbondanza di stimatori la domanda che vogliamo affrontare in questo paragrafo è la seguente: "Dati due stimatori esiste un modo per decidere quale dei due sia il migliore?". Per decidere questo è necessario chiarire in che senso uno stimatore è migliore di un altro. A tal fine definiremo alcune proprietà che uno stimatore può possedere oppure no. Esse ci aiuteranno a decidere se e in che misura uno stimatore è migliore di un altro.

Guardando a quanto discusso nei precedenti capitoli, notiamo che nel paragrafo 2.2 abbiamo definito la proprietà di correttezza (cf. Definizione 17) di uno stimatore. Se uno stimatore non è corretto è possibile quantificare il suo grado di distorsione.

Definizione 27. Distorsione (Bias). Sia $T(X_1, ..., X_n)$ uno stimatore di θ. La distorsione di $T(X_1, ..., X_n)$, notata $B(T)$, è definita come

$$B(T) := E\left(T(X_1, ..., X_n)\right) - \theta.$$

Il Bias di uno stimatore corretto è ovviamente 0. Per campioni finiti il bias di uno stimatore $T(X_1, ..., X_n)$ può essere funzione del numero n di osservazioni. Lo stimatore dei momenti $\widehat{\sigma}^2$ della varianza di una distribuzione $N(\mu, \sigma^2)$, μ e σ^2 entrambi sconosciuti, derivato nell'Esempio 36 lo dimostra. Nel medesimo esempio abbiamo studiato il comportamento asintotico (cioè quando $n \to \infty$) della distorsione. Lo studio del comportamento asintotico di uno stimatore e delle sue proprietà è molto importante in quanto per n finito la funzione di distribuzione di T_n non è generalmente conosciuta[7]. Per tale motivo si è costretti a confrontare gli stimatori sulla base delle loro *proprietà asintotiche* il cui calcolo è possibile grazie a teoremi quali la legge dei grandi numeri e il teorema del limite centrale (a tal proposito su veda il paragrafo 4.4 dedicato alla teoria asintotica).

La correttezza di uno stimatore è una proprietà importante ma non fondamentale per la determinazione del miglior stimatore. Come è stato visto nelle serie di esercizi esiste in genere più di uno stimatore corretto dello stesso parametro. Per tale motivo la correttezza non è un criterio sufficiente per individuare quale sia lo stimatore migliore fra due o più stimatori dello stesso parametro. Per trovare una soluzione al nostro problema consideriamo la seguente situazione. Sia X_1, X_2 e X_3 un campione aleatorio estratto da una popolazione Normale $N(\mu, 1)$. Sono proposti i seguenti stimatori

$$\begin{array}{ll}
\widehat{\mu}_1 = \frac{X_1+X_2}{2} & \Rightarrow E(\widehat{\mu}_1) = \mu, \\
\widehat{\mu}_2 = \frac{X_1+X_2+X_3}{3} & \Rightarrow E(\widehat{\mu}_2) = \mu, \\
\widehat{\mu}_3 = \frac{X_1+2X_2+X_3}{4} & \Rightarrow E(\widehat{\mu}_3) = \mu, \\
\widehat{\mu}_4 = X_1 & \Rightarrow E(\widehat{\mu}_4) = \mu, \\
\widehat{\mu}_5 = \frac{1}{9}\left(X_1 + X_2 + X_3\right) & \Rightarrow E(\widehat{\mu}_5) = \frac{1}{3}\mu.
\end{array}$$

I primi quattro stimatori sono tutti stimatori corretti di μ. L'ultimo per contro è distorto, con un termine di distorsione pari a $B(\widehat{\mu}_5) = \frac{1}{3}\mu - \mu = -\frac{2}{3}\mu$. Qual è dunque il migliore? Notiamo che i cinque stimatori hanno varianze diverse. Infatti

$$V\left(\widehat{\mu}_1\right) = \frac{1}{2},\ V\left(\widehat{\mu}_2\right) = \frac{1}{3},\ V\left(\widehat{\mu}_3\right) = \frac{3}{8},\ V\left(\widehat{\mu}_4\right) = 1,\ V\left(\widehat{\mu}_5\right) = \frac{1}{27}.$$

[7]Finora avete incontrato stimatori relativamente semplici quali la media campionaria $\overline{X}_n$ o S^2, per i quali è possibile calcolare i momenti pur non conoscendo la famiglia parametrica di densità utilizzata. Tuttavia essi sono dei casi particolari. In generale anche il semplice calcolo di $E(T(X_1, ..., X_n))$ può risultare assai complesso o addirittura impossibile.

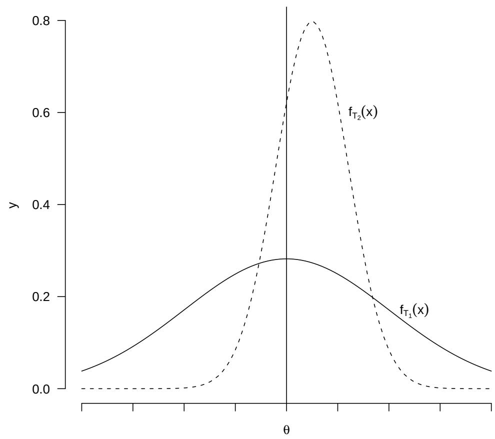


Figura 4.7: Confronto fra due stimatori

Fra i primi quattro stimatori corretti quello con la varianza minore è $\widehat{\mu}_2$. Per contro $\widehat{\mu}_5$ è distorto, ma con una varianza in assoluto inferiore, anche rispetto a $\widehat{\mu}_2$. Se decidessimo di restringere la ricerca dello stimatore migliore all'interno della classe di stimatori corretti, ovviamente $\widehat{\mu}_2$ sarebbe quello da preferire in quanto è il più preciso in termini di varianza. Tuttavia se estendessimo la ricerca anche a stimatori non necessariamente corretti, dovremmo trovare un criterio che ci permetta di misurare il "trade-off" tra la perdita in precisione causata dalla distorsione ed il guadagno in precisione dovuto ad una varianza inferiore. Tale criterio ci è dato dal cosiddetto *Errore Quadratico Medio (Mean Squared Error).*

Definizione 28. Errore Quadratico Medio (EQM). Sia $T(X_1, ..., X_n)$ uno stimatore di θ. L'Errore Quadratico Medio dello stimatore $T(X_1, ..., X_n)$ di θ è definito da

$$EQM_T = E\left((T-\theta)^2\right).$$

Il termine "errore quadratico medio" può essere giustificato se si pensa alla differenza $T-\theta$ come all'errore che si compie nella stima di θ. $E\left((T-\theta)^2\right)$ è dunque una misura di dispersione dei valori dello stimatore T attorno al *vero* valore θ. Se noi confrontassimo gli stimatori guardando i loro rispettivi errori quadratici medi, ne preferiremmo naturalmente uno con errore quadratico medio piccolo.

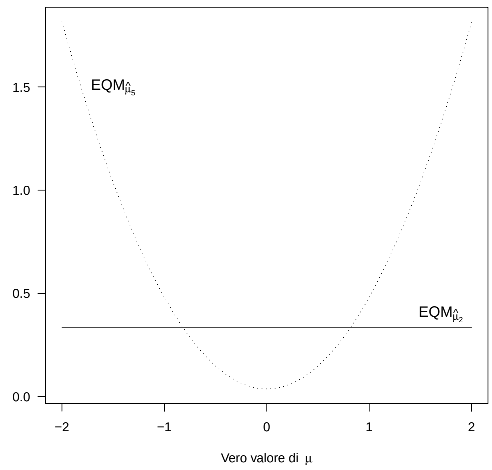


Figura 4.8: Confronto EQM $\hat{\mu}_2$ e $\hat{\mu}_5$.

Teorema 5. Sia $T(X_1, ..., X_n)$ uno stimatore di θ. Vale la seguente identità:

$$\begin{aligned} EQM_T &= \underbrace{V(T)}_{\text{Varianza di } T} + \underbrace{(E(T)-\theta)^2}_{\text{quadrato della distorsione}} \\ &= V(T) + (B(\theta))^2 . \end{aligned} \tag{4.17}$$

Osservazione 22. Per uno stimatore corretto T di θ avremo che il secondo termine è zero e quindi varrà semplicemente $EQM_T = V(T)$.

Tornando ai due stimatori $\widehat{\mu}_2$ e $\widehat{\mu}_5$ possiamo ora calcolare il rispettivo EQM. Poiché $\widehat{\mu}_2$ è uno stimatore corretto vale che

$$EQM_{\widehat{\mu}_2} = V(\widehat{\mu}_2) = \frac{1}{3}.$$

Per quanto riguarda $\widehat{\mu}_5$ abbiamo che

$$EQM_{\widehat{\mu}_5} = V(\widehat{\mu}_5) + \left(-\frac{2}{3}\mu\right)^2 = \frac{1}{27} + \frac{4}{9}\mu^2.$$

Lo stimatore $\widehat{\mu}_5$, pur essendo uno stimatore distorto di μ, è superiore allo stimatore $\widehat{\mu}_2$ in media quadratica per valori del parametro μ compresi nell'intervallo $(-\sqrt{2/3}, \sqrt{2/3})$ (confronta Figura 4.8).

4.4 Elementi di teoria asintotica

In questo paragrafo consideriamo due importanti caratteristiche di uno stimatore $T(X_1, ..., X_n)$ di θ. La prima è la consistenza, definita attraverso la convergenza in probabilità di $T(X_1, ..., X_n)$ a θ mentre la seconda è la normalità asintotica della quantità $\sqrt{n}\,(T(X_1, ..., X_n) - \theta)$. Entrambi i concetti sono già stati incontrati in precedenza: il primo ad esempio quando è stata trattata la convergenza dei momenti campionari $M_{r,n}$ ai rispettivi momenti teorici M_r, il secondo quando abbiamo considerato la distribuzione limite della media standardizzata $\frac{(\overline{X}_n-\mu)}{\sqrt{\sigma^2/n}}$.

4.4.1 Consistenza

La consistenza dello stimatore $T(X_1, ..., X_n)$ di θ è definita come segue.

Definizione 29. Consistenza. Sia $T(X_1, ..., X_n)$ uno stimatore del parametro sconosciuto θ. Diremo che T è uno stimatore consistente di θ se, per qualsiasi $\epsilon > 0$, T soddisfa la seguente condizione

$$\lim_{n\to\infty} P\left(T(X_1, ..., X_n) \in (\theta - \epsilon, \theta + \epsilon)\right) = 1. \tag{4.18}$$

Osservazione 23. Uno stimatore che soddisfa la (4.18) si dice convergente in probabilità a θ e si scrive $T(X_1, ..., X_n) \underset{n\to\infty}{\overset{P}{\to}} \theta$.

La consistenza di uno stimatore è dunque una proprietà asintotica. Al crescere dell'informazione disponibile la precisione della stima, misurata tramite la probabilità di trovarsi in prossimità del vero valore, deve essere totale e cioè la probabilità deve tendere a 1. Uno stimatore consistente ci garantisce che all'aumentare dell'informazione (numero di osservazioni) il valore stimato si avvicinerà con probabilità crescente al vero valore.

La Figura 4.9 mostra, dato uno stimatore $\hat{\theta}_n$ di θ avente funzione di densità continua, quello che accade al crescere della numerosità del campione. Con $n \to \infty$ la probabilità tende a concentrarsi sempre più attorno al vero valore θ.

Potremmo chiederci quale sia la relazione tra la consistenza di uno stimatore e la sua varianza. Per introdurre questa discussione presentiamo la presente definizione.

Definizione 30. Convergenza in media quadratica. Lo stimatore $T(X_1, ..., X_n)$ del parametro sconosciuto θ è convergente (o consistente) in media quadratica se

$$\lim_{n\to\infty} EQM_T = E((T-\theta)^2) = 0.$$

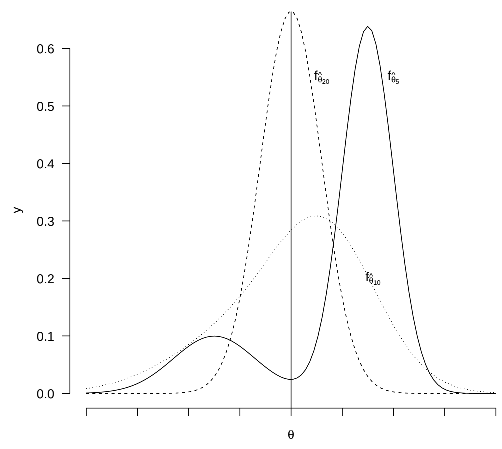


Figura 4.9: Convergenza in probabilità di $\hat{\theta}_n = T(X_1, \ldots, X_n)$.

Osservazione 24. Poiché, come visto nel teorema 28 vale che

$$EQM_T = V(T) + (B(\theta))^2 \ ,$$

lo stimatore T sarà convergente in media quadratica *se e solo se* sia la varianza che la sua distorsione tendono a 0 al crescere di n all'infinito. In questo modo abbiamo un criterio molto semplice per verificare la convergenza in media quadratica di T.

Possiamo ora caratterizzare la relazione che lega la consistenza di uno stimatore alla sua varianza.

Teorema 6. Se uno stimatore T di θ è consistente in media quadratica, esso sarà pure uno stimatore consistente di θ (non è necessariamente vero l'inverso).

Il precedente teorema è particolarmente utile in quanto permette di trovare un criterio semplice per verificare la consistenza di uno stimatore T di θ. Infatti, quando la distorsione e la varianza di T tendono entrambi a 0 al crescere di n, allora lo stimatore è consistente. Ovviamente per uno stimatore corretto è sufficiente che la sua varianza tenda a 0 al crescere di n essendo la sua distorsione uguale a 0.

4.4.2 Normalità asintotica

Il teorema 3 del limite centrale afferma che la media $\overline{X}_n$ delle n variabili aleatorie $X_1, ..., X_n$ opportunamente standardizzata converge per $n \to \infty$ in distribuzione ad una variabile aleatoria $Z \sim N(0,1)$

$$\frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \underset{n\to\infty}{\sim} Z. \tag{4.19}$$

Il lato sinistro della (4.19) che chiameremo Y_n può essere riscritto come

$$Y_n = \frac{\sqrt{n}\left(\overline{X}_n - \mu\right)}{\sigma}.$$

Una proprietà della convergenza in distribuzione è quella che se

$$Y_n \underset{n\to\infty}{\sim} Z$$

e noi moltiplichiamo Y_n per un numero, diciamo c, allora vale che

$$cY_n \underset{n\to\infty}{\sim} cZ.$$

Questa proprietà ci permette di studiare il comportamento asintotico della quantità $\sqrt{n}\left(\overline{X}_n - \mu\right)$. Infatti, per quanto appena detto avremo che

$$\sqrt{n}\left(\overline{X}_n - \mu\right) = \sigma\frac{\sqrt{n}\left(\overline{X}_n - \mu\right)}{\sigma} = \sigma Y_n \underset{n\to\infty}{\sim} \sigma Z.$$

Ma poiché Z è $N(0,1)$, σZ avrà una distribuzione $N(0,\sigma^2)$. Abbiamo dunque dimostrato che

$$\sqrt{n}\left(\overline{X}_n - \mu\right) \underset{n\to\infty}{\sim} N(0,\sigma^2). \tag{4.20}$$

Le due formule (4.19) e (4.20) sono del tutto equivalenti. Tuttavia la formula (4.20) è quella generalmente utilizzata per presentare la distribuzione asintotica dello stimatore di massima verosimiglianza (dopo le opportune modifiche di notazione).

Teorema 7. Distribuzione asintotica dello stimatore di massima verosimiglianza. Sia $X_1, ..., X_n$ un campione *i.i.d.* estratto da una popolazione X di densità $f(x;\theta)$. Sia $\widehat{\theta}_n := T_{mv}(X_1, ..., X_n)$ lo stimatore di massima verosimiglianza di θ con

$$\sum_{i=1}^{n} \frac{\partial ln\left(f(x_i;\theta)\right)}{\partial\theta} \mid_{\theta=\widehat{\theta}_n} = 0.$$

Allora vale che

1. Lo stimatore di massima verosimiglianza è consistente per il parametro θ. Vale dunque

$$\widehat{\theta}_n \underset{n\to\infty}{\overset{P}{\to}} \theta.$$

2. $\sqrt{n}\left(\widehat{\theta}_n-\theta\right)$ è asintoticamente Normale

$$\sqrt{n}\left(\widehat{\theta}_n-\theta\right) \underset{n\to\infty}{\sim} N(0,\frac{1}{I(\theta)})$$

dove $I(\theta) := E\left[\left(\frac{\partial ln(f(X;\theta))}{\partial\theta}\right)^2\right]$ è chiamata l'informazione di Fisher.

3. Alternativamente $I(\theta)$ può essere calcolata nel seguente modo

$$I(\theta) = -E\left(\frac{\partial^2 ln\left(f(X;\theta)\right)}{\partial\theta^2}\right).$$

Osservazione 25. Per il calcolo di $I(\theta)$ la funzione $ln\left(f(x;\theta)\right)$ è derivata rispetto a θ (la x è considerata costante!). La stessa cosa vale per la derivata seconda $\partial^2 ln\left(f(x;\theta)\right)/\partial\theta^2$.

Quest'ultimo teorema è particolarmente importante in quanto stabilisce che lo stimatore di massima verosimiglianza è

- uno stimatore consistente di θ;
- distribuito asintoticamente secondo la legge Normale.

Come nel caso della media campionaria $\overline{X}_n$, la distribuzione asintotica di $\sqrt{n}\left(\widehat{\theta}_n-\theta\right)$ è anch'essa Normale. La varianza della distribuzione asintotica di $\sqrt{n}\left(\widehat{\theta}_n-\theta\right)$ dipende da un'espressione conosciuta col nome di *informazione di Fisher*: il valore atteso del quadrato di $\frac{\partial ln(f(X;\theta))}{\partial\theta}$. Infine osserviamo che l'informazione di Fisher, notata $I(\theta)$, è altresì calcolabile per mezzo del valore atteso della seconda derivata di $ln\left(f(X;\theta)\right)$.

Esempio 39. Supponiamo di estrarre un campione casuale semplice di numerosità n da una popolazione Normale $N(\mu,\sigma^2)$ con solo μ sconosciuto. La famiglia parametrica è uguale a (σ^2 poiché conosciuto è ora da considerarsi alla stregua di un numero)

$$\mathcal{P} = \left\{f(x;\mu) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}\left(x-\mu\right)^2\right),\ \mu\in\mathbb{R}\right\}.$$

La funzione di massima verosimiglianza $L(\mu)$ è data dal prodotto delle n densità

$$\begin{aligned} L(\mu) &= \prod_{i=1}^{n} f(x_i;\mu) \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i-\mu\right)^2\right). \end{aligned}$$

Lo stimatore di massima verosimiglianza di μ sappiamo essere uguale alla media campionaria $\overline{X}_n$. Inoltre, poiché la somma di variabili aleatorie normali indipendenti è ancora Normale (è questa una particolarità della distribuzione Normale), la distribuzione di $\overline{X}_n$ è $N(\mu, \frac{\sigma^2}{n})$. La distribuzione di $\sqrt{n}\left(\overline{X}_n - \mu\right)$ è dunque $N(0, \sigma^2)$. Se il Teorema 7 è corretto, devono valere le seguenti tre uguaglianze:

$$\sigma^2 = E\left[\left(\frac{\partial ln\left(f(X;\mu)\right)}{\partial \mu}\right)^2\right]^{-1} = -\left[E\left(\frac{\partial^2 ln\left(f(X;\mu)\right)}{\partial \mu^2}\right)\right]^{-1}. \qquad (4.21)$$

Esercizio 11. Quale esercizio verificate voi stessi la correttezza della (4.21) nel caso in cui X è distribuita secondo la legge Normale $N(\mu, \sigma^2)$. Procedete come segue:

1. Calcolate la derivata $\frac{\partial ln(f(X;\mu))}{\partial \mu}$, il suo quadrato ed infine il valore atteso del suo quadrato, notato $E\left[\left(\frac{\partial ln(f(X;\mu))}{\partial \mu}\right)^2\right]$.

