



POLITECNICO DI MILANO

Statistica

Appunti

Stefano Invernizzi

Anno accademico 2010-2011

Sommario

Introduzione al corso.....	5
La statistica	5
Schema tipico di raccolta dei dati.....	5
Ripasso di Calcolo delle Probabilità.....	6
Variabile aleatoria	6
Funzione di ripartizione.....	6
Variabili aleatorie discrete.....	6
Variabili aleatorie discrete notevoli: binomiale	8
Variabili aleatorie discrete notevoli: geometrica	8
Variabili aleatorie continue	9
Indipendenza di variabili aleatorie	10
Distribuzione uniforme.....	11
Densità di variabili aleatorie derivate.....	11
Distribuzione esponenziale	12
Il modello di Weibull.....	14
Il modello gaussiano	15
Il teorema centrale del limite	17
Funzione generatrice di momenti	18
La distribuzione gamma	20
La distribuzione chi-quadro.....	22
La densità T-student	23
Stima di media e varianza di distribuzioni di probabilità	24
Terminologia.....	24
Valutazione di uno stimatore	24
La media campionaria	25
La varianza campionaria.....	26
Media e varianza campionaria nel caso gaussiano	27
Intervalli di confidenza	28
Metodi per la stima dei parametri: stima puntuale	33
Metodologie di stima dei parametri.....	33
Il metodo dei momenti.....	34
Il metodo di massima verosimiglianza	37
Confronto tra i due metodi.....	39

Ricerca dello stimatore ottimo	40
Disuguaglianza di Fréchet-Cramer-Rao	42
Proprietà degli stimatori di massima verosimiglianza.....	46
Il metodo della quantità pivotale	51
La verifica di ipotesi	53
Introduzione	53
I concetti fondamentali della verifica d'ipotesi	53
Errori.....	55
Il p-value	56
Lemma di Neyman-Pearson	57
Verifica d'ipotesi su popolazione gaussiana: gli Z-test e i T-test.....	59
Verifica d'ipotesi su popolazione gaussiana: test chi-quadro sulla varianza (metodo degli IC) – caso a media incognita	63
Verifica d'ipotesi su popolazione gaussiana: test chi-quadro sulla varianza (metodo degli IC) – caso a media nota.....	65
Test sui dati accoppiati – test di omogeneità sulle medie	66
Test sui dati accoppiati – test di indipendenza (dati gaussiani)	68
Test sui dati accoppiati: Test di Wilcoxon (omogeneità).....	70
Test di Wilcoxon-Mann-Whitney (omogeneità dati non accoppiati).....	71
Test di omogeneità su campioni gaussiani indipendenti	73
Test chi-quadro di Pearson – per il buon adattamento (goodness of fit)	78
Test chi-quadro di indipendenza	81
Test di Kolmogorov-Smirnov (buon adattamento)	82

Introduzione al corso

La statistica

La statistica può essere definita come "l'arte di imparare dai dati". Essa consiste quindi nell'individuare delle tecniche che facciano in modo che i dati ci forniscano le informazioni di cui abbiamo bisogno (vogliamo "far parlare i dati").

A tale scopo, bisogna per prima cosa essere in grado di sintetizzare i dati e di descriverli mediante quelle che vengono appunto chiamate *statistiche*, e che sono semplicemente delle descrizioni dei dati sottoforma, ad esempio, di tabelle, grafi, medie, Dopodiché, è necessario trarre delle conclusioni sui dati stessi.

Una volta ottenute tali conclusioni, bisogna anche domandarsi quanto le conclusioni raggiunte basandosi su un campione di dati siano realmente affidabili. A tale scopo occorre tener conto dell'incertezza dei dati: ad esempio, se si stanno rilevando i dati relativi all'occupazione di memoria di un web server, bisogna tener conto del fatto che rilevare i dati in un giorno diverso produrrebbe dati diversi. Un modo per "quantificare" l'incertezza è quello di utilizzare dei modelli probabilistici, e proprio per tale ragione non possiamo prescindere nello studio della statistica dal calcolo delle probabilità.

Schema tipico di raccolta dei dati

Un esempio

Analizziamo uno schema tipico per la raccolta dei dati, considerando ancora come esempio il web server al quale abbiamo accennato nel precedente paragrafo. Ipotizziamo per semplicità che tutte le azioni eseguite in un certo istante sul web server siano indipendenti tra loro e che tutte le operazioni siano dello stesso tipo (quindi richiedono lo stesso spazio di memoria).