2. Calcolate la derivata seconda $\partial^2 ln\left(f(X;\mu)\right)/\partial \mu^2$ ed in seguito il suo valore atteso.

3. Confrontate i risultati ottenuti con la formula (4.21).

4.4.3 Disuguaglianza di Cramèr-Rao

La disuguaglianza di Cramèr-Rao stabilisce un *limite inferiore* per quanto riguarda la varianza di un *qualsiasi* stimatore corretto del parametro θ. In altre parole, per quanto preciso possa essere lo stimatore in questione la sua varianza non sarà mai inferiore a tale limite.

Teorema 8. Disuguaglianza di Cramèr-Rao. Sia $X_1, ..., X_n$ un n-campione estratto da una popolazione X di densità $f(x;\theta)$ e $T(X_1, ..., X_n)$ uno stimatore corretto di θ. Vale la seguente disuguaglianza

$$V(T(X_1, ..., X_n)) \geq \frac{1}{n\ I(\theta)} \qquad (4.22)$$

dove $I(\theta)$ è l'informazione di Fisher definita nel Teorema 7.

Osservazione 26. Grazie alle proprietà della varianza la disuguaglianza (4.22) può essere riscritta come

$$V\left(\sqrt{n}\left(T(X_1, ..., X_n) - \theta\right)\right) \geq \frac{1}{I(\theta)}.$$

L'utilità di quest'ultimo teorema risiede nel fatto che esso rivela il limite massimo di efficienza nella stima del parametro sconosciuto θ. Come detto precedentemente, per n finito, può non essere possibile calcolare la varianza dello stimatore T e ovviamente non sarà neppure possibile verificare se il limite inferiore dato nella (4.22) è raggiunto[8] o meno. In tal caso si è soliti guardare alla varianza della distribuzione asintotica di

$$\sqrt{n}\,(T(X_1, ..., X_n) - \theta)$$

che nella maggior parte dei casi che incontrerete in econometria[9] sarà $N(0, \upsilon^2)$. Quando $T(X_1, ..., X_n)$ è lo stimatore di massima verosimiglianza di θ, il Teorema 7 afferma che υ^2 è $\frac{1}{I(\theta)}$. Grazie all'osservazione 26 possiamo quindi concludere che lo stimatore di massima verosimiglianza raggiunge *asintoticamente* il limite inferiore di Cramèr-Rao.

È possibile generalizzare la disuguaglianza di Cramèr-Rao anche al caso di uno stimatore T non corretto di θ. In tal caso il teorema 8 va modificato come segue.

Teorema 9. Disuguaglianza di Cramèr-Rao, caso generale. Sia $X_1, ..., X_n$ un campione *i.i.d.* (un n-campione) estratto da una popolazione X di densità $f(x;\theta)$ e $T(X_1, ..., X_n)$ uno stimatore di θ la cui distorsione è uguale a $B(\theta)$. Vale la seguente disuguaglianza

$$V(T(X_1, ..., X_n)) \geq \frac{(1 + \frac{\partial B(\theta)}{\partial \theta})^2}{n\ I(\theta)} \tag{4.23}$$

dove $I(\theta)$ è l'informazione di Fisher definita nel Teorema 7.

[8] In generale, per n fisso, il limite di Cramèr-Rao è un limite inferiore non raggiungibile.

[9] Il termine υ^2 corrisponde quindi alla varianza della distribuzione asintotica di

$$\sqrt{n}\,(T(X_1, ..., X_n) - \theta)\,.$$

Capitolo 5

Verifica d'ipotesi

Dopo aver trattato alcuni argomenti propri della teoria della stima affronteremo in questo capitolo il secondo grande tema della statistica inferenziale: la verifica d'ipotesi.

L'induzione statistica e la verifica di ipotesi sono alla base di tutta la ricerca scientifica. In economia così come nelle scienze naturali o in medicina accade spesso che l'oggetto del problema sia quello di capire se, a seguito di un particolare evento o di un intervento pianificato, vi sia un cambiamento nelle caratteristiche della popolazione in termini del suo comportamento e/o attributi. In medicina, ad esempio, si potrebbe voler determinare se un nuovo farmaco o una nuova terapia sia più efficace nella cura di una certa malattia rispetto a un farmaco o trattamento tradizionale. Anche in micro- e macroeconomia ci sono numerosi esempi dove la verifica d'ipotesi può essere utilizzata per appurare empiricamente la correttezza di una teoria economica, l'efficacia di una determinata politica economica o, come trattato nell'Esempio 8, verificare l'efficacia di una campagna pubblicitaria.

La costruzione di un test statistico richiede la definizione della famiglia parametrica $\mathcal{P}$ di densità che si intende utilizzare. In seguito si dovrà definire l'ipotesi che si desidera verificare in termini della distribuzione della popolazione obiettivo o della distribuzione di probabilità della variabile aleatoria in esame. Diamo ora la definizione formale di *ipotesi statistica*. I successivi esempi ne chiariranno il significato.

Definizione 31. Ipotesi statistica. Un'ipotesi statistica è un'affermazione o supposizione sulla distribuzione di una o più variabili casuali.

L'ipotesi statistica soggetta ad esame è chiamata ipotesi nulla ed è generalmente notata col termine "H_0 :" seguito dalla descrizione dell'ipotesi.

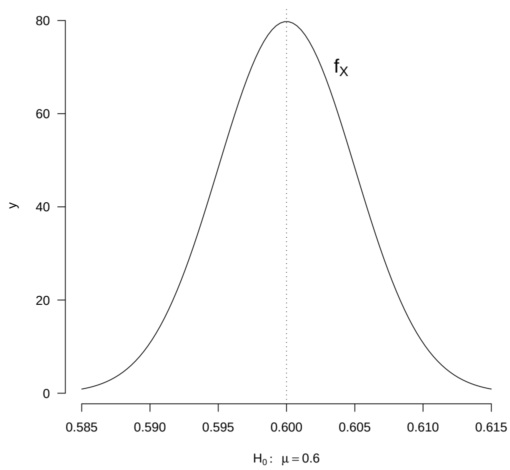


Figura 5.1: Densità della V.A. X sotto H_0 dell'Esempio 40.

Esempio 40. Un produttore di penne a sfera ha acquistato una macchina per la produzione delle sfere da inserire nella punta della penna. Una volta messa a punto, la macchina dovrebbe produrre punte dal diametro di 0.6 millimetri. Per il produttore è particolarmente importante che le sfere prodotte non siano né troppo grandi né troppo piccole. Dall'esperienza maturata in passato si sa che una macchina messa a punto correttamente produce sfere di diametro 0.6 millimetri con una variabilità di $\sigma = 0.005$ millimetri. Sia dunque X la variabile aleatoria che denota la misurazione del diametro di una sfera prodotta: $X \sim N(\mu, (0.005)^2)$. La famiglia parametrica di densità è dunque data da

$$\mathcal{P} = \left\{ f(x;\mu) = \frac{1}{\sqrt{2\pi}0.005} \exp\left(-\frac{1}{2 \cdot 0.005^2} (x-\mu)^2 \right), \ \mu \in \mathbb{R}^+ \right\}.$$

Sotto l'ipotesi che la macchina abbia la giusta messa a punto, μ è conosciuto ed uguale a 0.6. Per contro[1], se la macchina fosse mal funzionante il diametro medio della produzione potrebbe essere superiore o inferiore a 0.6. Poiché la prima ipotesi determina univocamente la distribuzione di X la utilizzeremo quale ipotesi nulla (vedi Figura 5.1):

$$H_0: \ \text{la macchina è tarata correttamente.}$$

[1]Implicitamente si assume che difetti nella taratura della macchina non influenzino la varianza della distribuzione.

L'alternativa all'ipotesi nulla, chiamata *ipotesi alternativa*, verrà indicata con "H_A :". In questo primo esempio essa può essere espressa matematicamente come

$$H_A: \ \mu \neq 0.6.$$

Esempio 41. Un caseificio ha il forte sospetto che un suo fornitore diluisca il latte con acqua. Dalle numerose misurazioni eseguite in passato dal Laboratorio cantonale si sa che la temperatura di gelo di un campione di latte è una variabile aleatoria X distribuita secondo la legge Normale di valore atteso -0.545 °C e deviazione standard 0.008 °C. La famiglia parametrica di densità è ancora Normale

$$\mathcal{P} = \left\{ f(x;\mu) = \frac{1}{\sqrt{2\pi}0.008} \exp\left(-\frac{1}{2 \cdot 0.008^2} (x-\mu)^2 \right), \ \mu \in \mathbb{R} \right\}.$$

Supponiamo che il latte non sia annacquato: in tal caso il valore di μ sarebbe conosciuto ed uguale a -0.545. Sia la distribuzione di X che la distribuzione campionaria $f(x_1, ..., x_n)$ sarebbero conosciuti. Per contro, qualora il latte venisse diluito con acqua, la sua temperatura di gelo sarebbe superiore a quella del latte "puro". Avremmo quindi[2] $\mu > -0.545$. Tuttavia, non conoscendo la percentuale di acqua aggiunta al latte, l'esatto valore di μ non sarebbe calcolabile. Inoltre la legge afferma che "chiunque è presunto innocente fino a quando la sua colpevolezza non sia dimostrata dinanzi alla Corte". Quindi la nostra ipotesi nulla sarà la seguente

$$H_0: \ \text{Il latte non è annacquato} \ (\Leftrightarrow \text{il fornitore è onesto})$$

che in base al nostro ragionamento può essere espressa in termini del parametro μ semplicemente come

$$H_0: \ \mu = -0.545.$$

Non rifiutare H_0 significa scagionare il fornitore dal sospetto di truffa nei confronti del caseificio. Per contro, l'alternativa di H_0 è che il fornitore effettivamente diluisca il latte. In tal caso il valore atteso della temperatura di gelo sarà superiore a -0.545. L'*ipotesi alternativa* H_A sarà dunque espressa come

$$H_A: \ \mu > -0.545.$$

Il caseificio decide di analizzare cinque campioni $(X_1, ..., X_5)$ di latte provenienti dal fornitore indagato. Il grosso vantaggio di scegliere quale ipotesi

[2]Si assume ragionevolmente che l'annacquamento del latte non modifichi σ^2.

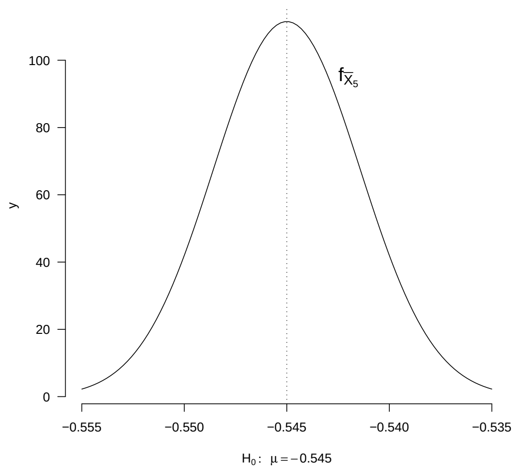


Figura 5.2: Funzione di densità di $\bar{X}_5$ sotto H_0.

nulla $\mu = -0.545$ è dato dal fatto che H_0 identifica in maniera esatta la distribuzione di X (e quindi indirettamente anche di $(X_1, ..., X_5)$ nonché $\bar{X}_5$, vedi Figura 5.2).

Nei due ultimi esempi abbiamo definito l'ipotesi nulla e l'ipotesi alternativa utilizzando le informazioni in nostro possesso parallelamente alle varie assunzioni riguardo la famiglia parametrica $\mathcal{P}$. In entrambi i casi l'ipotesi nulla è tale per cui la distribuzione della variabile aleatoria X è completamente specificata. In generale diremo che un'ipotesi statistica (nulla o alternativa che sia) è *semplice* se specifica completamente la distribuzione di X, come è appunto il caso per le due ipotesi nulle dei due ultimi esempi. In caso contrario diremo invece che l'ipotesi è *composta*. Ad esempio, l'ipotesi alternativa

$$H_A:\ \mu > -0.545$$

dell'Esempio 41 è un'ipotesi composta in quanto non identifica esattamente la distribuzione di X. H_A è l'unione di un'infinità di ipotesi semplici.

5.1 Test per un'ipotesi statistica

Eseguire un test di un'ipotesi statistica significa costruire una regola di decisione basata sulle osservazioni disponibili che ci permetta di concludere se l'ipotesi sotto esame debba essere rifiutata o meno. Vista la natura aleatoria di quanto osservato, tale conclusione varierà a seconda del campione osservato. Diamo innanzi tutto una definizione formale di quanto detto.

Definizione 32. Test per un'ipotesi statistica. Un test di una ipotesi statistica H è una regola o procedimento per decidere se rifiutare o meno H.

Il problema è come determinare tale regola. In pratica l'informazione disponibile è contenuta in un n-campione di osservazioni di X. La decisione dipenderà dai valori $(x_1, ..., x_n) \in \Omega^n$ che osserveremo: se quanto osservato sarà "in sintonia" con l'ipotesi non la rifiuteremo. Le osservazioni forniranno evidenza pro o contro l'ipotesi. Tale evidenza sarà generalmente misurata in termini di probabilità. Illustriamo quanto appena detto sviluppando i due esempi precedenti.

Esempio 42. Continuazione dell'Esempio 41. Il caseificio decide di calcolare la temperatura di gelo di un campione di 5 cartoni di latte provenienti dal fornitore in questione. Potremmo a questo punto domandarci qual è la probabilità, sotto l'ipotesi nulla, di osservare una determinata realizzazione $(x_1, ..., x_5)$. Tuttavia questa domanda ha poco senso: la distribuzione di $(X_1, ..., X_5)$ è una distribuzione continua e noi sappiamo che tale probabilità è 0. Per contro, anziché lavorare direttamente con i valori del campione per valutare l'attendibilità dell'ipotesi nulla, utilizziamo una statistica S, la media campionaria $\overline{X}$ ad esempio. Qual è il vantaggio di un simile approccio? Il vantaggio consiste nel fatto che lavorare con una statistica riduce la complessità del problema. Infatti, per decidere se rifiutare o non rifiutare H_0 è sufficiente guardare ad un unico valore anziché all'intero campione, che in questo esempio ha dimensione 5. Calcoliamo la media delle cinque osservazioni: $\overline{x} = -0.536$. Grazie alle proprietà della distribuzione Normale sappiamo che $\overline{X} \sim N(\mu, \frac{0.008^2}{5})$. Sotto l'ipotesi H_0 la distribuzione della statistica $\overline{X}$ è dunque $N(-0.545, \frac{0.008^2}{5})$.

A questo punto è possibile calcolare la probabilità di osservare un valore di $\overline{X}$ uguale o superiore a quello calcolato dal caseificio (vedi Figura 5.3):

$$\begin{aligned} P_{H_0}(\overline{X} \geq -0.536) &= P_{H_0}(\overline{X} + 0.545 \geq 0.009) \\ &= P_{H_0}(\frac{\sqrt{5}\,(\overline{X} + 0.545)}{0.008} \geq 2.515) \\ &= 0.59\%. \end{aligned}$$

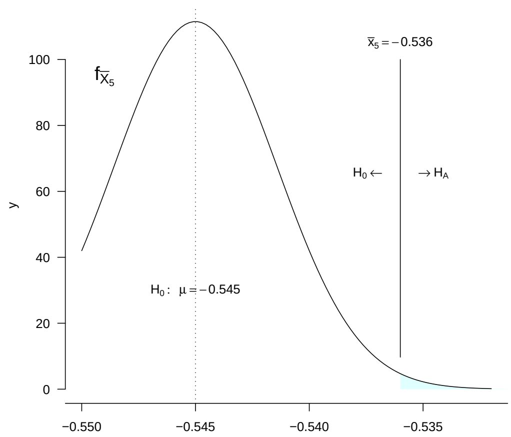


Figura 5.3: Calcolo del p-value (test unilaterale destro).

La probabilità è dunque 0.59%. Quindi se il latte fosse “puro” registreremmo una media $\overline{x}$ uguale o superiore a quanto osservato meno di una volta su cento. Ci troviamo di fronte ad un evento raro in quanto la sua probabilità è bassa. Una probabilità dello 0.59% genera evidenza contro l'ipotesi di onestà del fornitore. Si potrebbe dunque decidere di rifiutare l'ipotesi nulla a favore dell'ipotesi alternativa.

Definizione 33. p-value. La *probabilità sotto* H_0 di osservare un valore della statistica S almeno *tanto estremo* quanto il valore osservato è detta p-value.

Il p-value è dunque la probabilità di un particolare evento e costituisce un indice di evidenza empirica a favore dell'ipotesi nulla H_0. Il particolare evento a cui siamo interessati è costituito dall'osservazione di un valore della statistica S uguale o più estremo di quello effettivamente osservato. Qual è l'interpretazione del termine “più estremo”? Con “più estremo” si intende “più lontano”, “più distante” - rispetto al valore atteso μ sotto H_0 - in direzione dell'*ipotesi alternativa*. Il test statistico parlerà tanto più a favore dell'ipotesi nulla tanto più elevato sarà il p-value.

Calcolato il p-value, quando decideremo se rifiutare o non rifiutare H_0? La decisione finale di rifiutare o meno H_0 dipenderà dalla “quantità” minima di evidenza sotto la quale non siamo più disposti a mantenere H_0 come ipotesi valida. Tale soglia, notata α, è detta *livello di significatività*. Essa può variare a dipendenza delle circostanze. In economia si utilizza generalmente un livello di significatività del 5%. Assumendo dunque un livello di significatività $\alpha = 5\%$, rifiuteremo l'ipotesi nulla se il p-value $< \alpha$. In tal caso diremo che il test è significativo ad un livello di significatività dell' α per cento.

Osservazione 27. È importante sottolineare che a causa della natura aleatoria del fenomeno le nostre conclusioni saranno soggette ad errore. Infatti, la scarsa evidenza a favore di H_0 non significa che H_0 sia falsa. Come avremo modo di approfondire in seguito, il livello di significatività α misura la probabilità dell'errore di rifiutare H_0 quando questa è effettivamente vera.

Esempio 43. Continuazione dell'Esempio 40. Per verificare il corretto funzionamento della macchina il produttore produce un campione *i.i.d.* di $n = 25$ sfere. Come in precedenza utilizziamo quale statistica S la media campionaria $\overline{X}$. Supponiamo che il valore medio osservato sia $\overline{x} = 0.5985$. Sotto l'ipotesi H_0 la distribuzione di $\overline{X}$ è $N(0.6, 0.001^2)$. A differenza dell'Esempio 41, in questo caso l'ipotesi alternativa non ha una direzione ben precisa. H_A afferma che il valore di μ potrebbe essere superiore o inferiore a 0.6. Per questo motivo nel calcolo del p-value si terrà conto della probabilità di osservare un valore di S ($\overline{X}$ in questo caso) inferiore a 0.5985 (più estremo sulla sinistra della distribuzione) o superiore a 0.6015 (più estremo sulla destra della distribuzione). L'estremo destro 0.6015 è stato calcolato in maniera tale da mantenere la distanza $d =\mid \bar{x} - \mu \mid$ tra valore atteso e media campionaria anche alla destra di μ (vedi Figura 5.4):

$$0.6 + \mid 0.5985 - 0.6 \mid = 0.6015.$$

Il p-value è dunque uguale a

$$\begin{aligned} P_{H_0}(\overline{X} \leq 0.5985,\, \overline{X} \geqslant 0.6015) &= P_{H_0}(|\overline{X} - 0.6| \geq 0.0015) \\ &= P_{H_0}(\frac{|\overline{X} - 0.6|}{0.001} \geq 1.5) = 13.4\%. \end{aligned}$$

Decisamente questo non rappresenta un evento raro. Fissato un livello di significatività pari al 5% il p-value risulta essere più grande. Per tale motivo non rifiutiamo l'ipotesi nulla. La macchina è da considerarsi perfettamente funzionante.

Ricapitolando quanto visto nei due esempi precedenti abbiamo i seguenti elementi costitutivi di un test d'ipotesi:

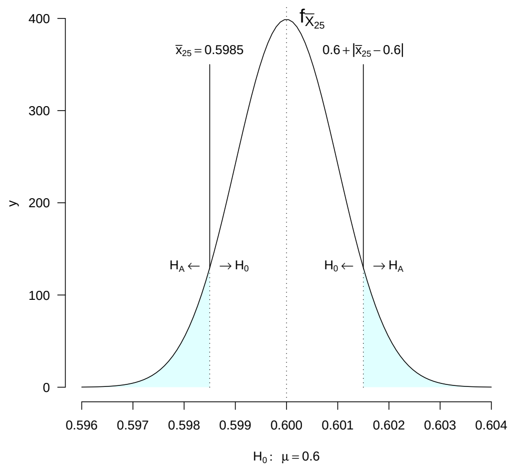


Figura 5.4: Calcolo del p-value (test bilaterale).

1. L'ipotesi, definita come un'affermazione su un determinato fenomeno. Nei nostri due esempi le ipotesi sotto esame concernono il funzionamento di una macchina per la produzione di sfere e la composizione del latte.

2. L'ipotesi nulla, H_0. È un'affermazione grazie alla quale è possibile ottenere la distribuzione della variabile aleatoria X nonché la distribuzione campionaria $f(x_1, \ldots, x_n)$. Spesso l'ipotesi nulla è qualcosa del tipo "... è uguale a ..." oppure "... nessun problema ..." oppure "... nessun effetto ..." oppure "... nessuna differenza ...".

3. L'ipotesi alternativa, H_A, è da intendersi come una "direzione contro l'ipotesi nulla". Nel primo esempio con $H_A : \mu \neq 0.6$ parleremo di test bilaterale, in quanto l'alternativa non specifica una direzione per il vero valore di μ rispetto a quello proposto dall'ipotesi nulla. Nel secondo esempio il test è unilaterale destro in quanto l'alternativa si trova alla destra del valore suggerito da H_0 ($\mu > \mu_0 = -0.545$) . Riassumendo:

   - *Test unilaterale*: l'alternativa è specificata in una sola direzione.
   - *Test bilaterale*: l'alternativa è specificata in due direzioni.

4. Statistica del test. In entrambi gli esempi la statistica S utilizzata nella costruzione del test è la media campionaria. La statistica S deve essere tale per cui, sotto l'ipotesi nulla H_0, la sua distribuzione è conosciuta.

5. $\alpha \in (0, 1)$ (ad esempio $\alpha = 5\%$). Esso corrisponde al livello di significatività desiderato. È la soglia o quantità di evidenza (in termini di probabilità) minima richiesta, sotto la quale si rigetta l'ipotesi nulla in quanto giudicata poco credibile.

6. Il p-value, calcolato tenendo conto dell'ipotesi alternativa (unilaterale o bilaterale). Esso corrisponde alla probabilità di osservare qualcosa di simile o di più estremo (nella direzione suggerita dall'ipotesi alternativa).

7. La decisione se rifiutare o meno il test. Se il p-value è inferiore al livello di significatività α, H_0 sarà rifiutata, altrimenti l'ipotesi nulla verrà mantenuta.

5.2 Regione critica di un Test d'ipotesi

Il test per un'ipotesi statistica presentato nel precedente paragrafo è basato sul concetto di p-value, ovvero la probabilità di osservare un valore della statistica S uguale o più estremo di quello calcolato tramite le osservazioni $(x_1, ..., x_n)$. Il p-value è un concetto relativamente nuovo, introdotto con l'avvento del personal computer grazie al quale il suo calcolo risulta essere immediato. Tuttavia, quella del p-value non è l'unica formulazione possibile di un test d'ipotesi. Infatti, studieremo ora un modo di formulare un test d'ipotesi che è alternativo a quello basato sul p-value, ma *equivalente* dal punto di vista delle conclusioni. La formulazione del test è diversa, non il suo risultato. Utilizzeremo l'Esempio 41 quale punto di partenza.

5.2.1 Test unilaterale destro

Abbiamo la seguente situazione:

1. Vogliamo testare l'ipotesi riguardante la composizione del latte acquistato da un certo fornitore sospettato di annacquare il latte.

2. L'ipotesi nulla H_0 è che il fornitore sia una persona onesta. Sotto questa ipotesi la distribuzione della temperatura di gelo del latte del fornitore, notata X, è identica a quella definita dal Laboratorio cantonale, ovvero $X \sim N(-0.545, 0.008^2)$. L'ipotesi nulla può essere quindi formalizzata come
$$H_0 : \mu = -0.545.$$

3. L'ipotesi alternativa è che il fornitore abbia aggiunto acqua. In tal caso il latte ghiaccerà in media ad una temperatura superiore

$$H_A : \mu > -0.545.$$

4. Abbiamo deciso di utilizzare la media campionaria $\overline{X}$ quale statistica per la costruzione del test. Sotto l'ipotesi nulla la nostra statistica è distribuita $N(-0.545, \frac{0.008^2}{5})$.

5. $\alpha = 5\%$ è il livello di significatività desiderato.

Fino a questo punto la costruzione del test è identica alla formulazione tramite il p-value. Ora, anziché calcolare il p-value, calcoliamo la soglia critica k_α, definita dalla seguente equazione

$$P_{H_0}(\overline{X} \geq k_\alpha) = \alpha.$$

La probabilità di osservare un valore della media campionaria almeno tanto estremo (in direzione dell'ipotesi alternativa) quanto k_α è uguale al livello di significatività α. In pratica, per un test unilaterale destro k_α corrisponde all' $1-\alpha$ quantile $q_{1-\alpha}$ della distribuzione della media campionaria $\overline{X}$ (vedi disegno 5.5). La costante k_α sarà in seguito utilizzata per sapere se rigettare o meno H_0. Riassumendo, poiché abbiamo scelto $\alpha = 5\%$ ed il test è unilaterale destro, la soglia $k_{5\%}$ soddisfa la seguente uguaglianza :

$$P_{H_0}(\overline{X} \geq k_{5\%}) = 0.05. \tag{5.1}$$

Il valore di $k_{5\%}$ può essere ricavato utilizzando le tavole della distribuzione Normale standard. Infatti

$$\begin{aligned}
P_{H_0}(\overline{X} + 0.545 \geq k_{5\%} + 0.545) &= 0.05 \\
P_{H_0}\left(\frac{\overline{X} + 0.545}{0.008/\sqrt{5}} \geq \frac{k_{5\%} + 0.545}{0.008/\sqrt{5}}\right) &= 0.05 \\
P\left(Z \geq \frac{k_{5\%} + 0.545}{0.008/\sqrt{5}}\right) &= 0.05 \\
P\left(Z \leq \frac{k_{5\%} + 0.545}{0.008/\sqrt{5}}\right) &= 0.95 \\
P\left(Z \leq z_{0.95}\right) &= 0.95
\end{aligned}$$

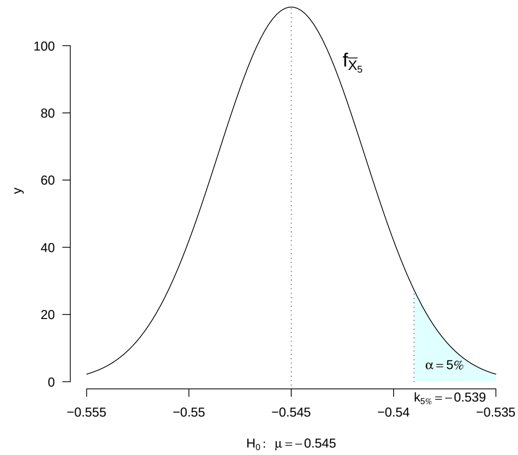


Figura 5.5: Valore critico di un test unilaterale destro.

Sotto l'ipotesi nulla la quantità $Z = \frac{\overline{X}+0.545}{0.008/\sqrt{5}}$ è distribuita $N(0,1)$. Dalle tavole sappiamo che $P(Z \leq 1.645) = 0.95$. Uguagliando 1.645 a $\frac{k_{5\%}+0.545}{0.008/\sqrt{5}}$ si ricava che[3]

$$k_{5\%} = -0.545 + 1.645 \cdot 0.008/\sqrt{5} = -0.539.$$

Per costruzione, la probabilità di osservare un valore di $\overline{X}$ superiore o uguale a -0.539 è dunque uguale 5%. Inoltre, come già osservato nel caso del p-value, l'osservazione di valori estremi in direzione dell'ipotesi alternativa è da considerarsi come evidenza sfavorevole all'ipotesi nulla. Per questo motivo utilizzeremo la seguente regola di decisione per un test unilaterale destro:

7 bis. Se il valore calcolato della statistica $\overline{X}$ (S in generale) cadrà "lontano" e in direzione dell'ipotesi alternativa, ovvero se

$$\overline{x} \geq k_{5\%} = -0.539 \tag{5.2}$$

rifiuteremo l'ipotesi nulla, altrimenti non la rifiuteremo.

Osservazione 28. Il "quanto lontano" deve cadere il valore della nostra statistica affinché H_0 venga rigettata dipende dal livello di significatività α. Tanto più piccolo il valore di α (che rappresenta il livello minimo di evidenza richiesta a favore di H_0) e tanto più estrema sarà la soglia k_α.

[3]Ricordiamo che data una variabile aleatoria $N(\mu, \sigma^2)$ l'α-quantile è dato dalla seguente formula

$$q_\alpha = \mu + z_\alpha \sigma$$

dove z_α rappresenta l'α-quantile della distribuzione Normale standard.

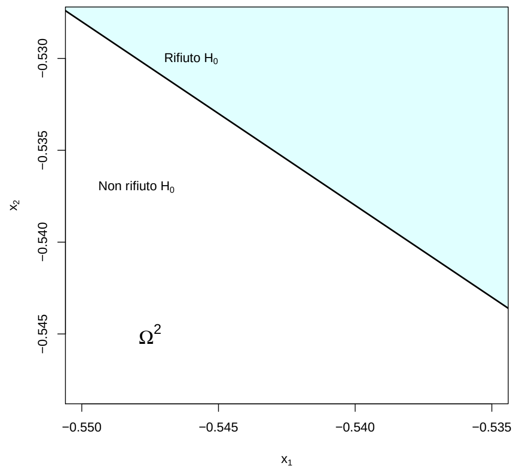


Figura 5.6: Esempio di regione critica (test unilaterale destro).

La decisione se rifiutare o meno il test dipende dalla disequazione (5.2). Essa definisce implicitamente la così detta *regione critica*.

Definizione 34. Regione critica (test unilaterale destro). L'insieme $R \subset \Omega^n$ degli esiti $(x_1, ..., x_n)$ dello spazio campionario per cui $S \geq k_\alpha$ (e quindi tali per cui il test è rifiutato) è chiamato la regione critica del test o semplicemente regione critica

$$R = \{(x_1, ..., x_n) \in \Omega^n \mid S(x_1, \ldots, x_n) \geq k_\alpha\} .$$

R è un evento del nostro esperimento statistico. Per un test unilaterale destro esso corrisponde alla preimmagine rispetto alla statistica S dell'evento $E' = [k_\alpha, \infty)$ dove k_α corrisponde all'$(1-\alpha)$-quantile della distribuzione di S. Nel nostro esempio abbiamo $\alpha = 5\%$, $S = \overline{X}$ e $k_{5\%} = -0.539$.

La regione critica è definita in termini delle possibili realizzazioni $(x_1, ..., x_5)$ del campione $(X_1, ..., X_5)$ e non rispetto all'intervallo $[-0.539, \infty)$. In situazioni molto semplici è possibile ottenere una rappresentazione grafica dell'evento R. Ad esempio, nella Figura 5.6 è rappresentato il grafico di $\overline{x} \geq -0.539$ per $n = 2$.

5.2.2 Test bilaterale

Ricapitolando quanto discusso nell'Esempio 40 concernente il produttore di penne a sfera ci troviamo nella seguente situazione.

1. Vogliamo testare l'ipotesi riguardante la corretta messa a punto della macchina che produce le sfere per la punta delle penne.

2. L'ipotesi nulla H_0 è che la macchina sia correttamente tarata, ovvero che il diametro $X \sim N(0.6, (0.005)^2)$. Matematicamente l'ipotesi nulla è espressa da
$$H_0: \ \mu = 0.6.$$

3. L'ipotesi alternativa è che la macchina non sia a punto e che di conseguenza il diametro medio delle sfere prodotte sia diverso da 0.6
$$H_A: \ \mu \neq 0.6.$$

4. Abbiamo deciso di utilizzare la media campionaria $\overline{X}$ quale statistica per costruire il test. Sotto l'ipotesi nulla la media campionaria delle 25 osservazioni è distribuita secondo la legge Normale $N(0.6, (0.001)^2)$.

5. $\alpha = 5\%$ è il livello di significatività desiderato.

Per quanto riguarda i punti 1-5 non vi è alcuna differenza con la specificazione del test basata sul p-value. A questo punto però anziché calcolare il p-value, costruiamo la regione critica. Essa verrà in seguito utilizzata per definire quando rifiutare il test.

Ma cos'è la regione critica? Rigetteremo l'ipotesi nulla se osserveremo dei valori della statistica estremi in direzione dell'ipotesi alternativa. Essendo questo un test bilaterale (il valore atteso del diametro delle sfere potrebbe essere maggiore o minore rispetto al valore indicato nell'ipotesi nulla), dobbiamo considerare valori estremi della statistica in ambo le direzioni. Ciò è possibile assegnando metà del livello di significatività al lato sinistro della distribuzione e l'altra metà al lato destro. Dovremo quindi calcolare due quantili della distribuzione della statistica S, $\overline{X}$ in questo caso. Il primo (lato sinistro) è il 2.5%-quantile (in generale sarà l'$\frac{\alpha}{2}$-quantile) ed il secondo (lato destro) è il 97.5%$-$quantile ($(1-\frac{\alpha}{2})$-quantile):

$$\begin{cases} P_{H_0}(\overline{X} \leq k_{s,5\%}) = 2.5\% \\ P_{H_0}(\overline{X} \geq k_{d,5\%}) = 2.5\% \end{cases} \tag{5.3}$$

$$\begin{cases} P_{H_0}\left(\overline{X} - 0.6 \leq k_{s,5\%} - 0.6\right) = 2.5\% \\ P_{H_0}\left(\overline{X} - 0.6 \geq k_{d,5\%} - 0.6\right) = 2.5\% \end{cases}$$

$$\begin{cases} P_{H_0}(\frac{\overline{X}-0.6}{0.001} \leq \frac{k_{s,5\%}-0.6}{0.001}) = 2.5\% \\ P_{H_0}(\frac{\overline{X}-0.6}{0.001} \geq \frac{k_{d,5\%}-0.6}{0.001}) = 2.5\% \end{cases}$$

Ancora una volta sotto l'ipotesi nulla la quantità $Z = \frac{\overline{X}-0.6}{0.001}$ è distribuita Normale standard. Definendo $z_{0.025} = \frac{k_{s,5\%}-0.6}{0.001}$ e $z_{0.975} = \frac{k_{d,5\%}-0.6}{0.001}$ il sistema si può riscrivere come

$$\begin{cases} P(Z \leq z_{0.025}) = 2.5\% \\ P(Z \geq z_{0.975}) = 2.5\% \end{cases} \Leftrightarrow P(|\frac{\overline{X}-0.6}{0.001}| \geq z_{0.975}) = 0.05$$

Utilizzando le tavole della distribuzione Normale standard e la sua simmetria attorno allo 0 deduciamo che $z_{0.975} = -z_{0.025} = 1.96$. Infine, si ottiene $k_{s,5\%} = 0.598$ e $k_{d,5\%} = 0.6019$. Siamo ora pronti a definire la regola per il rifiuto dell'ipotesi nulla.

7 bis. Se il valore calcolato della statistica $\overline{X}$ cadrà "lontano" in direzione dell'ipotesi alternativa, ovvero se

$$\overline{x} \leq k_{s,5\%} = 0.598 \text{ oppure } \overline{x} \geq k_{d,5\%} = 0.6019 \tag{5.4}$$

rifiuteremo l'ipotesi nulla. In caso contrario[4] H_0 sarà mantenuta (vedi Figura 5.7).

Anche nel caso di un'ipotesi bilaterale possiamo definire una regione critica R includendo in essa gli esiti (tutte quelle realizzazioni $(x_1, ..., x_{25})$) per cui una delle due disuguaglianze della (5.4) è soddisfatta.

Definizione 35. Regione critica (test bilaterale). L'insieme R degli esiti $(x_1, ..., x_n) \in \Omega^n$ per cui $S(x_1, ..., x_n) \leq k_{s,\alpha}$ oppure $S(x_1, ..., x_n) \geq k_{d,\alpha}$ (e quindi tali per cui il test è rifiutato) è chiamato regione critica del test

$$R := \left\{(x_1, ..., x_n) \in \Omega^2 \mid S(x_1, ..., x_n) \leq k_{s,\alpha} \text{ oppure } S(x_1, ..., x_n) \geq k_{d,\alpha}\right\}$$

o equivalentemente

$$R := S^{-1}(E') \text{ con } E' = (-\infty, k_{s,\alpha}] \cup [k_{d,\alpha}, \infty).$$

La regione critica R è dunque l'evento (sottoinsieme) di Ω^n definito dalla preimmagine rispetto alla statistica S dell'evento $E' = (-\infty, k_{s,\alpha}] \cup [k_{d,\alpha}, \infty)$ dove le costanti sono scelte in modo tale che (vedi Figure 5.8 e 5.9)

$$P_{H_0}(S \in (-\infty, k_{s,\alpha}]) = P_{H_0}(S \in [k_{d,\alpha}, \infty)) = \frac{\alpha}{2}.$$

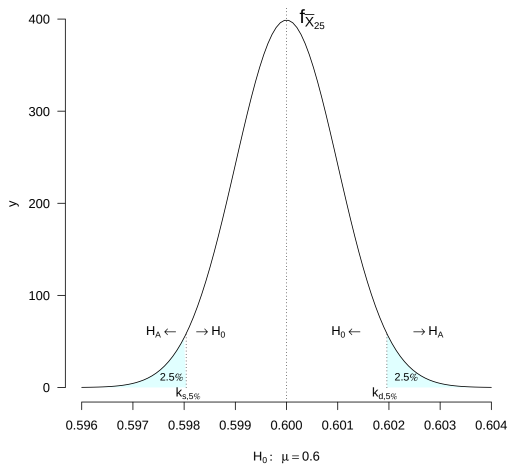


Figura 5.7: Test bilaterale e valori critici con $S = \bar{X}$ e $n = 25$.

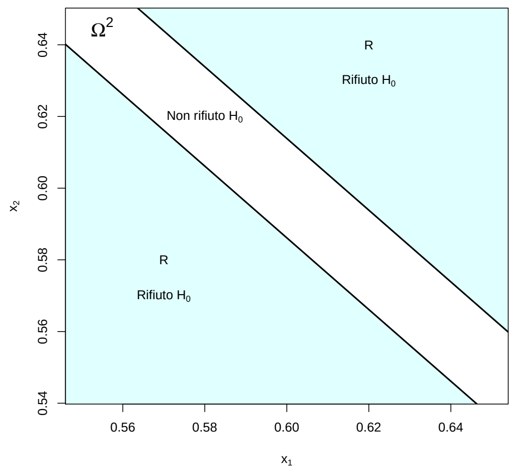


Figura 5.8: Regione critica di un test bilaterale con $S = \bar{X}$ $(n = 2)$.

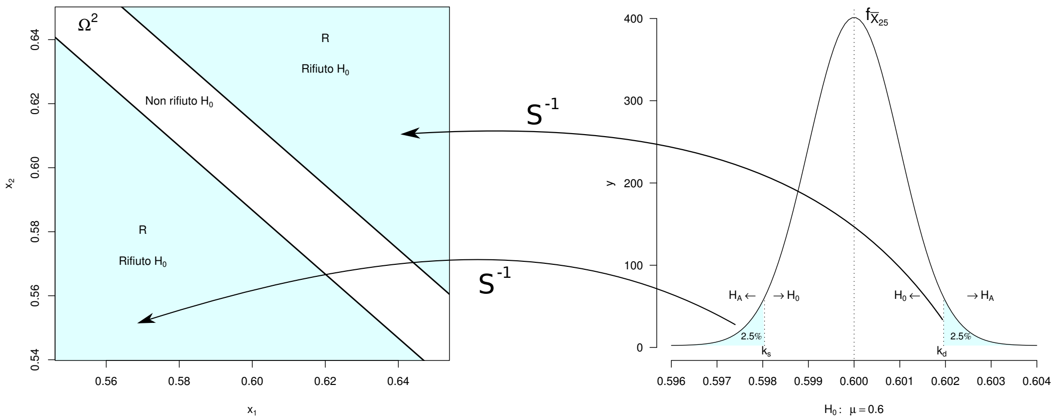


Figura 5.9: Test bilaterale con $S = \bar{X}$ e regione critica ($n = 2$).