Formalizzazione

La situazione illustrata nel precedente esempio consiste nel ripetere n volte uno stesso esperimento in condizioni analoghe. Avremo quindi:

$$\text{Risultati} := x_1, x_2, \dots, x_n$$

Dove i dati x_1, x_2, \dots, x_n sono sempre dei valori numerici. Questo è lo schema tipico di raccolta dei dati che adotteremo in seguito.

Si noti che l'esperimento eseguito non è deterministico, perciò i valori ottenuti come risultato, prima di eseguire gli esperimenti stessi, possono essere modellizzati mediante delle variabili aleatorie. In altri termini, i dati sono realizzazioni di n variabili aleatorie (va)

$$X_1, X_2, \dots, X_n$$

Come già affermato in maniera intuitiva, tali variabili aleatorie devono essere indipendenti tra loro:

$$X_1, X_2, \dots, X_n \quad i. i. d$$

Dove *i. i. d.* sta per *indipendenti identicamente distribuite*. Una famiglia di variabili aleatorie indipendenti e identicamente distribuite è detta **campione casuale**.

Variabili aleatorie indipendenti

Rimane però da definire il concetto di variabile aleatoria indipendente. Tale concetto, già studiato durante il corso di Calcolo delle Probabilità, è così formalizzato:

Le variabili X_1, X_2, \dots, X_n sono indipendenti se, per ogni insieme A tale che si possano calcolare tutte le probabilità che compaiono nella seguente espressione, si ha:

$$P(X_1 \in A | X_2 \in A, X_3 \in A, \dots, X_n \in A) = P(X_1 \in A)$$

Si noti che la definizione appena fornita è in realtà solamente di tipo concettuale: non si tratta infatti di una definizione operativa.

Ripasso di Calcolo delle Probabilità

Come si osserva dai pochi concetti finora introdotti, lo studio della statistica richiede l'utilizzo di molti concetti del Calcolo delle Probabilità. Avviamo quindi a questo punto un rapido ripasso di tale disciplina, mirato solamente a quei concetti che è necessario utilizzare in Statistica.

Variabile aleatoria

Una variabile aleatoria X è un numero casuale (cioè del quale non si conosce a priori il valore) del quale è possibile calcolare la probabilità:

$$P(X \leq x) \quad \forall x \in \mathbb{R}$$

Funzione di ripartizione

Funzione di ripartizione

La funzione $F(x) = P(X \leq x)$, definita per ogni x appartenente ad \mathbb{R} prende il nome di *funzione di ripartizione*.

Importanza della funzione di ripartizione

La funzione di ripartizione è un concetto di fondamentale importanza, in quanto ci consente di calcolare le probabilità di tutti gli eventi di interesse collegati alla variabile casuale. Tali eventi, che sono un'infinità numerabile, saranno di varie tipologie:

$$X \in (a, b), X \in [a, b], X \in (a, b], X \in [a, b), X = x$$

Se conosciamo la funzione di ripartizione di X per tutti i valori reali x , allora possiamo calcolare a partire da essa le probabilità di tutti gli eventi delle tipologie sopra elencate. Ad esempio:

$$P(X \in (a, b]) = F(b) - F(a)$$

In altri casi il procedimento è leggermente più complesso, perché è necessario un passaggio al limite, ma in ogni caso la funzione di ripartizione ci fornisce tutte le informazioni necessarie.

Variabili aleatorie discrete

Le variabili aleatorie possono essere classificate sulla base delle modalità che X può assumere (cioè dei valori che la variabile aleatoria stessa ammette). Iniziamo ora il ripasso di una prima categoria: quella delle variabili aleatorie discrete.

Variabile aleatoria discreta

Diciamo che X è una variabile aleatoria discreta se X è una variabile aleatoria che può assumere al più un'infinità numerabile di valori.

Funzione di densità di una variabile aleatoria discreta

Data una variabile aleatoria discreta X a valori in un insieme $S = \{a_1, a_2, \dots\}$ (detto **supporto**), chiamiamo *funzione di densità* di X la funzione così definita:

$$f(x) = \begin{cases} P(X = x) & \text{se } x \in S \\ 0 & \text{altrimenti} \end{cases}$$

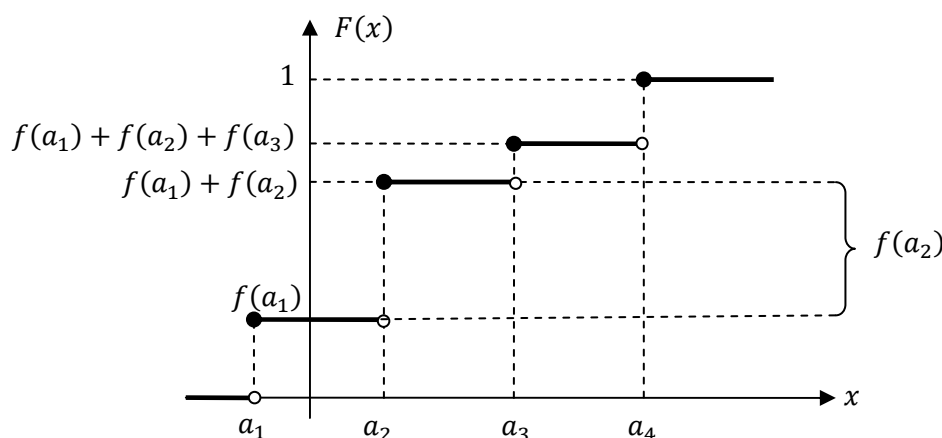
Si noti che spesso le funzioni di densità verranno indicate con $f(x, \vartheta_1, \dots, \vartheta_n)$ e non con la simbologia $f(x)$: ciò accadrà sempre all'interno di questo corso perché, come avremo modo di vedere, le funzioni di densità dipendono da diversi parametri e, in statistica, i parametri dai quali dipende la densità possono essere tutti o in parte non esattamente noti a priori.

Valgono le seguenti proprietà:

1. $f(x, \vartheta_1, \dots, \vartheta_n) \geq 0 \quad \forall x, \vartheta_1, \dots, \vartheta_n$
2. $\sum_{x \in S} f(x, \vartheta_1, \dots, \vartheta_n) = 1$
3. $P(X \in A) = \sum_{x \in A} f(x, \vartheta_1, \dots, \vartheta_n)$
4. $F(x_0) = \sum_{x \leq x_0} f(x, \vartheta_1, \dots, \vartheta_n)$

Andamento tipico della funzione di ripartizione

L'andamento tipico della funzione di ripartizione di una variabile aleatoria discreta è indicato nel grafico seguente:



Possiamo quindi affermare che la funzione di ripartizione di una variabile aleatoria discreta è costante a tratti, monotona non decrescente, continua da destra e che tale funzione tende a zero per $x \rightarrow -\infty$ e tende a 1 per $x \rightarrow +\infty$.

Inoltre, se a_1 e a_2 sono due valori tali che a_2 è il numero immediatamente successivo ad a_1 tra quelli appartenenti ad S , possiamo scrivere:

$$F(a_2) - F(a_1) = f(a_2)$$

In altri termini, dalla funzione di ripartizione è possibile ottenere tutti i valori di densità.

Media

La media di una variabile aleatoria discreta X a valori in S è definita come:

$$E(X) = \sum_{x \in S} [x \cdot f(x)]$$

Condizione necessaria per poter definire la media di X è che si abbia:

$$\sum_{x \in S} [|x| \cdot f(x)] < +\infty$$

Di fatto però non sarà mai necessario verificare tale condizione nelle applicazioni che andremo ad analizzare.

Varianza

La varianza di una variabile aleatoria discreta X a valori in S è definita come:

$$VAR(X) = \sum_{x \in S} [(x - E(X))^2 \cdot f(x)] = E[(x - E(X))^2]$$

Proprietà della media e della varianza

1. $E(a) = a, a \in \mathbb{R}$, costante
2. $E(aX + b) = a \cdot E(X) + b$
3. Se $a \leq X \leq b$, allora $a \leq E(X) \leq b$
4. Se $X \leq Y$, allora $E(X) \leq E(Y)$
5. $E(X + Y) = E(X) + E(Y)$
6. $VAR(aX + b) = a^2 \cdot VAR(X)$
7. $X = \text{costante} \Leftrightarrow VAR(X) = 0$
8. $VAR(X + Y) = VAR(X) + VAR(Y) + 2 \cdot E[(X - E(X))(Y - E(Y))] =$
 $= VAR(X) + VAR(Y) + 2 \cdot COV(X, Y)$
9. $VAR(X) = E(X^2) - E^2(X) \rightarrow E(X^2) = E^2(X) + VAR(X)$
10. $VAR(X) \geq 0$

Momenti

1. Momento primo

La media di una variabile aleatoria X viene detta *momento primo di X* .

2. Momento secondo

La media della variabile X^2 viene invece detta *momento secondo di X* .

E così via.