Osservazione 29. La costruzione di un test d'ipotesi statistica consiste quindi nel scegliere, oltre che l'ipotesi nulla e quella alternativa, una regione critica R. La regione critica R è un sottoinsieme dello spazio campionario del nostro esperimento statisico: $R \subset \Omega^n$. Se i valori $(x_1, ..., x_n)$ che osserveremo cadranno nella regione critica, ovvero se

- test unilaterale destro: $\overline{x} \geq k_\alpha$,
- test unilaterale sinistro: $\overline{x} \leq k_\alpha$,
- test bilaterale: $\overline{x} \leq k_{s,\alpha}$ oppure $\overline{x} \geq k_{d,\alpha}$,

rifiuteremo l'ipotesi nulla. Per costruzione la regione critica R soddisfa il vincolo concernente il livello di significatività prescelto. Nel caso di un test bilaterale la scelta di $k_{s,\alpha}$ e $k_{d,\alpha}$ è tale per cui dato $E' = (-\infty, k_{s,\alpha}] \cup [k_{d,\alpha}, \infty)$

$$P_{H_0}(S \in E') = P_{H_0}(S^{-1}(E')) = P_{H_0}(R) = \alpha. \tag{5.5}$$

5.3 Scelta della statistica S

La scelta della statistica S utilizzata nella costruzione di un test di ipotesi è importante tanto quanto lo è la scelta dello stimatore T nell'ambito della

[4]Ciò corrisponde al caso in cui

$$\frac{|\overline{X} - 0.6|}{0.001} \leq z_{0.975}.$$

stima puntuale. Nei due esempi precedenti abbiamo utilizzato la media $\overline{X}$ quale statistica S per la costruzione del test[5]. Grazie ad essa è stato possibile definire la regione critica R, ovvero l'evento che definisce il rifiuto dell'ipotesi nulla. Se l'esito dell'esperimento statistico sarà contenuto in R, ovvero se

$$(x_1, x_2, ..., x_n) \in R$$

allora l'ipotesi nulla verrà rifiutata altrimenti verrà mantenuta. Cosa succederebbe se, anziché la media $\overline{X}$, utilizzassimo un'altra statistica? Vedremo che cambiando la statistica si va a modificare la regione critica R. Eseguendo gli stessi passi studiati con la media $\bar{X}$, ovvero

1. definire l'ipotesi nulla e ricavare la distribuzione della nuova statistica S sotto l'ipotesi nulla (passo 1);

2. definire l'ipotesi alternativa, cioé la direzione contro l'ipotesi nulla e la tipologia del test (unilaterale o bilaterale) (passo 2);

3. scegliere il livello α di significatività desiderato (passo 3);

4. calcolare il nuovo p-value e confrontarlo con il livello α di significatività prescelto (passo 4), oppure

5. calcolare le nuove soglie critiche k_α(test unilaterale) o $k_{s,\alpha}$ e $k_{d,\alpha}$ (test bilaterale) ed il nuovo evento critico E' corrispondente (passo 5),

otterremo una nuova regione critica definita come in precedenza da

$$R = S^{-1}(E').$$

Vogliamo considerare questa possibilità per l'Esempio 41 e la statistica $S = \min(X_1, ..., X_5)$. Supponiamo che i cinque valori osservati siano uguali a[6]

$$(x_1, ..., x_5) = (-0.532, -0.54, -0.525, -0.541, -0.542).$$

La realizzazione di S con questo campione è dunque $s = -0.542$. La distribuzione della statistica S dipende dalla distribuzione congiunta di $(X_1, ..., X_5)$. Per eseguire il test (indipendentemente se formulato col p-value o con la regione critica) dobbiamo conoscere la funzione di ripartizione di S.

[5]Dunque nei due casi visti S coincide con lo stimatore $\overline{X}$ del valore atteso della popolazione.

[6]La media dei cinque valori è uguale a -0.536 che corrisponde al valore utilizzato precedentemente per $\overline{x}$.

5.3.1 Distribuzione di $S = \min(X_1, ..., X_5)$

Prima di proseguire nella costruzione del test di H_0 dell'Esempio 41 calcoliamo[7] $F_S(s)$, la distribuzione della statistica S (passo 1), cioè del minimo fra $(X_1, ..., X_5)$. Essa è per definizione uguale a

$$F_S(s) = P(\min(X_1, ..., X_5) \leq s).$$

Anziché tentare di calcolare direttamente la probabilità $P(\min(X_1, ..., X_5) \leq s)$, la riscriviamo come

$$F_S(s) = 1 - P(\min(X_1, ..., X_5) > s). \tag{5.6}$$

Il motivo per cui si percorre questa strada è semplice: prendiamo ad esempio $s = 2$. $\min(X_1, ..., X_5) > 2$ se e solo se tutte e 5 le variabili aleatorie saranno maggiori di 2. Quindi l'evento $\{\min(X_1, ..., X_5) > 2\}$ corrisponde (è identico) all'evento che tutte e 5 le variabili aleatorie sono maggiori di 2. Per tale motivo e per un valore qualsiasi s

$$P(\min(X_1, ..., X_5) > s) = P(X_1 > s, X_2 > s, ..., X_5 > s).$$

Utilizzando l'indipendenza degli X_i

$$P(X_1 > s, ..., X_5 > s) = \prod_{i=1}^{5} P(X_i > s).$$

Infine, poiché $P(X_i > s) = 1 - P(X_i \leq s)$, otteniamo la seguente identità

$$P(\min(X_1, ..., X_5) > s) = P(X_1 > s, ..., X_5 > s) = \prod_{i=1}^{5} (1 - P(X_i \leq s))$$

che quando inserita nella (5.6) ci permette di ottenere la formula finale

$$F_S(s) = 1 - \prod_{i=1}^{5} \left[1 - P(X_i \leq s)\right]. \tag{5.7}$$

Questa formula vale per il minimo di qualsiasi campione $(X_1, ..., X_5)$ di variabili aleatorie indipendenti. Per ottenere la formula (5.7) abbiamo utilizzato

[7] Al fine di distinguere fra la funzione di ripartizione (densità) della statistica S e la funzione di ripartizione (densità) delle variabile aleatoria X, aggiungeremo la lettera S o X al piede di F (f).

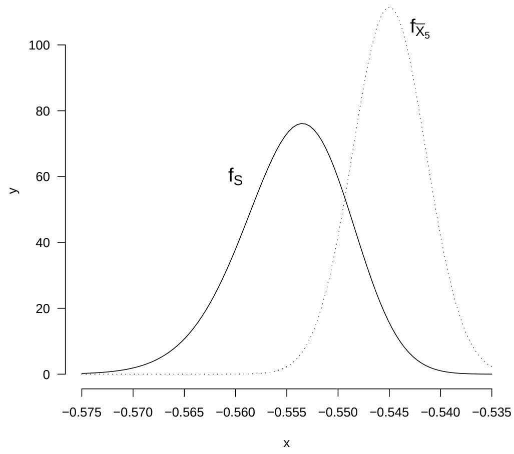


Figura 5.10: Densità di $S = \min(X_1, ..., X_5)$ e $\overline{X}_5$ sotto H_0.

esclusivamente l'indipendenza degli X_i. Poiché $(X_1, ..., X_5)$ sono oltre che indipendenti anche identicamente distribuiti la (5.7) è semplificabile:

$$\begin{aligned} F_S(s) &= 1 - [1 - P(X \leq s)]^5 \\ &= 1 - [1 - F_X(s)]^5 . \end{aligned} \tag{5.8}$$

La formula (5.8) è la formula finale che mette in relazione la funzione di ripartizione di $S = \min(X_1, ..., X_5)$ con la funzione di ripartizione di X, la popolazione dalla quale sono campionate le variabili aleatorie X_i del nostro esperimento statistico. Da questa formula è immediato derivare la funzione di densità di S. Derivando entrambi i lati rispetto ad s otteniamo

$$\frac{\partial}{\partial s} F_S(s) = f_S(s) = 5 \, [1 - F_X(s)]^4 \; f_X(s). \tag{5.9}$$

Abbiamo ora tutto quanto occorre per continuare l'Esempio 41.

5.3.2 Formulazione del test

Desideriamo testare l'ipotesi riguardante la composizione del latte acquistato da un certo fornitore sospettato di aggiungere acqua al suo latte. I passi necessari al test sono

1. L'ipotesi nulla H_0 è che il fornitore sia onesto

$$H_0 : \; \mu = -0.545.$$

Utilizziamo $S = min(X_1, \ldots, X_5)$ quale statistica per costruire il test. Sotto l'*ipotesi nulla* ogni singola $X_i \sim N(-0.545, 0.008^2)$. La funzione di ripartizione di S è data da

$$F_S(s) = 1 - [1 - F_X(s)]^5$$

mentre la funzione di densità è uguale

$$f_S(s) = 5\,[1 - F_X(s)]^4 \; f_X(s).$$

dove

$$f_X(s) = \frac{1}{\sqrt{2\pi}0.008} \exp\left(-\frac{1}{2 \cdot 0.008^2}\,(s + 0.545)^2\right)$$

e

$$F_X(s) = \int_{-\infty}^{s} f_X(x)dx$$

Notate che per dato valore di s sia F_S che f_S potrebbero essere calcolate in Excel o Openoffice utilizzando la funzione NORMDIST().

2. L'ipotesi alternativa è che il fornitore abbia aggiunto acqua

$$H_A: \; \mu > -0.545.$$

Fino a questo punto nulla è mutato. La novità consiste nell'utilizzare la statistica $S = min(X_1, \ldots, X_5)$ anziché $S = \overline{X}$. Grazie al paragrafo 5.3.1 ora conosciamo la distribuzione di S.

3. Come con la media campionaria $\overline{X}$ prendiamo un livello di significatività $\alpha = 5\%$. La scelta del livello di significatività è soggettiva e spetta al ricercatore.

4. Calcolo del p-value. In questo caso il p-value è da calcolarsi rispetto alla nuova statistica $S = min(X_1, \ldots, X_5)$. Come detto in precedenza, il suo valore realizzato è $s = -0.542$. Il p-value sarà dunque

$$\begin{aligned} P_{H_0}(S \geq -0.542) &= 1 - P_{H_0}(S < -0.542) \\ &= 1 - F_S(-0.542) = 0.5545\%. \end{aligned}$$

È questa una probabilità assai bassa ed inferiore al livello di significatività prefissato del 5%. L'evidenza a favore di H_0 è come nel caso della media campionaria $\overline{X}$ insufficiente[8] e per tale motivo rifiutiamo l'ipotesi nulla.

[8] Ricordiamo che in precedenza per la media campionaria $\overline{X}$ era stato calcolato un p-valore dello 0.59%.

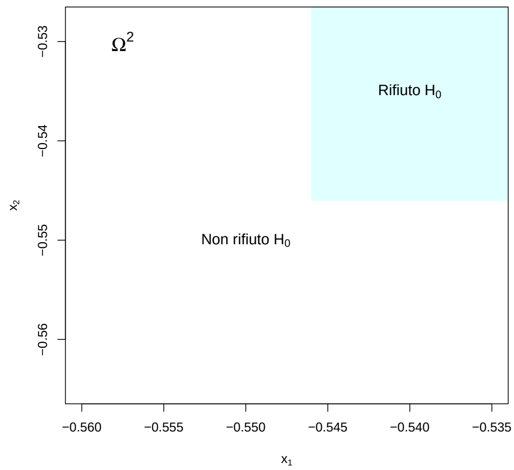


Figura 5.11: Regione critica con $S = min(X_1, \ldots, X_n)$ e $n = 2$.

5. Calcolo della regione critica. Essendo questo un test d'ipotesi unilaterale destro, la regione critica è definita tramite l'$1 - \alpha$ quantile della distribuzione di S, che notiamo k_α.

$$\begin{aligned} P_{H_0}(S \geq k_{5\%}) &= 0.05 & (5.10)\\ F_S(k_{5\%}) &= 0.95 & (5.11)\\ k_{5\%} &= F_S^{-1}(0.95)\\ k_{5\%} &= -0.546 \end{aligned}$$

Per valori della statistica S superiori a $k_{5\%} = -0.546$ (estremi in direzione dell'ipotesi alternativa) rigetteremo l'ipotesi nulla. In questo caso la regione critica R è definita dall'insieme delle possibili realizzazioni $(x_1, ..., x_5)$ tali che $\min(x_1, ..., x_5) \geq k_{5\%} = -0.546$,

$$R = \{(x_1, ..., x_5) \mid \min(x_1, ..., x_5) \geq k_{5\%} = -0.546\}.$$

Il valore realizzato $s = -0.542$ è effettivamente maggiore di -0.546 e pertanto $(x_1, ..., x_5) = (-0.532, -0.54, -0.525, -0.541, -0.542) \in R$. L'ipotesi nulla va rigettata! Per $n = 2$ è istruttivo eseguire il grafico di $\min(x_1, x_2) \geq -0.546$ e confrontarlo con il grafico di $\overline{x} \geq -0.539$. Ovviamente le due regioni critiche sono diverse a conferma che i due test, sebbene costruiti ad uno stesso livello di significatività del 5%, sono diversi.

In questo esempio i due test, l'uno basato sulla media campionaria e l'altro basato sul minimo del campione, hanno condotto alla medesima conclusione. Tuttavia ciò non è sempre il caso. È allora spontaneo chiedersi quale dei due test sia il migliore. Per costruzione, entrambi i test rigetteranno l'ipotesi nulla con una probabilità del 5% quando questa è effettivamente vera. Da questo punto di vista essi sono equivalenti. Per il fornitore dunque, potrebbe essere indifferente l'uso dell'uno o dell'altro al fine di valutare la sua onestà. Riprenderemo questo discorso all'interno del prossimo paragrafo.

5.4 Errore di prima e seconda specie

Se osserviamo la verifica d'ipotesi come un problema decisionale possiamo identificare due tipi di decisioni e due tipi di errori possibili. Per quanto riguarda le decisioni possiamo

- rifiutare H_0,
- non rifiutare H_0 (mantenere l'ipotesi in vista di ulteriori verifiche tramite altri test).

Per quanto riguarda invece gli errori abbiamo due tipi di errori evidenziati in tabella con un asterisco:

	H_0 vera	H_0 falsa
Rifiutare H_0	decisione incorretta* α	decisione corretta β
Non rifiutare H_0	decisione corretta $1-\alpha$	decisione incorretta* $1-\beta$

Abbiamo quindi tipi di errore con probabilità associate (α e $1-\beta$).

- α è chiamato livello di significatività o *errore di prima specie* e corrisponde alla probabilità di rifiutare H_0 quando H_0 è vera. Si tratta di una probabilità condizionata:

$$\alpha = P(\text{Rifiutare } H_0 \mid H_0 \text{ è vera}) = P_{H_0}(\text{Rifiutare } H_0).$$

- β è invece la probabilità di rifiutare H_0 quando questa è falsa ($\Leftrightarrow$ quando H_A vera)

$$\beta = P(\text{Rifiutare } H_0 \mid H_0 \text{ è falsa}) = P_{H_A}(\text{Rifiutare } H_0).$$

- $1-\beta$ è chiamato errore di seconda specie e corrisponde all'errore di non rifiutare H_0 quando questa è falsa.

Osservazione 30. Nell'Esempio 41 del caseificio abbiamo quale ipotesi nulla

$$H_0 : \text{"Il fornitore è innocente"}.$$

ed i seguenti due errori

- Errore di 1^a specie: fornitore dichiarato colpevole quando in realtà è innocente.

- Errore di 2^a specie: fornitore dichiarato innocente quando in realtà è colpevole.

Idealmente non vorremmo mai rifiutare H_0 quando questa è vera (vorremmo un α uguale a 0) e rifiutarla sempre quando essa è falsa (un $\beta = 1$). Questo però non è possibile. Occorre studiare come si possono controllare queste due probabilità per ridurre al minimo le possibilità d'errore.

Possiamo dunque chiederci come si decide di "Rifiutare H_0" sulla base dei dati disponibili. Nei paragrafi precedenti abbiamo definito dei test d'ipotesi basati su una statistica S ed una regione critica R, quest'ultima definita in un test unilaterale tramite una soglia critica k_α

- test unilaterale destro

$$R = \{(x_1, ..., x_n) \mid S(x_1, ..., x_n) \geq k_\alpha\},$$

- test unilaterale sinistro

$$R = \{(x_1, ..., x_n) \mid S(x_1, ..., x_n) \leq k_\alpha\},$$

o una soglia critica inferiore $k_{s,\alpha}$ ed una superiore $k_{d,\alpha}$

- test bilaterale

$$R \;=\; \{(x_1, ..., x_n) \mid S(x_1, ..., x_n) \leq k_{s,\alpha} \text{ oppure } S(x_1, ..., x_n) \geq k_{d,\alpha}\}.$$

La decisione se rifiutare H_0 dipende dal fatto che il campione osservato appartenga o meno alla regione critica. Se (test unilaterale destro)

$$S(x_1, ..., x_n) \geq k_\alpha$$

rifiuteremo H_0 altrimenti non la rifiuteremo. La costante k_α, che definisce la soglia fra rifiuto e non rifiuto di H_0, è fissata *per costruzione* in modo tale che la probabilità *sotto l'ipotesi nulla* H_0 di osservare valori più estremi di k_α sia uguale al livello di significatività α prescelto[9]. L'errore di primo tipo è dunque *per costruzione* uguale ad α.

[9] Proprio per tale motivo si è soliti aggiungere al piede di k, k_d e k_s la probabilità α.

Esempio 44. Tornando all'Esempio 41 del caseificio (test unilaterale destro basato sulla media campionaria)

$$\begin{aligned}\alpha &= P_{H_0}(\text{rifiutare } H_0) = P_{H_0}(\overline{X} \geq k_\alpha) \\ &= P_{H_0}\left(\frac{\overline{X}+0.545}{0.008/\sqrt{5}} \geq \frac{k_\alpha+0.545}{0.008/\sqrt{5}}\right) \\ &= P\left(\underset{\sim N(0,1)}{Z} \geq \frac{k_\alpha+0.545}{0.008/\sqrt{5}}\right)\end{aligned}$$

Utilizzando la solida proprietà $P\left(Z \geq a\right) = 1 - P(Z < a) = 1 - P(Z \leq a)$

$$P\left(Z \leq \frac{k_\alpha+0.545}{0.008/\sqrt{5}}\right) = 1-\alpha$$

da cui

$$\begin{aligned}\frac{k_\alpha+0.545}{0.008/\sqrt{5}} &= z_{1-\alpha} \\ k_\alpha &= -0.545 + z_{1-\alpha}0.008/\sqrt{5}\end{aligned}$$

Utilizzando la tabella sottostante è possibile calcolare le varie soglie della regione critica in funzione del livello di significatività (errore di prima specie) desiderato:

α	5%	2.5%	1%
$1-\alpha$	95%	97.5%	99%
$z_{1-\alpha}$	1.645	1.96	2.33

Ricordiamo che grazie alla simmetria della distribuzione Normale avremo per un test unilaterale sinistro i seguenti valori di z_α :

α	5%	2.5%	1%
z_α	−1.645	−1.96	−2.33

Osservazione 31. Come già accennato, il livello "tipico" di significatività è del 5%. In medicina si indica con

- n.s. : se il risultato non è significativo al 5% (H_0 non rifiutata),
- * : se il risultato è significativo al 5% (H_0 rifiutata),
- ** : se il risultato è significativo all'1% (H_0 rifiutata),
- *** : se il risultato è significativo allo 0.1% (H_0 rifiutata).