Variabili aleatorie discrete notevoli: binomiale

Significato

Una variabile aleatoria X è binomiale se è del tipo:

$X = \text{"n° di successi su } n \text{ ripetizioni indipendenti di una prova con probabilità di successo } p\text{"}$

Si noti che è necessario che tutte le prove siano indipendenti tra loro e che la probabilità di successo sia uguale ad ogni prova. Una variabile binomiale si indica:

$$X \sim \text{Bin}(n, p)$$

Funzione di densità

La funzione di densità di una variabile aleatoria binomiale è data dalla formula:

$$f(k, p) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}, \quad k = 0, 1, \dots, n$$

Media e varianza

- La media della binomiale è: $E(X) = n \cdot p$
- La varianza della binomiale è: $VAR(X) = n \cdot p \cdot (1 - p)$

Variabili aleatorie discrete notevoli: geometrica

Significato

Una variabile aleatoria X è geometrica se è del tipo:

$X = \text{"n° di prove indep., con probabilità di successo } p, \text{ necessarie per ottenere il 1° successo"}$

Scriviamo allora:

$$X \sim \text{Geom}(p), \quad p \in (0, 1)$$

Funzione di densità

La funzione di densità di una variabile aleatoria geometrica è data dalla formula:

$$f(k, p) = (1 - p)^{k-1} \cdot p, \quad k = 1, 2, \dots$$

Media e varianza

- La media della geometrica è: $E(X) = \frac{1}{p}$
- La varianza della geometrica è: $VAR(X) = \frac{1 - p}{p^2}$

Variabili aleatorie continue

Nel caso in cui i dati rilevati siano delle misurazioni effettuate nel continuo, chiaramente non è possibile rappresentarli per mezzo di variabili aleatorie discrete: servono perciò delle variabili aleatorie che assumano valori reali all'interno di un intervallo continuo. Tali variabili sono dette *variabili aleatorie continue* e, in particolare, ci occuperemo solamente di studiare *variabili aleatorie assolutamente continue* (dal momento che studieremo solo tale tipologia di variabili, utilizzeremo i due termini come sinonimi, anche se in realtà hanno significati leggermente diversi).

Variabile aleatoria continua

Diciamo che X è una variabile aleatoria continua se X è una variabile aleatoria la cui funzione di ripartizione è del tipo:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt$$

Dove $f: \mathbb{R} \rightarrow [0, +\infty)$ è una funzione integrabile, detta *densità di probabilità* di X . L'integrale della funzione di densità su tutto \mathbb{R} è sempre pari ad 1:

$$\int_{-\infty}^{+\infty} f(t) dt = 1$$

Analogie e differenze rispetto alle variabili aleatorie continue

Si nota facilmente che in entrambi i casi la funzione di ripartizione è definita allo stesso modo; inoltre, sia per le variabili aleatorie continue che per quelle discrete è definita una *densità di probabilità*, che però assume significati ben diversi: mentre nel caso delle variabili aleatorie discrete tale funzione assume effettivamente il significato di una probabilità, nelle variabili aleatorie continue non è così (basti pensare che il suo codominio non è limitato ad 1, perciò f può assumere qualsiasi valore positivo).

Proprietà della funzione di ripartizione

Per la funzione di ripartizione di una variabile aleatoria continua valgono proprietà molto simili a quelle viste per la funzione di ripartizione di una variabile aleatoria discreta:

1. Per $x \rightarrow -\infty$, la funzione di ripartizione tende a zero.

$$\lim_{x \rightarrow -\infty} F_X(x) = 0$$

2. Per $x \rightarrow +\infty$, la funzione di ripartizione tende ad uno.

$$\lim_{x \rightarrow +\infty} F_X(x) = 1$$

3. La funzione di ripartizione è continua ed è monotona crescente.

Probabilità che la variabile aleatoria assuma valori in un certo insieme

In maniera molto intuitiva possiamo ricavare che:

$$P(a < X \leq b) = \int_a^b f(x) dx = F_X(b) - F_X(a)$$

Tramite semplici calcoli si ricava inoltre che:

$$P(X = k) = 0, \quad \forall k \in \mathbb{R}$$

E quindi possiamo concludere che:

$$P(a < X < b) = P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = \int_a^b f(x) dx = F_X(b) - F_X(a)$$

Questo risultato può essere interpretato da un punto di vista statistico: infatti, ogni rilevazione di grandezze continue deve essere interpretata non come un valore preciso, ma come l'indicazione dell'appartenenza del valore vero della grandezza ad un certo intervallo di valori, che dipende dalla sensibilità dello strumento di misura utilizzato.

Legame tra funzione di ripartizione e funzione di densità

Sulla base delle definizioni date, la funzione di ripartizione viene calcolata a partire da quella di densità, perciò è chiaro che una conoscenza completa della prima è sufficiente per ottenere tutte le informazioni che riguardano la seconda.

Vogliamo a questo punto domandarci se sia vero anche il contrario. A tale scopo, è sufficiente osservare che la funzione di ripartizione potrebbe non essere derivabile in tutti i punti reali.