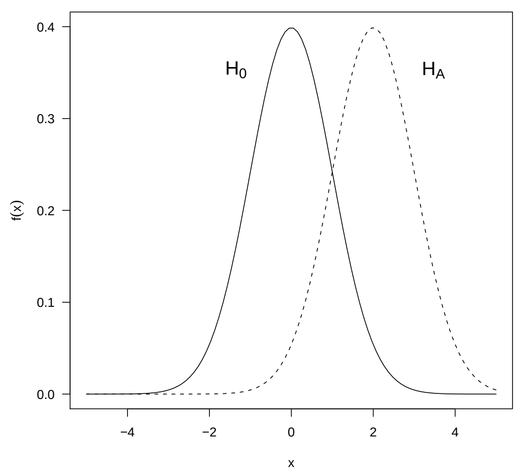


Figura 5.12: Distribuzione Normale sotto l'ipotesi nulla e alternativa.

5.4.1 Relazione tra errore di prima e di seconda specie

Al fine di comprendere la relazione fra l'errore di prima specie α e quello di seconda specie $1-\beta$ analizziamo il seguente esempio.

Esempio 45. Supponiamo che la variabile aleatoria X sia distribuita secondo la legge Normale $N(\mu, 1)$ e che l'ipotesi nulla sia $H_0 : \mu = 0$ mentre l'ipotesi alternativa[10] $H_A : \mu = 2$. Si osserva un'unica realizzazione di X, notata semplicemente x. L'errore di prima specie (livello di significatività) è fissato inizialmente ad $\alpha = 5\%$. Il grafico della funzione di densità di X sotto le due ipotesi è riportato nella Figura 5.12. Poiché in questo caso l'ipotesi alternativa è situata alla destra di H_0, la soglia $k_{5\%}$ della regione critica (ad un livello di significatività del 5% quindi) è definita da

$$P_{H_0}(X \geq k_{5\%}) = 0.05 \Rightarrow k_{5\%} = z_{95\%} = 1.645.$$

La regione critica[11] è

$$R = \{x \in [1.645, \infty)\}.$$

[10]Sono entrambe ipotesi semplici in quanto identificano esattamente la distribuzione di X.

[11]In questo esempio ad uso didattico la statistica S utilizzata è semplicemente il valore osservato. La regione critica R coincide quindi con l'intervallo $[k_{5\%}, \infty)$. Tuttavia è questo un caso più unico che raro che mai si verifica nelle applicazioni reali.

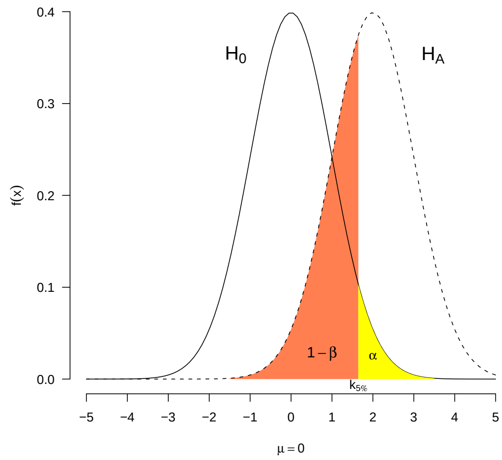


Figura 5.13: Errore di prima e seconda specie, $\alpha = 5\%$.

Graficamente la probabilità $P_{H_0}(X \geq k_{5\%})$ è rappresentata dalla superficie α della Figura 5.13. Determinata la regione critica, il test di H_0 è dunque il seguente:

- si osserva la realizzazione x di X.
- Se x cade nella regione critica rifiutiamo H_0. Condizionatamente alla correttezza di H_0 ciò accadrà, per costruzione, con una probabilità $\alpha = 5\%$.
- Se x cade al di fuori della regione critica non rifiutiamo H_0. Condizionatamente alla correttezza di H_0 questo accadrà, per costruzione, con una probabilità uguale ad $1 - \alpha = 95\%$.

Fissata la regione critica R sulla base del livello di significatività desiderato (errore di prima specie), possiamo chiederci quanto sarebbe la probabilità di rifiutare l'ipotesi nulla (o di non rifiutarla) se la vera distribuzione di X fosse data da H_A.

Osservazione 32. È importante sottolineare che la regione critica non varia: noi rifiuteremo H_0 se l'osservazione di x cadrà nell'intervallo $[1.645, \infty)$ mentre non la rifiuteremo se $x \in (-\infty, 1.645)$. Ciò che desideriamo fare a questo punto è calcolare la probabilità dell'evento $X \in [1.645, \infty)$ dato che la distribuzione di X è ora determinata dall'ipotesi alternativa H_A.

La probabilità di rigettare H_0 è quindi definita dalla probabilità dell'evento $X \geq 1.645$ e la sua probabilità è semplicemente

$$P\left(X \geq 1.645\right). \tag{5.12}$$

Questa probabilità deve essere calcolata sotto l'ipotesi alternativa H_A, vale a dire sotto l'ipotesi che $X \sim N(2,1)$. Facciamo questo aggiungendo H_A alla P della (5.12) che quindi diventa

$$P_{H_A}\left(X \geq 1.645\right)$$

Per calcolare questa probabilità utilizziamo la solita tecnica di standardizzazione, tenendo presente che sotto l'ipotesi alternativa H_A il valore atteso di X è uguale 2 e la varianza 1.

$$P_{H_A}\left(\underbrace{X-2}_{Z\sim N(0,1)} \geq 1.645-2\right) = P\left(Z \geq -0.355\right) = P(Z \leq 0.355) = 63.87\%.$$

Abbiamo così calcolato la probabilità di rifiutare H_0 condizionatamente al fatto che essa sia falsa e che sia invece vera H_A. Questa probabilità è stata indicata con la lettera β e corrisponde alla probabilità dell'azione corretta nel caso che H_A sia vera. In Figura 5.13 essa corrisponde alla superficie rispetto alla funzione di densità *tratteggiata*[12] alla destra di k_α. In teoria noi vorremmo che questa probabilità fosse uguale 1, ma come è evidente dal disegno questo è impossibile per $\alpha < 1$. Siamo dunque costretti ad accettare il fatto che con una probabilità pari a $1-\beta$ (36.13% in questo caso) noi non rifiuteremo H_0 quando questa andrebbe invece rifiutata. Come descritto precedentemente $1-\beta$ è l'errore di seconda specie. In questo semplice esempio ad un errore di prima specie del 5% corrisponde un errore di seconda specie del 36.13%.

Potremmo decidere a questo punto di diminuire l'errore di prima specie (il livello di significatività) e portarlo al 2.5%. Cosa capiterà a β ed indirettamente all'errore di seconda specie $1-\beta$? Calcoliamo innanzi tutto il nuovo valore critico $k_{2.5\%}$, ovvero

$$P_{H_0}(X \geq k_{0.025}) = 0.025 \Rightarrow k_{2.5\%} = z_{97.5\%} = 1.96.$$

Abbiamo ridotto α, l'errore di prima specie ed il valore della soglia critica si è spostato sulla destra. Cosa implica questo per β? Sappiamo che

$$\beta = P_{H_A}\left(\text{Rifiuto } H_0\right) = P_{H_A}\left(X \geq 1.96\right).$$

[12]La densità tratteggiata corrisponde alla funzione di densità di X quando H_A è vera.

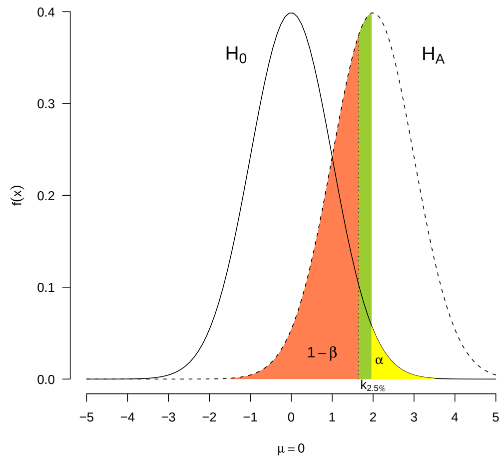


Figura 5.14: Errore di prima e seconda specie, $\alpha = 2.5\%$.

Per calcolare β usiamo la solita tecnica di standardizzazione tenendo presente che sotto l'ipotesi alternativa $X \sim N(2,1)$:

$$\begin{aligned}\beta &= P_{H_A}(X \geq 1.96) = P_{H_A}\left(\underbrace{X-2}_{Z\sim N(0,1)} \geq -0.04\right) \\ &= P(Z \leq 0.04) = 51.6\%\end{aligned}$$

Confrontando la nuova Figura 5.14 alla Figura 5.13 notiamo che la superficie a sinistra di $k_{2.5\%}$ delimitata dalla densità tratteggiata è aumentata rispetto al valore precedente. L'incremento è dato dalla superficie verde. L'errore di seconda specie $1-\beta$ è dunque aumentato, passando dal 36.13% al 48.4%.

Da questo esempio deduciamo che esiste un relazione inversa fra errore di prima specie ed errore di seconda specie nel senso che qualora si decida di diminuire la probabilità d'errore di prima specie si dovrà pagare un costo in termini di una maggiore probabilità d'errore di seconda specie e viceversa. Possiamo semplicemente scrivere quanto detto a parole con

$$\alpha \downarrow \Rightarrow 1-\beta \uparrow .$$

Per tale motivo non è possibile scendere a zero con entrambe le probabilità. Si è pertanto soliti fissare il livello di significatività α ad un livello appropriato. Come detto, in econometria si utilizza un $\alpha = 5\%$. Il nostro sistema giuridico per contro assegna maggiore importanza all'errore di 1^a specie (dichiarare colpevole un innocente) e per tale motivo i valori di α saranno di molto inferiori (radar detection $\alpha = 10^{-10}\%$).

5.4.2 Potenza di un test

J. Neyman e E. Pearson, due famosi statistici del ventesimo secolo, propongono di fissare α, il livello di significatività del test e di massimizzare[13] la probabilità $\beta = P(\text{Rifiutare } H_0 \mid H_0 \text{ è falsa})$. La probabilità β è chiamata potenza del test.

Definizione 36. Potenza di un test (caso unilaterale destro). La potenza di un test, notata β, è definita come la probabilità di rifiutare H_0 quando essa è falsa[14]

$$\beta = P_{H_A}(S \geq k_\alpha).$$

Esempio 46. Nell'Esempio 41 del caseificio abbiamo costruito due possibili test d'ipotesi, il primo basato sulla media $\overline{X}$ del campione ed il secondo basato sul minimo $min(X_1, \ldots, X_n)$ del campione. Per un livello di significatività del 5% i due test avevano condotto alla medesima conclusione. Nel primo caso con $S = \overline{X}$ il valore critico $k_{5\%} = -0.539$, infatti

$$P_{H_0}(\overline{X} \geq -0.539) = 5\%.$$

Con $S = min(X_1, \ldots, X_n)$ il valore critico è $k_{5\%} = -0.546$ poiché

$$P_{H_0}(S \geq -0.546) = 5\%.$$

Vogliamo ora calcolare la potenza β per entrambi i test. Per definizione la potenza del test è

$$\beta = P_{H_A}(\overline{X} \geq -0.539)$$

e rispettivamente

$$\beta = P_{H_A}(S \geq -0.546).$$

[13] Che corrisponde a minimizzare l'errore di seconda specie $1 - \beta$.

[14] Per un test unilaterale sinistro la potenza è definita da

$$\beta = P_{H_A}(S \leq k_\alpha)$$

e per un test bilaterale

$$\beta = P_{H_A}(S \leq k_{s,\alpha} \text{ oppure } S \geq k_{d,\alpha}).$$

Tuttavia l'ipotesi alternativa $H_A: \mu > -0.545$ non identifica con precisione la distribuzione, essa si limita a dire che il valore atteso della popolazione X è superiore a -0.545. Per tale motivo eseguiamo dapprima il calcolo per un particolare valore di $\mu > -0.545$, diciamo $\mu = -0.54$ ed in seguito per un valore di $\mu > -0.545$ qualsiasi.

5.4.2.1 Potenza quando $S = \overline{X}$

• Caso particolare
L'ipotesi alternativa è $X \sim N(-0.54, 0.008^2)$ da cui segue grazie alla proprietà della distribuzione Normale che $\overline{X} \sim N(-0.54, 0.008^2/5)$. Calcolando abbiamo

$$
\begin{aligned}
\beta &= P_{H_A}(\overline{X} \geq -0.539) = P_{H_A}\left(\underbrace{\frac{\overline{X} - (-0.54)}{0.008/\sqrt{5}}}_{Z \sim N(0,1)} \geq \frac{-0.539 - (-0.54)}{0.008/\sqrt{5}}\right) \\
&= P\left(Z \geq 0.2795\right) = 1 - P\left(Z \leq 0.2795\right) = 0.39
\end{aligned}
$$

• Caso generale, ipotesi alternativa $X \sim N(\mu, 0.008^2) \Rightarrow \overline{X} \sim N(\mu, 0.008^2/5)$ con $\mu > -0.545$
Il calcolo è esattamente uguale al precedente, basta sostituire -0.54 con μ

$$
P_{H_A}(\overline{X} \geq -0.539) = P_{H_A}\left(\underbrace{\frac{\overline{X} - \mu}{0.008/\sqrt{5}}}_{Z \sim N(0,1)} \geq \frac{-0.539 - \mu}{0.008/\sqrt{5}}\right) = P\left(Z \geq \frac{-0.539 - \mu}{0.008/\sqrt{5}}\right).
$$

Il valore assunto da β dipende dal valore di μ. In Figura 5.15 è possibile osservare β quale funzione del valore del parametro μ.

5.4.2.2 Potenza quando $S = min(X_1, \dots, X_n)$

• Caso particolare
L'ipotesi alternativa è che $X \sim N(-0.54, 0.008^2)$. Nel paragrafo 5.3.2, punto 4, abbiamo derivato la distribuzione di S sotto l'ipotesi nulla. La funzione di ripartizione di S sotto l'ipotesi alternativa sarà identica, cioè

$$
\beta = P_{H_A}(S \geq -0.546) = 1 - F_S(-0.546) = \left[1 - F_X(-0.546)\right]^5,
$$

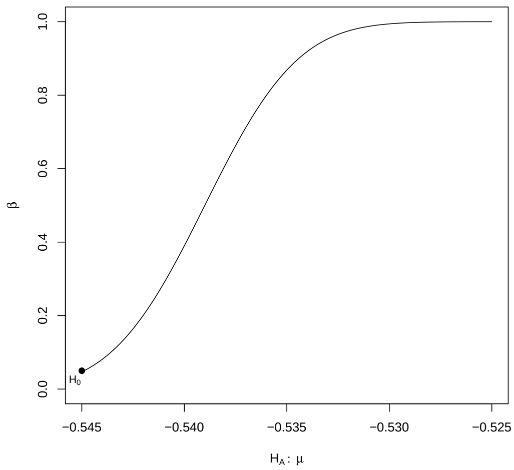


Figura 5.15: Potenza del test dell'Esempio 41 con $S = \overline{X}$.

ma con $F_X(s)$ che rappresenta ora la funzione di ripartizione di X sotto H_A. Il valore di β è dunque

$$\beta = [1 - F_X(-0.546)]^5 = 0.277.$$

Osserviamo che il valore calcolato è inferiore a quello ottenuto utilizzando la media $\overline{X}$ del campione. Ciò significa che quando l'ipotesi nulla è falsa, il test basato sulla statistica $S = min(X_1, \ldots, X_n)$ la rigetterà con una probabilità di solo il 27.7% mentre il test basato sulla statistica $S = \overline{X}$ la rigetterà con una probabilità ben superiore ed uguale al 39%. Il test sulla media $\overline{X}$ è dunque da preferirsi rispetto al test sul $min(X_1, \ldots, X_n)$ poiché *più potente* nel testare H_0 contro l'ipotesi alternativa $H_A : \mu = -0.54$. È interessante chiedersi se questo vale solo per $\mu = -0.54$ o la superiorità del test sulla media è uniforme per qualsiasi $\mu > -0.545$.

• Caso generale
L'ipotesi alternativa è $X \sim N(\mu, 0.008^2)$. In questo caso come con la media calcoleremo $\beta = P_{H_A}(S \geq -0.546)$ per un qualsiasi valore di $\mu > -0.545$. In Figura 5.16 rappresentiamo graficamente la potenza β del test in funzione di μ. Sovrapponendo nel medesimo grafico le curve di potenza dei due test notiamo che effettivamente il test sulla media è da preferire al test sul minimo per qualsiasi ipotesi alternativa semplice $H_A : \mu = \widetilde{\mu} > -0.545$. Vogliamo ora metterci nei panni del fornitore di latte e ragionare rispetto alla scelta del test ottimale. Supponiamo dapprima che egli sia innocente, cioè che H_0

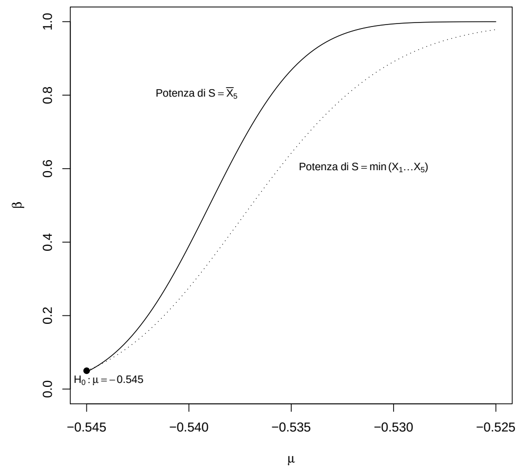


Figura 5.16: Potenza del test dell'Esempio 41 con $S = min$.

sia corretta. Per il fornitore è indifferente che la corte utilizzi il test basato sulla media campionaria $\overline{X}$ o sul minimo $min(X_1, \ldots, X_n)$. Questo perché, in entrambi i casi, la probabilità di essere ingiustamente condannato (errore di prima specie) è uguale a 5%. Per contro, supponiamo che egli sappia di essere colpevole o, in altre parole, che H_A sia vera. In tal caso egli non sarà più indifferente rispetto alla scelta del test. Infatti, qualora la corte utilizzasse il test sulla media, la probabilità β di essere (giustamente) condannato sarebbe uguale al 39% mentre per il test basato su $min(X_1, \ldots, X_n)$ la probabilità sarebbe solo del 27.7%. È chiaro quindi che il fornitore preferirà il test meno potente poiché meno efficace nello scoprire la verità. Il tribunale per contro deciderà di utilizzare il test più potente a disposizione che in questo caso è il test basato sulla media campionaria $\overline{X}$.

In generale potremmo chiederci se, dati

- un'ipotesi nulla $H_0 : \mu = \mu_0$,
- un'ipotesi alternativa $H_A : \mu = \mu_A$,
- un livello prefissato α di significatività (errore di prima specie),

sia possibile trovare (costruire) un test che, fra tutti i possibili test, sia il più potente in assoluto. La risposta a questo quesito ci è data dall'importante lemma di Neyman e Pearson.

5.5 Il Lemma di Neyman e Pearson

In questo paragrafo daremo una soluzione alla domanda che ci siamo posti al termine del paragrafo precedente, ovvero come trovare il test di massima potenza (ricordiamo che la potenza di un test è la probabilità di rifiutare H_0 quando questa è falsa). Prima di dare la soluzione cerchiamo di formalizzare il problema i cui elementi essenziali sono i seguenti.

1. Una popolazione X la cui distribuzione, caratterizzata tramite la sua funzione di densità, è sconosciuta. Si suppone che la densità di X sia contenuta nella famiglia parametrica
$$\mathcal{P} = \{f(x;\theta),\ \theta \in \Theta\}.$$
2. Un'ipotesi nulla H_0 riguardante la distribuzione di X. Tale ipotesi è espressa in termini del parametro sconosciuto θ nel seguente modo
$$H_0 :\ \theta = \theta_0$$
dove θ_0 rappresenta un particolare valore del parametro (θ_0 è quindi un numero).
3. Un'ipotesi alternativa semplice
$$H_A :\ \theta = \theta_A$$
dove θ_A è il valore che il parametro θ assume sotto l'ipotesi alternativa.
4. Il livello α di significatività desiderato (errore di prima specie).

Sulla base di questi elementi desideriamo costruire il test ottimale. Ricordiamo che un qualsiasi test basato sulla statistica S sarà caratterizzato da una regione critica R definita tramite[15]
$$R = \{(x_1, ..., x_n) \mid S(x_1, ..., x_n) \geq k_\alpha\} \tag{5.13}$$
dove la soglia k_α è stata scelta in maniera tale che la regione critica R soddisfi[16] il vincolo
$$P_{H_0}(R) = P_{H_0}(S \geq k_\alpha) = \alpha. \tag{5.14}$$

[15]Se i valori osservati danno un valore della statistica S superiore alla soglia k_α si rifiuta l'ipotesi nulla. Si veda gli esempi con $\overline{X}$ e $min(X_1, \ldots, X_n)$.

[16]Una volta scelta la statistica S, il valore della soglia k_α è automaticamente fissato tramite l'equazione (5.14). In pratica quindi il problema si riduce alla scelta ottimale di S.

Il problema risiede quindi nel trovare la statistica S^* e la soglia k^*_α ad essa associata tale per cui

$$P_{H_0}(S^* \geq k^*_\alpha) = \alpha$$

e

$$P_{H_A}(R^*) = P_{H_A}(S^* \geq k^*_\alpha) = P(\text{rifiuto } H_0 \mid H_0 \text{ è falsa}) = \beta \tag{5.15}$$

sia massimo rispetto a qualsiasi altra statistica S.

Esempio 47. State passeggiando tranquillamente nel bosco quando notate l'entrata di una grotta ed incuriositi decidete di entrarvi. Nel suo interno la luce è molto debole, tanto che dopo alcuni metri la visibilità è praticamente nulla. A questo punto decidete di tornare indietro ma all'improvviso andate ad inciampare contro un baule che una volta aperto risulta contenere 100 kg di gioielli e pietre preziose di ogni genere, colore e dimensione. Volendo approfittare della situazione, decidete di riempire lo zaino con le gemme e i gioielli più preziosi. Purtroppo per voi però, lo zaino ha una capacità massima di $\alpha = 5$ kg. Vi trovate dunque nella spiacevole situazione di dover selezionare, tramite un'opportuna strategia di selezione S, quali gioielli mettere nello zaino. L'insieme dei gioielli e delle pietre preziose che verranno da voi scelti è notato R (raccolto). Ogni gioiello/pietra g ha un peso $p_s(g)$ ed un prezzo di mercato $p_r(g)$. Il vostro scopo è dunque quello di massimizzare il valore della raccolto, sotto il vincolo che il suo peso sia uguale ad α.

Supponiamo che i gioielli contenuti nel baule siano quelli della Tabella (5.1). Per massimizzare il valore del contenuto dello zaino dobbiamo semplicemente ordinare i gioielli in ordine decrescente rispetto al rapporto prezzo/peso. Così facendo ci assicuriamo di prendere gli oggetti di maggior valore per unità di peso.

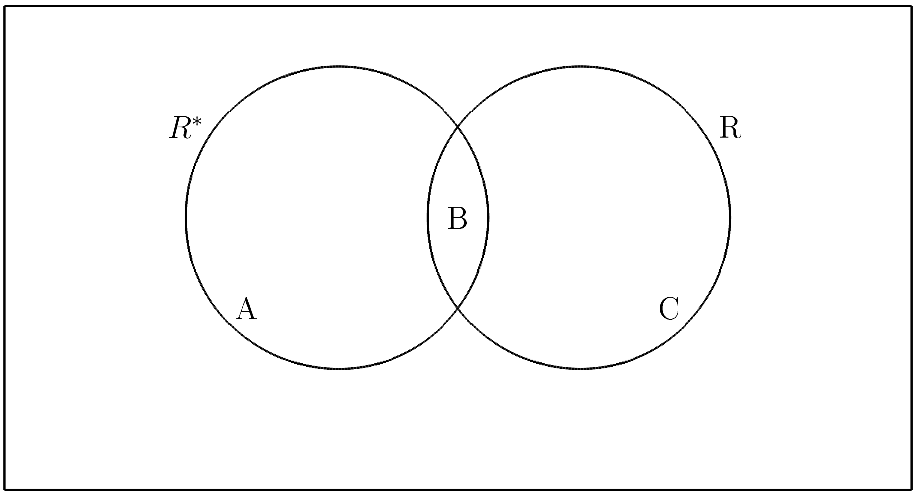
La Tabella (5.2) contenente la lista dei gioielli ordinata rispetto al rapporto prezzo/peso permette di costruire l'insieme R^* del raccolto ottimale, che corrisponde a

$$R^* = \{g_{32}, g_{24}, g_1, g_{21}, g_{36}, g_{16}, g_{27}\}.$$

Notiamo che la strategia di selezione ottimale $S^*(x) = \frac{p_r(g)}{p_s(g)}$ prevede di prendere solo gli oggetti aventi un rapporto prezzo/peso superiore o uguale ad una soglia k^*_α, che nel nostro caso è uguale a 8'888.90. Tale soglia è stata scelta in maniera che il peso complessivo della regione critica sia uguale ad α ovvero 5 kg. Formalmente abbiamo determinato k^*_α tramite la seguente relazione (P_s denota il peso totale)

$$P_s(R^*) = P_s(g \mid S^*(g) \geq k^*_\alpha) = 5 \text{ kg}.$$

Teorema 10. La strategia di selezione $S^* = \frac{p_r(g)}{p_s(g)} \geq k^*_\alpha$ massimizza il valore (prezzo totale) del raccolto R rispetto a tutte le possibili strategie S aventi una raccolto R di peso 5 kg.



Indichiamo con R^* e R l'insieme dei gioielli e delle pietre preziose selezionati utilizzando rispettivamente la strategia di selezione S^* ed una qualsiasi altra strategia S tale per cui il $P_s(R) = 5$ kg. I prezzi di mercato dei raccolti R^* ed R, indicati rispettivamente con $P_r(R^*)$ e $P_r(R)$ sono calcolabili semplicemente come la somma del prezzo di ogni singolo gioiello in essi contenuti, ovvero

$$P_r(R^*) = \sum_{g \in R^*} p_r(g) = \sum_{g \in A} p_r(g) + \sum_{g \in B} p_r(g)$$

e

$$P_r(R) = \sum_{g \in R} p_r(g) = \sum_{g \in B} p_r(g) + \sum_{g \in C} p_r(g).$$

Vogliamo ora dimostrare che

$$P_r(R^*) > P_r(R).$$

A tal fine consideriamo la differenza

$$P_r(R^*) - P_r(R) = \sum_{g \in A} p_r(g) - \sum_{g \in C} p_r(g). \tag{5.16}$$

Poiché la strategia S^* prevede di includere g in R^* se e solo se $\frac{p_r(g)}{p_s(g)} \geq k^*_\alpha$, per ogni $g \in R^*$ (e quindi anche per ogni $g \in A \subset R^*$) sarà vero che $p_r(g) \geq$

$k_\alpha^* p_s(g)$. Di conseguenza varrà la seguente disuguaglianza

$$P_r(R^*) - P_r(R) \geq k_\alpha^* \sum_{g\in A} p_s(g) - \sum_{g\in C} p_r(g).$$

D'altra parte, visto che l'insieme C è disgiunto da R^*, per qualsiasi $g \in C$ deve valere che $\frac{p_r(g)}{p_s(g)} \leq k_\alpha^*$, ovvero $p_r(g) \leq p_s(g) k_\alpha^*$ e di conseguenza

$$\begin{aligned} P_r(R^*) - P_r(R) &\geq k_\alpha^* \sum_{g\in A} p_s(g) - \sum_{g\in C} p_r(g) \\ &\geq k_\alpha^* \sum_{g\in A} p_s(g) - k_\alpha^* \sum_{g\in C} p_s(g) \\ &= k_\alpha^* \left(\sum_{g\in A} p_s(g) - \sum_{g\in C} p_s(g) \right) \end{aligned}$$

Infine, notiamo che la quantità fra parentesi è uguale 0. Infatti

$$\begin{aligned} \sum_{g\in A} p_s(g) - \sum_{g\in C} p_s(g) &= \sum_{g\in A} p_s(g) + \sum_{g\in B} p_s(g) - \sum_{g\in B} p_s(g) - \sum_{g\in C} p_s(g) \\ &= \sum_{g\in R^*} p_s(g) - \sum_{g\in R} p_s(g) \\ &= P_s(R^*) - P_s(R) = 5 - 5 = 0 \end{aligned}$$

Abbiamo dunque dimostrato che

$$P_r(R^*) \geq P_r(R).$$

L'Esempio 47 può essere ripreso ed adattato al problema di scelta del test ottimale. Infatti valgono le seguenti analogie:

- L'insieme dei gioielli e delle pietre prezione (raccolto) R messe nel sacco corrispondono alla regione critica R, ovvero all'insieme delle osservazioni campionarie tali per cui l'ipotesi nulla H_0 va rifiutata.
- Il limite di peso $\alpha = 5$ kg che coincide col peso del raccolto $P_s(R) = \sum_{g\in R} p_s(g)$ corrisponde al livello di significatività α ovvero a $P_{H_0}(R)$.
- P_s che nell'esempio denota la funzione peso di un qualsiasi insieme di pietre preziose corrisponde alla probabilità sotto l'ipotesi nulla P_{H_0}. p_s dal canto suo corrisponde alla funzione di densità sotto l'ipotesi nulla.

- Allo stesso modo P_r, che corrisponde al prezzo di un qualsiasi insieme di pietre preziose, rappresenta la probabilità sotto l'ipotesi alternativa H_A di rifiutare l'ipotesi nulla. Tale probabilità è rappresentata da $P_{H_A}(R)$ e va massimizzata per dato α. Nell'Esempio 47 abbiamo massimizzato il valore del raccolto, ovvero $P_r(R)$. Analogamente a p_s, p_r corrisponde alla funzione di densità sotto l'ipotesi alternativa.

- La strategia S, ovvero il criterio utilizzato per la costruzione del raccolto (selezione dei gioielli da inserire nel sacco) corrisponde alla statistica S da utilizzarsi per la costruzione della regione critica R. La strategia ottimale dell'Esempio 47 corrisponde alla strategia

$$S^*(g) = \frac{p_r(g)}{p_s(g)}$$

utilizzata per la costruzione della regione critica (riempire il sacco)

$$R^* = \{g \mid S^*(g) \geq k^*_\alpha\}$$

con k^*_α scelto in maniera tale che

$$P_s(R) = \alpha = 5 \text{ kg}.$$

Se adattato alla scelta del test ottimale la statistica ottimale S^* diventa semplicemente il rapporto

$$S^*(x_1, ..., x_n) = \frac{f(x_1, ..., x_n; \theta_A)}{f(x_1, ..., x_n; \theta_0)}$$

delle due funzioni di densità del campione sotto l'ipotesi alternativa (numeratore) e sotto l'ipotesi nulla (denominatore). Il criterio per la costruzione della regione critica R^* corrispondente è dunque

$$R^* = \left\{ (x_1, ..., x_n) \mid S^*(x_1, ..., x_n) = \frac{f(x_1, ..., x_n; \theta_A)}{f(x_1, ..., x_n; \theta_0)} \geq k^*_\alpha \right\}. \quad (5.17)$$

L'interpretazione della relazione (5.17) è molto semplice ed intuitiva. Essa afferma che nella costruzione della regione critica (l'insieme delle osservazioni per le quali rigettare l'ipotesi nulla) occorre tener conto del rapporto delle probabilità sotto le due ipotesi. La regione critica dovrà essere costituita da quelle osservazioni che forniscono la maggiore evidenza a favore dell'ipotesi alternativa (contro l'ipotesi nulla). La regione critica ottimale sarà dunque composta dalle realizzazioni con un alto rapporto $\frac{f(x_1,\ldots,x_n;\theta_A)}{f(x_1,\ldots,x_n;\theta_0)}$. La soglia k_α^* è determinata dalla condizione che l'errore di prima specie sia uguale al livello di significatività desiderato

$$P_{H_0}(R^*) = P_{H_0}(S^* \geq k_\alpha^*) = \alpha.$$

Così facendo, per un dato livello di errore di primo tipo α, la potenza β del test è massima e l'errore di secondo tipo $1-\beta$ minimo.

Tabella 5.1: Elenco gioielli

Gioiello	**Peso (grammi)**	**Prezzo in Fr.**	**Prezzo/Peso**
g_1	671	8'000	11'923
g_2	1'717	1'000	582
g_3	1'034	7'000	6'770
g_4	1'756	7'000	3'986
g_5	1'199	2'000	1'668
g_6	886	1'000	1'129
g_7	1'166	1'000	858
g_8	1'040	2'000	1'923
g_9	932	5'000	5'365
g_{10}	1'380	7'000	5'072
g_{11}	1'769	10'000	5'653
g_{12}	1'686	2'000	1'186
g_{13}	902	6'000	6'652
g_{14}	1'257	3'000	2'387
g_{15}	1'768	3'000	1'697
g_{16}	528	5'000	9'470
g_{17}	1'676	6'000	3'580
g_{18}	1'523	10'000	6'566
g_{19}	1'666	9'000	5'402
g_{20}	1'456	3'000	2'060
g_{21}	871	10'000	11'481
g_{22}	988	4'000	4'049
g_{23}	1'190	6'000	5'042
g_{24}	540	7'000	12'963
g_{25}	831	5'000	6'017
g_{26}	802	6'000	7'481
g_{27}	900	8'000	8'889
g_{28}	1'679	6'000	3'574
g_{29}	1'166	6'000	5'146
g_{30}	1'980	5'000	2'525
g_{31}	850	2'000	2'353
g_{32}	676	10'000	14'793
g_{33}	1'399	3'000	2'144
g_{34}	1'883	3'000	1'593
g_{35}	2'535	7'000	2'761
g_{36}	814	9'000	11'057
g_{37}	853	2'000	2'345
g_{38}	1'488	2'000	1'344
g_{39}	1'430	2'000	1'399
g_{40}	1'113	1'000	898
Totale	50'000	202'000	

Tabella 5.2: Elenco gioielli ordinati secondo P_r/P_s

Gioiello	**Prezzo/Peso**	**Peso Cumulato**	**Prezzo Cumulato**
g_{32}	14'792.9	676	10'000
g_{24}	12'963.0	1'216	17'000
g_1	11'922.5	1'887	25'000
g_{21}	11'481.1	2'758	35'000
g_{36}	11'056.5	3'572	44'000
g_{16}	9'469.7	4'100	49'000
g_{27}	**8'888.9**	**5'000**	**57'000**
g_{26}	7'481.3	5'802	63'000
g_3	6'769.8	6'836	70'000
g_{13}	6'651.9	7'738	76'000
g_{18}	6'566.0	9'261	86'000
g_{25}	6'016.8	10'092	91'000
g_{11}	5'652.9	11'861	101'000
g_{19}	5'402.2	13'527	110'000
g_9	5'364.8	14'459	115'000
g_{29}	5'145.8	15'625	121'000
g_{10}	5'072.5	17'005	128'000
g_{23}	5'042.0	18'195	134'000
g_{22}	4'048.6	19'183	138'000
g_4	3'986.3	20'939	145'000
g_{17}	3'580.0	22'615	151'000
g_{28}	3'573.6	24'294	157'000
g_{35}	2'761.3	26'829	164'000
g_{30}	2'525.3	28'809	169'000
g_{14}	2'386.6	30'066	172'000
g_{31}	2'352.9	30'916	174'000
g_{37}	2'344.7	31'769	176'000
g_{33}	2'144.4	33'168	179'000
g_{20}	2'060.4	34'624	182'000
g_8	1'923.1	35'664	184'000
g_{15}	1'696.8	37'432	187'000
g_5	1'668.1	38'631	189'000
g_{34}	1'593.2	40'514	192'000
g_{39}	1'398.6	41'944	194'000
g_{38}	1'344.1	43'432	196'000
g_{12}	1'186.2	45'118	198'000
g_6	1'128.7	46'004	199'000
g_{40}	898.5	47'117	200'000
g_7	857.6	48'283	201'000
g_2	582.4	50'000	202'000

5.6 Esempio

Vogliamo applicare il lemma di Neyman e Pearson alla seguente, classica situazione. Supponiamo di osservare un campione di numerosità n estratto da una popolazione $X \sim N(\mu, \sigma^2)$ con σ^2 conosciuto. La famiglia parametrica $\mathcal{P}$ è dunque data da

$$\mathcal{P} = \left\{ f(x;\mu) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2\sigma^2}\,(x-\mu)^2),\ \mu \in \mathbb{R} \right\}.$$

Supponiamo di voler testare l'ipotesi nulla

$$H_0:\ \mu = \mu_0$$

contro l'ipotesi alternativa

$$H_A:\ \mu = \mu_1,$$

dove $\mu_1 > \mu_0$. Qual è sotto queste condizioni il test più potente? Per rispondere a questa domanda applichiamo i risultati del lemma di Neyman e Pearson. La funzione di densità congiunta di $(X_1, ..., X_n)$ è semplicemente il prodotto delle densità di $X_1, ..., X_n$

$$f(x_i;\mu) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2\sigma^2}\,(x_i-\mu)^2)$$

e

$$\begin{aligned} f(x_1, ..., x_n;\mu) &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp(-\frac{1}{2\sigma^2}\left[(x_1-\mu)^2 + ... + (x_n-\mu)^2\right]) \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2). \end{aligned}$$

La statistica ottimale S definita dal rapporto fra le densità del campione sotto l'ipotesi alternativa e quella nulla è data da

$$
\begin{aligned}
S(x_1,...,x_n) &= \frac{f_{H_A}(x_1,...,x_n)}{f_{H_0}(x_1,...,x_n)} \\
&= \frac{\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp(-\frac{1}{2\sigma^2}\sum\limits_{i=1}^{n}(x_i-\mu_1)^2)}{\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp(-\frac{1}{2\sigma^2}\sum\limits_{i=1}^{n}(x_i-\mu_0)^2)} \\
&= \exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(x_i-\mu_1)^2-\sum_{i=1}^{n}(x_i-\mu_0)^2\right]\right) \\
&= \exp\left(-\frac{1}{2\sigma^2}\left[n(\mu_1^2-\mu_0^2)-2\underbrace{\sum_{i=1}^{n}x_i}_{n\overline{x}}\mu_1+2\underbrace{\sum_{i=1}^{n}x_i}_{n\overline{x}}\mu_0\right]\right)
\end{aligned}
$$

La regione critica sappiamo essere definita dalla seguente formula.

$$R = \{(x_1,...,x_n) \mid S(x_1,...,x_n) \geq k_\alpha\}.$$

Otteniamo quindi

$$R = \left\{(x_1,...,x_n) \mid \exp\left(-\frac{1}{2\sigma^2}\left[n(\mu_1^2-\mu_0^2)-2n(\mu_1-\mu_0)\overline{x}\right]\right) \geq k_\alpha\right\}, \tag{5.18}$$

con k_α da determinare tramite il vincolo $P_{H_0}(R)=\alpha$. Poiché n, μ_0, μ_1 sono conosciuti possiamo riscrivere la (5.18) come

$$
\begin{aligned}
R &= \left\{(x_1,...,x_n) \mid -\frac{1}{2\sigma^2}\left[n(\mu_1^2-\mu_0^2)-2n(\mu_1-\mu_0)\overline{x}\right] \geq \ln(k_\alpha)\right\} \\
&= \left\{(x_1,...,x_n) \mid \left[n(\mu_1^2-\mu_0^2)-2n(\mu_1-\mu_0)\overline{x}\right] \leq -2\sigma^2\ln(k_\alpha)\right\} \\
&= \left\{(x_1,...,x_n) \mid -2n(\mu_1-\mu_0)\overline{x} \leq -2\sigma^2\ln(k_\alpha)-n(\mu_1^2-\mu_0^2)\right\} \\
&= \left\{(x_1,...,x_n) \mid \overline{x} \geq \left[2\sigma^2\ln(k_\alpha)+n(\mu_1^2-\mu_0^2)\right]/2n(\mu_1-\mu_0)\right\} \\
&= \left\{(x_1,...,x_n) \mid \overline{x} \geq \widetilde{k}_\alpha\right\}.
\end{aligned}
$$

Abbiamo sorprendentemente ottenuto il test basato sulla media campionaria incontrato ripetutamente in questo capitolo. La soglia $\widetilde{k}_\alpha$ deve essere semplicemente scelta in modo che $P_{H_0}(R)=\alpha$. Il passaggio da k_α a $\widetilde{k}_\alpha$ è stato causato dalle trasformazioni che abbiamo applicato alla (5.18). Tuttavia queste trasformazoni non modificano la regione critica R, modificano solo la formula utilizzata per rappresentarla.