Tuttavia, l'insieme dei punti nei quali la funzione di ripartizione non è derivabile è necessariamente finito o al più numerabile, perciò è possibile calcolare la derivata della funzione di ripartizione in tutti i punti nei quali essa è derivabile, completando poi tale funzione in modo che valga zero in tutti i restanti punti.

Media di una variabile aleatoria continua

La media di una variabile aleatoria X è definita come:

$$\mu = E(X) = \int_{-\infty}^{+\infty} x \cdot f(x) dx$$

A patto che l'integranda risulti assolutamente integrabile:

$$\int_{-\infty}^{+\infty} |x| \cdot f(x) dx < +\infty$$

Varianza di una variabile aleatoria continua

Come nel caso delle variabili aleatorie discrete, la varianza è definita nel modo seguente:

$$VAR(X) = E[(X - E(X))^2] = \int_{-\infty}^{+\infty} (x - \mu)^2 \cdot f(x) dx$$

Naturalmente, sempre a patto che si abbia:

$$\int_{-\infty}^{+\infty} (x - \mu)^2 \cdot f(x) dx < +\infty$$

Proprietà di media e varianza

Le proprietà della media e della varianza di variabili aleatorie continue sono sostanzialmente le stesse che abbiamo enunciato nel caso di variabili aleatorie discrete: proprio per tale motivo esse sono state elencate utilizzando il simbolo $E(X)$, in modo tale che non risulti a questo punto necessario elencarle nuovamente.

Indipendenza di variabili aleatorie

Abbiamo già dato una prima definizione di indipendenza tra variabili aleatorie. A questo punto possiamo però dare una definizione operativa dell'indipendenza tra due variabili aleatorie X ed Y . Il procedimento che andremo a descrivere può poi essere facilmente generalizzato ad un numero qualsiasi di variabili.

Densità congiunta

Per prima cosa dobbiamo introdurre il concetto di *densità congiunta* tra due variabili aleatorie X ed Y :

$$f(x, y): \mathbb{R}^2 \rightarrow \mathbb{R}$$

1. Nel caso discreto, la densità congiunta è definita come:

$$f(x, y) = P(X = x, Y = y)$$

2. Nel caso continuo, la densità congiunta è definita come quella funzione $f(x, y)$ integrabile su \mathbb{R}^2 e tale che si abbia:

$$P(X \leq x, Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y f(s, t) dt ds$$

Indipendenza

Diremo allora che X ed Y sono indipendenti se la densità congiunta fattorizza nel prodotto delle due densità marginali, ovvero:

$$f(x, y) = f_X(x) \cdot f_Y(y)$$

Distribuzione uniforme

Definizione

Una variabile aleatoria continua X è detta *uniforme tra a e b* ($b > a$) se ha una distribuzione di probabilità nulla fuori dall'intervallo $[a, b]$ e costante pari a $\frac{1}{b-a}$ all'interno di tale intervallo:

$$f(x, a, b) = \begin{cases} \frac{1}{b-a} & \text{se } a \leq x \leq b, \\ 0 & \text{altrimenti} \end{cases} \quad b > a$$

La stessa funzione può essere rappresentata per mezzo della *funzione indicatore* così definita:

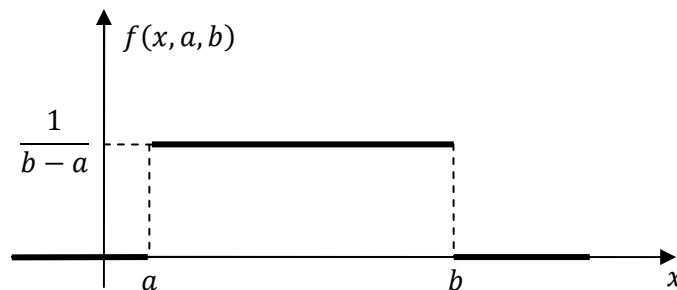
$$\mathbb{I}_A(x) = \begin{cases} 0 & \text{se } x \notin A \\ 1 & \text{se } x \in A \end{cases}$$

Dove A è un qualsiasi sottoinsieme di \mathbb{R} . Utilizzando l'indicatore si ottiene un'espressione più compatta:

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

La variabile uniforme si indica con:

$$X \sim \mathcal{U}(a, b)$$



È molto semplice verificare che la media e la varianza della distribuzione di probabilità si ottengono calcolando:

$$E[X] = \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{a+b}{2}$$

$$E[X^2] = \int_a^b \frac{x^2}{b-a} dx = \frac{1}{b-a} \left[\frac{x^3}{3} \right]_a^b = \frac{b^3 - a^3}{3(b-a)} = \frac{(b-a)(b^2 + ab + a^2)}{3(b-a)} = \frac{b^2 + ab + a^2}{3}$$

$$\begin{aligned} \text{VAR}[X] &= E[X^2] - E^2[X] = \frac{b^2 + ab + a^2}{3} - \frac{a^2 + 2ab + b^2}{4} = \frac{4b^2 + 4ab + 4a^2 - 3a^2 - 6ab - 3b^2}{12} = \\ &= \frac{b^2 - 2ab + a^2}{12} = \frac{(b-a)^2}{12} \end{aligned}$$

Densità di variabili aleatorie derivate

Sia data una variabile aleatoria X con distribuzione di probabilità $f_X(x)$ e supporto S . Sia inoltre data una funzione $g(X)$ invertibile sul supporto di X , con inversa g^{-1} . Allora:

$$Y = g(X)$$

è una variabile aleatoria la cui densità di probabilità è data dalla formula:

$$f_Y(y) = \left| \frac{d}{dy} g^{-1}(y) \right| \cdot f_X(g^{-1}(y))$$

Distribuzione esponenziale

Definizione

Una variabile aleatoria continua X è detta *esponenziale* (o *esponenziale negativa*) se ha una distribuzione di probabilità del tipo:

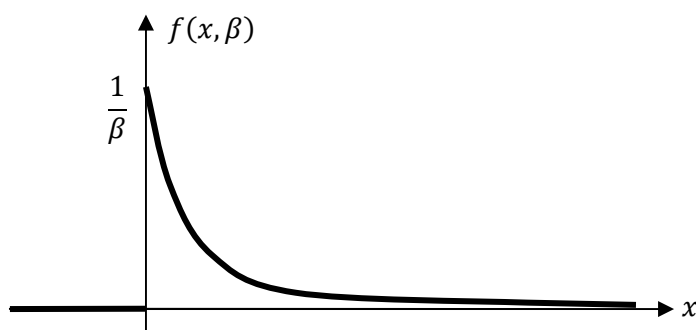
$$f(x, \beta) = \begin{cases} \frac{1}{\beta} e^{-\frac{x}{\beta}} & \text{se } x > 0, \\ 0 & \text{se } x \leq 0 \end{cases}, \quad \beta > 0$$

Utilizzando l'indicatore si ottiene un'espressione più compatta:

$$f(x, \beta) = \mathbb{I}_{x>0}(x) \cdot \frac{1}{\beta} e^{-\frac{x}{\beta}}$$

Grafico della funzione di densità

L'andamento della $f(x, \beta)$ è perciò del tipo:



Per dire che una variabile aleatoria X ha funzione di densità esponenziale con parametro β si scrive:

$$X \sim \mathcal{E}(\beta)$$

Funzione di ripartizione

La funzione di ripartizione, secondo la definizione data, verrà calcolata nel modo seguente:

1. Se $x \leq 0$:

$$F_X(x) = 0$$

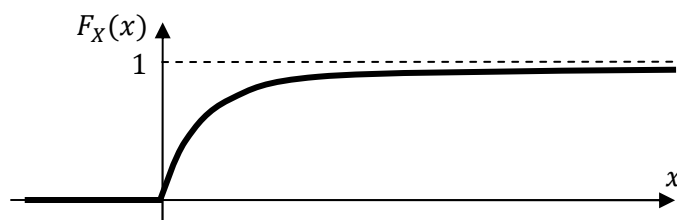
2. Se $x > 0$:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt = \int_0^x \frac{1}{\beta} e^{-\frac{t}{\beta}} dt = -e^{-\frac{t}{\beta}} \Big|_0^x = 1 - e^{-\frac{x}{\beta}}$$

Ovvero:

$$F_X(x) = \left(1 - e^{-\frac{x}{\beta}}\right) \cdot \mathbb{I}_{x>0}(x)$$

Il grafico della funzione è dunque quello rappresentato nella figura seguente:



Osserviamo quindi che la funzione di ripartizione possiede tutte le proprietà precedentemente elencate.