Il risultato così ottenuto ha la seguente interpretazione. Quando la popolazione è distribuita secondo la legge Normale con σ^2 conosciuto il test sulla media è il miglior test (il più potente). Non a caso avevamo visto nel lungo Esempio 41 che il test basato su $S = \overline{X}$ risultava essere più potente rispetto al test basato su $S = min(X_1, \ldots, X_n)$. Notiamo inoltre che il risultato è indipendente dal valore di μ_1. Questo significa che sotto l'ipotesi di normalità il test basato sulla media è il più potente non solo rispetto all'ipotesi alternativa semplice $H_A : \mu = \mu_1$ ma anche rispetto all'ipotesi alternativa multipla $H_A : \mu > \mu_0$.

5.6.1 Calcolo della potenza

Il calcolo della potenza del test basato sulla media è particolarmente istruttivo per analizzare la dipendenza di β da μ_0, μ_1, n ed α. Sia dunque $X \sim N(\mu, \sigma^2)$ con σ^2 conosciuto e

$$H_0: \ \mu = \mu_0$$

mentre l'ipotesi alternativa

$$H_A: \ \mu = \mu_1, \qquad \mu_1 > \mu_0.$$

Prendendo il test basato sulla media abbiamo per il caso unilaterale destro che

$$R = \{(x_1, ..., x_n) \mid \overline{x} \geq k_\alpha\}$$

con la soglia critica k_α determinata tramite la solita equazione

$$P_{H_0}(\overline{X} \geq k_\alpha) = \alpha.$$

Poiché sotto l'ipotesi nulla $\overline{X} \sim N(\mu_0, \frac{\sigma^2}{n})$, applichiamo la standardizzazione seguente

$$P_{H_0}\left(\underbrace{\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}}_{Z \sim N(0,1)} \geq \frac{k_\alpha - \mu_0}{\sigma/\sqrt{n}}\right) = \alpha$$

Utilizzando la relazione $P\left(Z \geq \frac{\sqrt{n}(k_\alpha - \mu_0)}{\sigma}\right) = 1 - P\left(Z \leq \frac{\sqrt{n}(k_\alpha - \mu_0)}{\sigma}\right)$ otteniamo

$$P\left(Z \leq \frac{k_\alpha - \mu_0}{\sigma/\sqrt{n}}\right) = 1 - \alpha. \tag{5.19}$$

Poiché l'$1 - \alpha$ quantile $z_{1-\alpha}$ è soluzione della precedente equazione, cioé

$$P\left(Z \leq z_{1-\alpha}\right) = 1 - \alpha,$$

ricaviamo il valore della soglia critica k_α uguagliando l'espressione nella (5.19) a $z_{1-\alpha}$

$$\begin{aligned} \frac{k_\alpha - \mu_0}{\sigma/\sqrt{n}} &= z_{1-\alpha} \\ k_\alpha &= \mu_0 + z_{1-\alpha}\frac{\sigma}{\sqrt{n}}. \end{aligned} \tag{5.20}$$

La soglia critica k_α dipende dunque

- da μ_0, il valore di μ sotto l'ipotesi nulla,
- dal livelo di significatività α che determinina indirettamente $z_{1-\alpha}$,
- dalla numerosità del campione (quantità d'informazione disponibile).

Una volta che k_α è stato fissato *esso rimane costante*. Il calcolo della potenza del test presuppone il calcolo della probabilità $P(\overline{X} \geq k_\alpha \mid H_A \text{ è vera})$, cioè

$$\begin{aligned}
\beta &= P_{H_A}(\overline{X} \geq k_\alpha) \\
&= P_{H_A}\left(\underbrace{\frac{\overline{X}-\mu_1}{\sigma/\sqrt{n}}}_{Z\sim N(0,1)} \geq \frac{k_\alpha-\mu_1}{\sigma/\sqrt{n}}\right) \\
&= P\left(Z \geq \frac{k_\alpha-\mu_1}{\sigma/\sqrt{n}}\right) \\
&= 1-P\left(Z \leq \frac{k_\alpha-\mu_1}{\sigma/\sqrt{n}}\right)
\end{aligned}$$

Sostituendo l'espressione ricavata in precedenza per k_α otteniamo la formula finale

$$\begin{aligned}
\beta &= 1-P\left(Z \leq z_{1-\alpha} - (\mu_1-\mu_0)\frac{\sqrt{n}}{\sigma}\right) \\
&= 1-\Phi\left(z_{1-\alpha} - (\mu_1-\mu_0)\frac{\sqrt{n}}{\sigma}\right)
\end{aligned} \tag{5.21}$$

dove Φ rappresenta la funzione di ripartizione della Normale standard, cioé

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx.$$

Osserviamo che la potenza β del test dipende oltre che

- dal valore di μ sotto l'ipotesi nulla H_0,
- dal livello di significatività α
- dalla numerosità del campione

anche da

- μ_1, il valore di μ sotto l'ipotesi alternativa H_A.

La formula (5.21) ci permette di analizzare il comportamento della potenza in funzione dei vari parametri μ_1, n, α.

1. μ_1, n fissi: modifichiamo solo α.

   Quando il livello di significatività α aumenta, $1-\alpha$ diminuisce e con esso $z_{1-\alpha}$. Poiché Φ è una funzione monotona crescente, β aumenterà. Quindi quando
   $$\alpha \uparrow \rightarrow \beta \uparrow \text{ e } 1-\beta \downarrow .$$

2. α, n fissi: cambia solo μ_1.

   Quando μ_1 aumenta l'alternativa si allontana dall'ipotesi nulla. Caeteris paribus Φ diminuirà e quindi β tenderà ad aumentare. L'errore di secondo tipo $1-\beta$ diminuirà. Più l'ipotesi alternativa si allontana dall'ipotesi nulla e più la potenza aumenta (l'errore di seconda specie diminuisce): il test discrimina meglio fra l'ipotesi nulla e l'alternativa.

3. μ_1, α fissi: cambia solo n.

   Con l'aumentare del numero n di osservazioni la funzione Φ diminuisce e la potenza β del test aumenta. L'errore di seconda specie diminuisce. Ciò è in linea con l'intuizione. Se ho più informazione a disposizione riuscirò a discriminare in maniera migliore tra le due ipotesi.

Questi calcoli di potenza vengono eseguiti prima di eseguire una stima per identificare la numerosità del campione. Ad esempio, supponiamo che per dato errore di prima specie $\alpha = 5\%$ e ipotesi alternativa μ_1 si desideri avere un errore di seconda specie $1-\beta = 20\%$. La relazione data dalla formula (5.21) ci permette di determinare la numerosità del campione necessario

$$\begin{aligned} \beta &= 1-\Phi\left(z_{1-\alpha} - (\mu_1 - \mu_0)\frac{\sqrt{n}}{\sigma}\right) \\ 1-\beta &= \Phi\left(z_{1-\alpha} - (\mu_1 - \mu_0)\frac{\sqrt{n}}{\sigma}\right). \end{aligned}$$

Poiché $z_{1-\beta}$ soddisfa $1-\beta = \Phi(z_{1-\beta})$ uguagliamo $z_{1-\beta}$ a $z_{1-\alpha} - (\mu_1 - \mu_0)\frac{\sqrt{n}}{\sigma}$ ed otteniamo

$$n = \left(\frac{z_{1-\beta} - z_{1-\alpha}}{(\mu_1 - \mu_0)/\sigma}\right)^2$$

che in questo caso per $(\mu_1 - \mu_0)/\sigma = 0.1$ diventa

$$n = \left(\frac{-0.841 - 1.645}{0.1}\right)^2 = 618.$$

Capitolo 6

TEST

6.1 Test sulla media con X non Normale

Il test sulla media fino ad ora considerato si basa sull'ipotesi di normalità della popolazione X. Vogliamo ora rinunciare all'ipotesi di normalità ed assumere unicamente che la popolazione sia distribuita secondo una certa distribuzione F (non necessariamente Normale quindi) con $E(X) = \mu$ e $V(X) = \sigma^2$. Scriveremo semplicemente $X \sim F(\mu, \sigma^2)$. Per contro manteniamo l'ipotesi di indipendenza e stessa distribuzione (detta ipotesi *i.i.d.*) delle osservazioni $X_1, ..., X_n$. Come andrà effettuato ora il test sulla media? Supponiamo di voler testare l'ipotesi nulla

$$H_0 : \mu = \mu_0$$

contro l'ipotesi alternativa

$$H_A : \mu > \mu_0$$

ad un livello di significatività α del 5%. Sebbene grazie alle proprietà del valore atteso e della varianza possiamo affermare che $\overline{X} \sim (\mu, \frac{\sigma^2}{n})$, la distribuzione di $\overline{X}$ non è conosciuta. Questo pone un problema per la determinazione della regione critica ed in particolare della soglia k_α

$$P_{H_0}(\overline{X} \geq k_{0.05}) = 0.05.$$

Tuttavia proviamo ugualmente a standardizzare:

$$P_{H_0}(\overline{X} - \mu_0 \geq k_{0.05} - \mu_0) = 0.05$$

e dividendo per $\sigma/\sqrt{n}$

$$P_{H_0}(\frac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}} \geq \frac{k_{0.05} - \mu_0}{\sigma/\sqrt{n}}) = 0.05.$$

Pur non conoscendo l'esatta distribuzione di $\frac{\overline{X}_n-\mu_0}{\sigma/\sqrt{n}}$ sappiamo che, per n sufficientemente grande, il teorema del limite centrale viene in nostro soccorso permettendoci di calcolare il valore di $k_{0.05}$ come se X fosse $N(\mu,\sigma^2)$ e quindi come se $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$. Infatti, dato un campione $X_1, ..., X_n$ di n osservazioni indipendenti ed identicamente distribuite $X \sim (\mu,\sigma^2)$

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \underset{n\to\infty}{\sim} N(0,1).$$

Ma questo significa che, sotto l'ipotesi nulla

$$P_{H_0}\left(\underbrace{\frac{\overline{X}_n-\mu_0}{\sigma/\sqrt{n}}}_{\text{circa } Z} \geq \frac{k_{0.05}-\mu_0}{\sigma/\sqrt{n}}\right) \underset{\text{circa}}{=} P\left(Z \geq \frac{k_{0.05}-\mu_0}{\sigma/\sqrt{n}}\right) = 0.05.$$

Uguagliando $z_{0.95}$ a $\frac{k_{0.05}-\mu_0}{\sigma/\sqrt{n}}$ è infine possibile ottenere il valore della soglia critica $k_{0.05}$. Riassumendo, grazie al Teorema del limite centrale abbiamo che quando

- la varianza della popolazione è nota;
- le osservazioni sono indipendenti ed identicamente distribuite;
- il numero n di osservazioni è sufficientemente grande;

il test sulla media è identico al caso in cui la popolazione è distribuita secondo la legge Normale.

6.2 Test di Student (t-test)

Vogliamo riprendere nuovamente il test sulla media riportandoci al caso in cui la popolazione X è Normale $N(\mu,\sigma^2)$. Questa volta consideriamo la situazione in cui la varianza σ^2 della popolazione è sconosciuta. In questo contesto σ^2 è detto parametro di disturbo (nuisance parameter) in quanto non è il parametro di immediato interesse. Le nostre ipotesi sono sempre

$$\begin{aligned} H_0 &: \quad \mu = \mu_0, \\ H_A &: \quad \mu > \mu_0. \end{aligned}$$

Bisogna stimare μ e σ^2 simultaneamente

$$\begin{aligned} \widehat{\mu} &= \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ (buon stimatore di } \mu), \\ \widehat{\sigma}^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2 \text{ (buon stimatore di } \sigma^2,\ E(\widehat{\sigma}^2)=\sigma^2). \end{aligned}$$

Sappiamo che se σ^2 fosse conosciuto potremmo effettuare il test abituale andando a calcolare per dato livello di significatività α il valore critico k_α:

$$P_{H_0}\left(\underbrace{\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}}_{Z\sim N(0,1)} \geq \frac{k_\alpha-\mu_0}{\sigma/\sqrt{n}}\right)=\alpha. \tag{6.1}$$

Potremmo pensare di fare la stessa cosa sostituendo a σ sconosciuto la stima $\widehat{\sigma}$. In tal caso però la quantità $\frac{\overline{X}-\mu_0}{\widehat{\sigma}/\sqrt{n}}$ consiste nel rapporto di due variabili casuali che possiamo riscrivere come segue

$$\frac{\overline{X}-\mu_0}{\widehat{\sigma}/\sqrt{n}}=\frac{\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}}{\widehat{\sigma}/\sigma}$$

- Numeratore: variabile casuale $\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$ distribuita sotto l'ipotesi nulla H_0 come una Normale standard, $\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\sim N(0,1)$.
- Denominatore: variabile casuale $\widehat{\sigma}/\sigma \underset{H_0}{\sim}$? (la distribuzione è sconosciuta!)

Sostituendo $\widehat{\sigma}$ a σ nella (6.1) e considerando quanto detto otteniamo

$$P_{H_0}\left(\underbrace{\frac{\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}}{\widehat{\sigma}/\sigma}}_{\sim\ ?} \geq \frac{k_\alpha-\mu_0}{\widehat{\sigma}/\sqrt{n}}\right)=\alpha. \tag{6.2}$$

La nostra statistica rappresenta quindi il rapporto tra 2 variabili casuali. La soluzione a questo problema è in due tappe:

1. Derivazione della distribuzione del denominatore
2. Derivazione della distribuzione del rapporto

Vedremo che la distribuzione del denominatore è data dalla distribuzione χ^2_m (Chi-2).

6.2.1 La distribuzione χ^2_m

La distribuzione Chi-2, notata χ^2_m, con m gradi di libertà è definita come la somma del quadrato di m variabili aleatorie normali standard indipendenti Z_1, Z_2, ..., $Z_m \sim N(0,1)$.

$$X = \sum_{n=1}^{m} Z_n^2 \sim \chi^2_m.$$

- Funzione di densità:

$$f(x,m) = \frac{1}{2^{\frac{m}{2}}\Gamma(\frac{m}{2})} \exp(-\frac{x}{2}) x^{\frac{m}{2}-1}$$

  dove

$$\Gamma(y) = \int_0^\infty \exp(-u) u^{y-1} du$$

- Calcolo:

$$P(X \leq x) = \int_0^x f(y,m) dy$$

  oppure in Excel o Openoffice tramite la funzione $CHIDIST(x,m)$. Attenzione: la funzione $CHIDIST(x,m) = P(V > x)$.

6.2.2 La distribuzione di Student (o distribuzione-t)

Definizione 37. Distribuzione-t. Siano $Z \sim N(0,1)$ e $X \sim \chi^2_m$ due variabili aleatorie indipendenti. Il raporto

$$\frac{Z}{\sqrt{\frac{X}{m}}} \sim t_m$$

segue una distribuzione di student a m gradi di libertà, abbreviato m g.l..

La distribuzione di Student è simile alla distribuzione Normale standard. Essa è simmetrica rispetto allo zero, ma possiede delle code più pesanti (tendono meno velocemente a zero rispetto alla distribuzione Normale). Come per la distribuzione Normale standard esistono delle tavole per il calcolo dell'

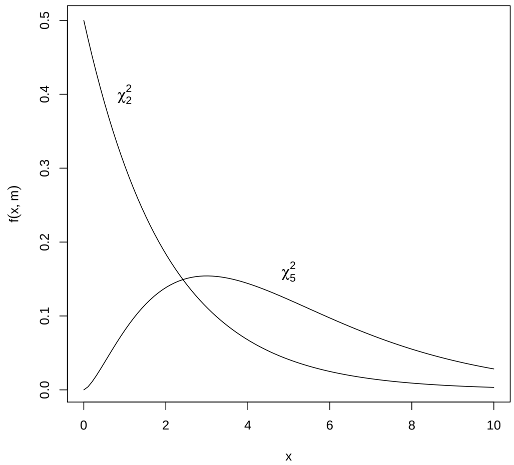


Figura 6.1: Funzione di densità della distribuzione χ^2_m, $m = 2, 5$.

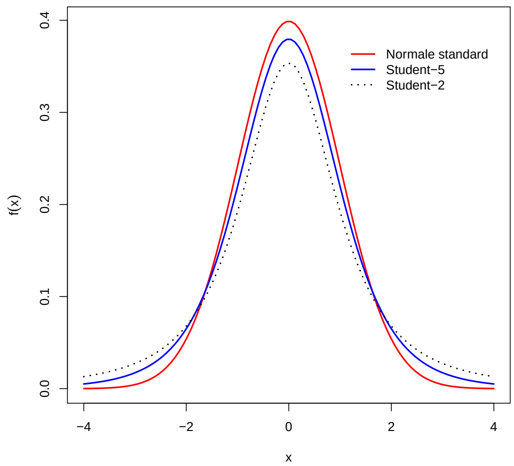


Figura 6.2: Confronto tra la densità Normale standard e la Student.

$\alpha-$quantile. A differenza della distribuzione Normale standard, per la distribuzione t occorre specificare oltre alla probabilità α anche i gradi di libertà m desiderati. Una caratteristica della distribuzione di Student consiste nel fatto che all'aumentare dei gradi di libertà essa si avvicina sempre più alla distribuzione Normale standard.

6.2.3 Il t-test

Torniamo ora al rapporto della statistica utilizzata nella (6.2)

$$\frac{\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}}{\widehat{\sigma}/\sigma} = \frac{Z \sim N(0,1)}{\sqrt{\frac{\widehat{\sigma}^2}{\sigma^2}}}.$$

Per quanto riguarda il denominatore, ed in particolare il rapporto $\widehat{\sigma}^2/\sigma^2$ notiamo che

$$\frac{\widehat{\sigma}^2}{\sigma^2} = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2} = \frac{\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\overline{X})^2}{n-1} = \frac{V \sim \chi^2_{n-1}}{n-1}.$$

Riassumendo abbiamo dunque che

$$\frac{\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}}{\widehat{\sigma}/\sigma} = \frac{N(0,1)}{\sqrt{\frac{\chi^2_{n-1}}{n-1}}} \sim t_{n-1}.$$

La nostra statistica è dunque distribuita come una variabile casuale t a $n-1$ gradi di libertà[1].

[1] È immediato vedere che la statistica

$$\begin{aligned} W &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \\ &= \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \end{aligned}$$

è una χ^2_n. Tuttavia μ è sconosciuto e va sostituito con $\overline{X}$. Poiché

$$\sum_{i=1}^{n}\left(X_i-\overline{X}\right)=0$$

abbiamo in realtà solo $n-1$ gradi di libertà nella "scelta" di $X_1-\overline{X}$, $X_2-\overline{X}$, ..., $X_n-\overline{X}$: l'ultima osservazione non può essere scelta come si vuole in quanto la somma deve essere 0.

La struttura del test di Student è dunque identica a quella del test sulla media con la sola differenza che la varianza sconosciuta va sostituita con lo stimatore $\widehat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$. Il calcolo della soglia critica k_α andrà poi modificato come segue. Sotto l'ipotesi nulla

$$\begin{aligned} P_{H_0}\left(\underbrace{\frac{\overline{X}-\mu_0}{\widehat{\sigma}/\sqrt{n}}}_{\sim t_{n-1}} \geq \frac{k_\alpha - \mu_0}{\widehat{\sigma}/\sqrt{n}}\right) &= \alpha \\ P\left(t_{n-1} \geq \frac{k_\alpha - \mu_0}{\widehat{\sigma}/\sqrt{n}}\right) &= \alpha \end{aligned}$$

Il procedimento è *identico* a quello con la Normale:

$$\begin{aligned} 1 - P\left(t_{n-1} \leq \frac{k_\alpha - \mu_0}{\widehat{\sigma}/\sqrt{n}}\right) &= \alpha \\ P\left(t_{n-1} \leq \frac{k_\alpha - \mu_0}{\widehat{\sigma}/\sqrt{n}}\right) &= 1-\alpha. \end{aligned}$$

È dunque necessario derivare l' $1-\alpha$-quantile della distribuzione di Student a $n-1$ gradi di libertà che notiamo con $t_{1-\alpha,n-1}$. Infine, uguagliando $t_{1-\alpha,n-1}$ a $\frac{k_\alpha-\mu_0}{\sigma/\sqrt{n}}$ si ottiene il valore di k_α:

$$k_\alpha = \mu_0 + t_{1-\alpha,n-1}\widehat{\sigma}/\sqrt{n}. \tag{6.3}$$

Confrontanto la (6.3) con la versione corrispondente (5.20) nel caso in cui la varianza σ^2 è conosciuta notiamo che

- la vera deviazione standard σ è stata sostituita con la deviazione standard stimata e
- l' $1-\alpha$-quantile della distribuzione Normale è stato rimpiazzato dall'$1-\alpha$-quantile della distribuzione t a $n-1$ gradi di libertà.

6.3 Altri test

6.3.1 Test su due campioni: uguaglianza di due medie

Il test che tratteremo in questo paragrafo riguarda il confronto fra due gruppi di osservazioni campionate da due popolazioni distinte. Le osservazioni della prima popolazione sono indicate con $X_1, ..., X_m$ mentre quelle della seconda con $Y_1, ..., Y_n$. Gli interi m ed n indicano la numerosità del primo e rispettivamente del secondo campione. In medicina, psicologia, economia e nelle scienze sociali capita sovente di dover determinare l'efficacia di un nuovo medicamento, la validità di un nuovo metodo d'insegnamento, l'effetto di una campagna pubblicitaria e via dicendo. Per questo motivo si è soliti raccogliere dati su due popolazioni fra loro ben distinte[2], notate appunto X e Y.

	"treatment"	controllo
medicina	nuovo farmaco	placebo
didattica	nuovo programma	vecchio programma
marketing	dopo la campagna pubb.	prima della campagna pubb.

Il parametro d'interesse in questo caso è il valore atteso delle due popolazioni. Infatti, se le due popolazioni hanno valore atteso diverso significa che l'azione da noi intrapresa ha avuto l'effetto sperato. Generalmente si è soliti scegliere quale ipotesi nulla l'ipotesi che le due popolazioni possiedono lo stesso valore atteso. La speranza è dunque quella di rifiutare l'ipotesi nulla. Studieremo ora come effettuare il test quando le due popolazioni possiedono la stessa varianza (parleremo in tal caso di omoschedasticità) e quando invece possiedono varianza diversa (caso questo di eteroschedasticità).

6.3.1.1 Caso 1: homoschedasticità delle due popolazioni

Assumiamo le seguenti ipotesi sulle due popolazioni e sui rispettivi campioni:

1. Le osservazioni $X_1, ..., X_m$ formano un campione di osservazioni *i.i.d.* (indipendenti ed identicamente distribuite) estratte dalla popolazione $X \sim N(\mu_x, \sigma_x^2)$

2. Le osservazioni $Y_1, ..., Y_n$ formano un campione di osservazioni *i.i.d.* (indipendenti ed identicamente distribuite) estratte dalla popolazione $Y \sim N(\mu_y, \sigma_y^2)$

3. La varianza della popolazione X è la stessa della popolazione Y:
$$\sigma_x^2 = \sigma_y^2 = \sigma^2.$$
Inoltre σ^2 è sconosciuta.

[2] Può capitare che la popolazione sia la stessa, ma i campionamenti siano effettuati ad istanti temporali diversi. In tal caso si desidera verificare se il valore atteso della distribuzione della popolazione sotto esame ha subito un cambiamento.

4. Per ogni X_i in $X_1, ..., X_m$ e Y_j in $Y_1, ..., Y_n$ vale che X_i e Y_j sono fra loro indipendenti.

Calcoliamo ora la varianza di $\overline{X}$, $\overline{Y}$ e $cov(\overline{X}, \overline{Y})$:

$$\begin{aligned} V(\overline{X}) &= \frac{\sigma_x^2}{m} = \frac{\sigma^2}{m} \\ V(\overline{Y}) &= \frac{\sigma_y^2}{n} = \frac{\sigma^2}{n} \end{aligned}$$

$$\begin{aligned} cov(\overline{X}, \overline{Y}) &= cov(\frac{1}{m}\sum_{i=1}^{m} X_i, \frac{1}{n}\sum_{j=1}^{n} Y_j) \\ &= \frac{1}{m} cov(\sum_{i=1}^{m} X_i, \frac{1}{n}\sum_{j=1}^{n} Y_j) \\ &= \frac{1}{m}\frac{1}{n} cov(\sum_{i=1}^{m} X_i, \sum_{j=1}^{n} Y_j) \\ &= \frac{1}{m}\frac{1}{n}\sum_{i=1}^{m} cov(X_i, \sum_{j=1}^{n} Y_j) \\ &= \frac{1}{m}\frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n} \underbrace{cov(X_i, Y_j)}_{0 \text{ per ogni } i,j} \\ &= 0 \end{aligned}$$

La varianza di $\overline{X} - \overline{Y}$ sarà dunque uguale a

$$\begin{aligned} V(\overline{X} - \overline{Y}) &= V(\overline{X}) + V(\overline{Y}) - 2cov(\overline{X}, \overline{Y}) \\ &= \frac{\sigma^2}{m} + \frac{\sigma^2}{n} = \left(\frac{1}{m} + \frac{1}{n}\right)\sigma^2 \ . \end{aligned}$$

Sotto l'ipotesi nulla la differenza $\overline{X} - \overline{Y}$ sarà distribuita secondo la legge Normale, con valore atteso 0 e varianza $\left(\frac{1}{m} + \frac{1}{n}\right)\sigma^2$. Poiché σ^2 è sconosciuto andrà stimato. Lo stimatore da utilizzarsi in questo caso è

$$\widehat{\sigma}^2 = \frac{1}{m+n-2}\left(\sum_{i=1}^{m}(x_i - \overline{x})^2 + \sum_{j=1}^{n}(y_i - \overline{y})^2\right)$$

La statistica

$$\tau = \frac{\overline{X} - \overline{Y}}{\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right)\widehat{\sigma}^2}}$$

sotto l'ipotesi nulla $H_0: \mu_x = \mu_y$ è distribuita secondo la distribuzione di Student

$$\tau \sim t_{m+n-2}$$

con $m + n - 2$ gradi di libertà.

6.3.1.2 Caso 2: eteroschedasticità delle due popolazioni

Le ipotesi riguardanti la normalità delle due popolazioni X e Y così come le ipotesi di indipendenza fra i rispettivi campioni rimangono invariate. Solo l'ipotesi numero tre riguardante l'uguaglianza delle varianze delle due popolazioni viene a cadere nel senso che non sarà necessariamente vero che $\sigma_x^2 = \sigma_y^2$. L'ipotesi nulla rimane invariata, cioè

$$H_0 : \mu_x = \mu_y.$$

In questo caso avremo

$$\begin{aligned} V(\overline{X}) &= \frac{\sigma_x^2}{m} \\ V(\overline{Y}) &= \frac{\sigma_y^2}{n} \end{aligned}$$

$$cov(\overline{X}, \overline{Y}) = 0.$$

La varianza di $\overline{X} - \overline{Y}$ sarà dunque uguale a

$$\begin{aligned} V(\overline{X} - \overline{Y}) &= V(\overline{X}) + V(\overline{Y}) - 2cov(\overline{X}, \overline{Y}) \\ &= \frac{\sigma_x^2}{m} + \frac{\sigma_y^2}{n}. \end{aligned}$$

Le varianze delle due popolazioni saranno stimate utilizzando lo stimatore usuale, ovvero

$$\begin{aligned} \widehat{\sigma}_x^2 &= \frac{1}{m-1}\sum_{i=1}^{m}(x_i - \overline{x})^2 \\ \widehat{\sigma}_y^2 &= \frac{1}{n-1}\sum_{j=1}^{n}(y_j - \overline{y})^2 \end{aligned}$$

Sotto l'ipotesi nulla, la distribuzione della statistica

$$\tau = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\widehat{\sigma}_x^2}{m} + \frac{\widehat{\sigma}_y^2}{n}}}$$

non è conosciuta. Tuttavia vale il seguente risultato:

$$\tau \text{ circa} \sim t_k$$

dove $k = \min(m-1, n-1)$.

Osservazione 33. Riassumendo, in caso di eteroschedasticità e varianze sconosciute la distribuzione della statistica τ non è nota ma è approssimabile tramite una distribuzione t a $k = \min(m-1, n-1)$ gradi di libertà (problema di Behrens-Fisher).

6.3.2 Test per la dispersione su singolo campione

È questo un test utilizzato per verificare un'ipotesi sulla varianza della popolazione. Supponiamo che il campione composto dalle n osservazioni indipendenti ed identicamente distribuite $X_1, ..., X_n$ sia stato estratto dalla popolazione $X \sim N(\mu, \sigma^2)$ con μ e σ^2 entrambi sconosciuti. Supponiamo inoltre di voler testare l'ipotesi nulla

$$H_0 : \sigma = \sigma_0$$

contro l'ipotesi alternativa

$$H_A : \sigma > \sigma_0.$$

In questo caso σ è il parametro d'interesse mentre μ è il parametro di disturbo (nel t-test visto in precedenza era σ ad essere il parametro di disturbo e μ quello d'interesse). Prendiamo quale statistica del test

$$S(X_1, ..., X_n) = \frac{\widehat{\sigma}^2}{\sigma_0^2},$$

con σ^2 stimata col "solito" stimatore

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

Un quantità osservata $\frac{\widehat{\sigma}^2}{\sigma_0^2} > 1$ sarà evidenza negativa contro l'ipotesi nulla (tanto più estremo verso destra sarà questo rapporto e tanto maggiore sarà

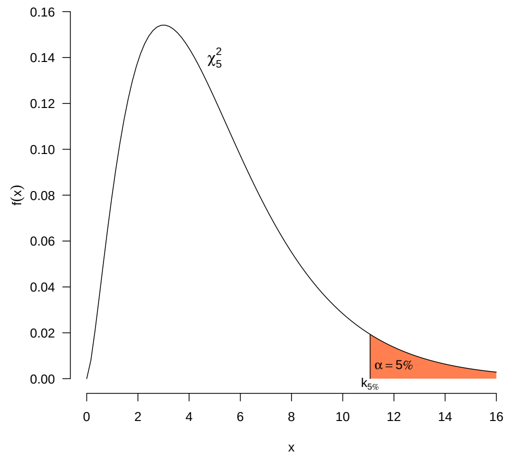


Figura 6.3: Test per la dispersione su singolo campione ($n = 6, \alpha = 5\%$).

l'evidenza contro H_0) mentre se $\frac{\widehat{\sigma}^2}{\sigma_0^2} \approx 1$ non si rifiuterà l'ipotesi nulla. Si può dimostrare che sotto l'ipotesi nulla

$$(n-1)S(X_1, ..., X_n) = (n-1)\,\frac{\widehat{\sigma}^2}{\sigma_0^2} = \sum_{i=1}^{n}\left(\frac{X_i - \overline{X}}{\sigma_0}\right)^2 \underset{H_0}{\sim} \chi^2_{n-1}$$

Costruzione della regione critica (zona di rifiuto di H_0) ad un livello di significatività α:

$$R = \{(x_1, ..., x_n) \mid (n-1)S \geq k_\alpha\}$$

con

$$P_{H_0}\left(\underbrace{(n-1)S}_{\chi^2_{n-1}} \geq k_\alpha\right) = \alpha,$$

il che equivale a dire che k_α corrisponde all'$1-\alpha$ quantile della distribuzione Chi-2 con $n-1$ gradi di libertà: $k_\alpha = \chi^2_{n-1,1-\alpha}$. Per concludere, se $(n-1)\,\frac{\widehat{\sigma}^2}{\sigma_0^2} \geq k_\alpha$ rifiuto l'ipotesi nulla H_0 altrimenti non la rifiuto.

6.3.3 Test per la dispersione su due campioni (F-Test)

Assumiamo le seguenti ipotesi sulle due popolazioni e sui rispettivi campioni:

1. Le osservazioni $X_1, ..., X_m$ formano un campione di osservazioni *i.i.d.* (indipendenti ed identicamente distribuite) estratte dalla popolazione $X \sim N(\mu_x, \sigma_x^2)$

2. Le osservazioni $Y_1, ..., Y_n$ formano un campione di osservazioni *i.i.d.* (indipendenti ed identicamente distribuite) estratte dalla popolazione $Y \sim N(\mu_y, \sigma_y^2)$

3. Per ogni X_i in $X_1, ..., X_m$ e Y_j in $Y_1, ..., Y_n$ vale che X_i e Y_j sono fra loro indipendenti.

4. μ_x, σ_x^2 e μ_y, σ_y^2 sono sconosciuti.

Desideriamo testare l'ipotesi nulla:

$$H_0 : \ \sigma_x = \sigma_y$$

contro l'ipotesi alternativa

$$H_A : \ \sigma_x > \sigma_y.$$

Stimiamo le varianze σ_x^2 e σ_y^2 delle due popolazioni con lo stimatore abituale:

$$\begin{aligned} \widehat{\sigma}_x^2 &= \frac{1}{m-1} \sum_{i=1}^{m} \left(X_i - \overline{X}\right)^2 \\ \widehat{\sigma}_y^2 &= \frac{1}{n-1} \sum_{j=1}^{n} \left(Y_j - \overline{Y}\right)^2 . \end{aligned}$$

Sotto l'ipotesi nulla la quantità

$$S(X_1, ..., X_m; Y_1, ..., Y_n) = \frac{\widehat{\sigma}_x^2}{\widehat{\sigma}_y^2} \underset{H_0}{\sim} F_{m-1,n-1}.$$

Definizione 38. Distribuzione $F_{l,k}$. Chiamiamo Distribuzione F con l gradi di libertà a numeratore e k gradi di libertà a denominatore il rapporto fra due variabili aleatorie U e V standardizzate e tali che

- $U \sim \chi_l^2$
- $V \sim \chi_k^2$
- U e V fra loro indipendenti

$$h = \frac{\frac{U}{l}}{\frac{V}{k}} \sim F_{l,k} \text{ (rapporto di due } \chi^2 \text{ standardizzate).}$$

Ragionando sulla struttura della statistica S ($\widehat{\sigma}_x^2$ a numeratore e $\widehat{\sigma}_y^2$ a denominatore) deduciamo che l'osservazione di un valore elevato costituisce evidenza a favore dell'ipotesi alternativa e contro l'ipotesi nulla. Si tratta ancora una volta di un test unilaterale destro. La costruzione della regione critica (zona di rifiuto di H_0) ad un livello di significatività α è la seguente:

$$R = \left\{ (x_1, ..., x_n, y_1, ..., y_n) \mid \frac{\widehat{\sigma}_x^2}{\widehat{\sigma}_y^2} \geq k_\alpha \right\}$$

con

$$P_{H_0} \left(\underbrace{\frac{\widehat{\sigma}_x^2}{\widehat{\sigma}_y^2}}_{F_{m-1,n-1}} \geq k_\alpha \right) = \alpha \text{ cioè } k_\alpha = F_{m-1,n-1;1-\alpha}.$$

k_α è dunque l'$1-\alpha$ quantile della distribuzione F con $m-1$ e $n-1$ gradi di libertà.

6.3.4 Test di conformità

Il test di conformità ("goodness-of-fit" test) non è un test su un'ipotesi parametrica. Infatti, con il test di conformità si desidera verificare se la distribuzione della popolazione dalla quale è estratto il campione corrisponde ad una distribuzione data (ad esempio Normale, di Pareto, esponenziale, ecc.) che indicheremo qui con F.

Esempio 48. Distribuzione dei redditi. Vogliamo vedere se la distribuzione dei dati $X_1, ..., X_n$ è compatibile con l'ipotesi che la popolazione $X \sim F$, dove in questo caso F è la distribuzione di Pareto la cui funzione di densità sappiamo essere uguale a

$$f(x) = k \frac{x_0^k}{x^{k+1}} \quad \text{per } x \geq x_0,$$

e x_0 e k sono due parametri supposti conosciuti. La funzione di ripartizione di X è data da

$$F(x) = 1 - \left(\frac{x}{x_0} \right)^{-k}.$$

Al fine di costruire il test dovremo, partendo dalle osservazioni, costruire degli intervalli (o classi) di reddito, che noteremo con $c_1, c_2, ..., c_K$.

Classe	Intervallo Mia Fr.	# osservazioni	# osservazioni atteso sotto Ho	Deviazioni	$\frac{(O_i-E_i)^2}{E_i}$
c_1	$[0,30)$	O_1	$n\ p_1 = E_1$	$O_1 - E_1$	$\frac{(O_1-E_1)^2}{E_1}$
c_2	$[30,50)$	O_2	$n\ p_2 = E_2$	$O_2 - E_2$	$\frac{(O_2-E_2)^2}{E_2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
c_k	$[100,120)$	O_k	$n\ p_k = E_k$	$O_k - E_k$	$\frac{(O_k-E_k)^2}{E_k}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
c_K	$[200,+)$	O_K	$n\ p_K = E_K$	$O_K - E_K$	$\frac{(O_K-E_K)^2}{E_K}$
Totale		$n = \sum_{k=1}^K O_k$	n	0	$\chi^2 = \sum_{k=1}^K \frac{(O_k-E_k)^2}{E_k}$

Le probabilità p_k sono le probabilità teoriche di osservare una realizzazione nell'intervallo corrispondente, ovvero

$$p_k = P_{H_0}(l_k \leq X < l_{k+1}).$$

Su n estrazioni indipendenti il numero atteso di osservazioni nell'intervallo $[l_k, l_{k+1})$, notato E_k, sarà semplicemente $n\ p_k$. Il test consiste nel confrontare le frequenze assolute empiriche con quelle teoriche. Se le differenze sono "grandi", avremo evidenza contro l'ipotesi nulla. Per misurare tali discrepanze si calcolano gli scarti al quadrato tra le frequenze assolute empiriche e quelle teoriche, debitamente "standardizzate". Al fine di eseguire correttamente il test occorre ancora una volta conoscere la distribuzione della statistica $S = \sum_{k=1}^K \frac{(O_k-E_k)^2}{E_k}$. Fortunatamente è possibile dimostrare che sotto l'ipotesi nulla la quantità

$$\sum_{k=1}^{K} \frac{(O_k - E_k)^2}{E_k} \sim \chi^2_{K-1.}$$

Osservazione 34. Affinché la statistica così costruita sia valida occorre selezionare le classi in maniera tale che le frequenze teoriche E_i siano maggiori o uguali a 5. Per tale motivo è talvolta necessario ridurre il numero di classi aumentando l'ampiezza degli intervalli.

Esempio 49. Osserviamo n incidenti della circolazione. Ci chiediamo se la distribuzione degli incidenti è uniforme rispetto ai mesi dell'anno. Il numero di classi K sarà quindi uguale a 12. Definiamo con O_k := "# incidenti osservati nel mese k". La statistica

$$\chi^2 = \sum_{k=1}^{12} \frac{\left(O_k - \frac{n}{12}\right)^2}{\frac{n}{12}}$$

è distribuita secondo una χ^2_{11}.

Scelto il valore critico $k_\alpha = \chi^2_{11,1-\alpha}$ al livello di significatività desiderato, rigetteremo l'ipotesi nulla se

$$\sum_{k=1}^{12} \frac{\left(O_k - \frac{n}{12}\right)^2}{\frac{n}{12}} \geq k_\alpha$$

e non la rigetteremo altrimenti.

6.3.5 Test d'indipendenza

Questo test verrà presentato alla lavagna.