Media e varianza

1. Calcoliamo la media dell'esponenziale:

$$E(X) = \int_{-\infty}^{+\infty} x \cdot f(x, \beta) dx = \int_0^{+\infty} x \cdot \frac{1}{\beta} e^{-\frac{x}{\beta}} dx$$

Integrando per parti:

$$E(X) = -xe^{-\frac{x}{\beta}} \Big|_0^{+\infty} + \int_0^{+\infty} e^{-\frac{x}{\beta}} dx = 0 + \int_0^{+\infty} e^{-\frac{x}{\beta}} dx = \int_0^{+\infty} \frac{1}{\beta} e^{-\frac{x}{\beta}} dx$$

Ricordando che calcolare:

$$\int_0^{+\infty} \frac{1}{\beta} e^{-\frac{x}{\beta}} dx$$

Equivale di fatto ad integrare su tutto l'insieme dei reali la funzione di densità dell'esponenziale, e sapendo che tale integrale deve essere sempre pari ad uno, per ogni variabile aleatoria continua:

$$E(X) = \int_0^{+\infty} e^{-\frac{x}{\beta}} dx = \beta \int_0^{+\infty} \frac{1}{\beta} e^{-\frac{x}{\beta}} dx = \beta$$

2. Calcoliamo ora la varianza dell'esponenziale, utilizzando la proprietà:

$$VAR(X) = E(X^2) - E^2(X)$$

A tale scopo, dobbiamo calcolare per prima cosa il momento secondo:

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 \cdot f(x, \beta) dx = \int_0^{+\infty} x^2 \cdot \frac{1}{\beta} e^{-\frac{x}{\beta}} dx$$

Mediante un'integrazione per parti, che omettiamo per brevità, otteniamo:

$$E(X^2) = 2\beta^2$$

Perciò:

$$VAR(X) = E(X^2) - E^2(X) = 2\beta^2 - \beta^2 = \beta^2$$

Proprietà dell'assenza di memoria

La distribuzione di probabilità esponenziale è l'unica distribuzione continua che gode della proprietà di assenza di memoria. Ciò significa che:

$$P(X \geq x_0 + t | X \geq x_0) = P(X \geq t)$$

Usi della variabile aleatoria esponenziale

Le variabili aleatorie esponenziali vengono utilizzate per modellare i tempi di vita (o di guasto) di apparecchiature di vario genere (o anche per i tempi di vita dei pazienti, in ambito medico). In particolare si considera un'apparecchiatura che inizialmente funziona correttamente e che non è soggetta né ad usura, né a rodaggio, e la si monitora continuamente; la variabile aleatoria rappresenta poi l'istante nel quale l'apparecchiatura si guasta. L'assenza di fenomeni di usura e rodaggio è rappresentata matematicamente per mezzo della proprietà dell'assenza di memoria, che abbiamo appena enunciato.

Il modello di Weibull

Proviamo ora ad introdurre una variabile aleatoria che rappresenti il tempo di vita di un'apparecchiatura tenendo conto dei fenomeni di usura o del suo rodaggio.

La variabile aleatoria può essere ottenuta semplicemente come trasformazione continua della variabile esponenziale: trattandosi della trasformazione continua di una variabile continua, si tratterà ancora di una variabile continua. In particolare, la trasformazione necessaria è di tipo esponenziale:

$$Y = X^c, \quad \text{con } X \sim \mathcal{E}(\beta) \text{ e } c > 0$$

Avremo quindi:

$$F_Y(y) = P(Y \leq y) = P(X^c \leq y)$$

Possiamo ora distinguere due sottocasi:

1. Se $y \leq 0$:

$$F_Y(y) = 0$$

2. Se invece $y > 0$, siccome tutte le grandezze in gioco sono positive, avremo:

$$F_Y(y) = P(X^c \leq y) = P\left(X \leq y^{\frac{1}{c}}\right) = F_X\left(y^{\frac{1}{c}}\right)$$

Anche senza proseguire oltre nei calcoli (che, come vedremo, sarebbe inutile per i nostri scopi), abbiamo:

$$F_Y(y) = \begin{cases} 0 & \text{se } y \leq 0 \\ F_X\left(y^{\frac{1}{c}}\right) & \text{se } y > 0 \end{cases}$$

Ciò che ci interessa è invece determinare la funzione di densità di Y , che possiamo calcolare come derivata di F_Y :

1. Se $y < 0$, naturalmente avremo:

$$\frac{dF_Y}{dy} = 0$$

2. Se $y > 0$, possiamo calcolare la derivata mediante la regola della derivazione di funzione composta:

$$\frac{dF_Y}{dy} = \frac{dF_X\left(y^{\frac{1}{c}}\right)}{dy} = \frac{1}{c} \cdot y^{\frac{1}{c}-1} \cdot F'_X\left(y^{\frac{1}{c}}\right) = \frac{1}{c} \cdot y^{\frac{1}{c}-1} \cdot f_X\left(y^{\frac{1}{c}}\right) = \frac{1}{c\beta} \cdot y^{\frac{1-c}{c}} \cdot e^{-\frac{y^{\frac{1}{c}}}{\beta}}$$

In conclusione, completando la funzione di densità così trovata, abbiamo:

$$f_Y(y) = \begin{cases} 0 & \text{se } y \leq 0 \\ \frac{1}{c\beta} \cdot y^{\frac{1-c}{c}} \cdot e^{-\frac{y^{\frac{1}{c}}}{\beta}} & \text{se } y > 0 \end{cases}$$

Questo risultato poteva anche essere ottenuto più semplicemente utilizzando la formula relativa alla funzione di densità di una variabile aleatoria derivata.

Si nota che:

1. Se $c > 1$, Y rappresenta il tempo di usura di un'apparecchiatura con rodaggio.
2. Se $c = 1$, Y è l'esponenziale già analizzata (cioè non si tiene conto né di rodaggi, né di usure).
3. Se $c < 1$, Y rappresenta il tempo di usura di un'apparecchiatura soggetta ad usura.

Il modello gaussiano

Variabile aleatoria gaussiana (o normale)

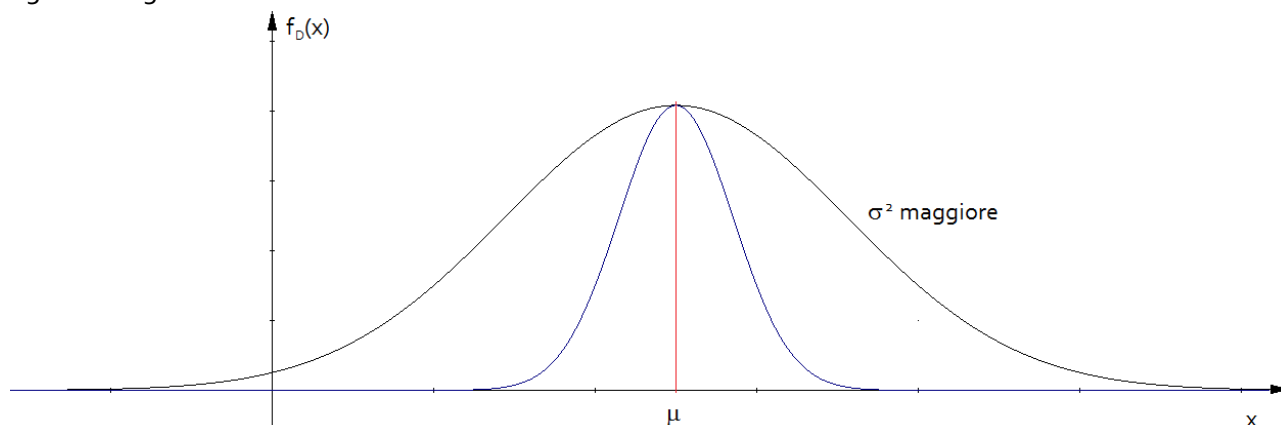
Diciamo che una variabile aleatoria X è normale di media μ e varianza σ^2 , e scriviamo:

$$X \sim \mathcal{N}(\mu, \sigma^2), \quad \mu, \sigma^2 \in \mathbb{R}, \sigma^2 > 0$$

Se la distribuzione di probabilità di X è:

$$f(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Il grafico seguente mostra due distribuzioni normali con la stessa media ma con varianze diverse:



Variabile aleatoria normale standard

Diciamo che una variabile aleatoria X è una variabile aleatoria normale standard se X ha una distribuzione normale di media $\mu = 0$ e varianza $\sigma^2 = 1$:

$$X \sim \mathcal{N}(0,1)$$

La distribuzione di probabilità di X è allora:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2}$$

La funzione di densità di X si indica invece con la lettera Φ . Si ricorda che i valori della funzione di ripartizione delle normale standard vengono ottenuti utilizzando le apposite tabelle (o, nelle applicazioni pratiche, utilizzando opportuni software).

Proprietà delle variabili aleatorie normali

Data una generica variabile aleatoria $X \sim \mathcal{N}(\mu, \sigma^2)$, valgono sempre le proprietà seguenti:

1. $f_D(x)$ è simmetrica rispetto alla retta $x = \mu$.
2. La funzione di ripartizione in μ vale sempre $\frac{1}{2}$, ovvero: $F_D(\mu) = 0.5$.
3. La variabile aleatoria $Y = aX + b$, $a, b \in \mathbb{R}$ è una variabile aleatoria normale:

$$Y \sim \mathcal{N}(a \cdot \mu + b, a^2 \cdot \sigma^2)$$

Nel caso particolare in cui la variabile aleatoria sia una normale standard, cioè $Z \sim \mathcal{N}(0,1)$:

1. φ è una funzione pari.
2. $\Phi(0) = 0.5$
3. Per ogni valore reale di z , si ha $\Phi(-z) = 1 - \Phi(z)$.
4. Per ogni valore reale di z , si ha:

$$P(|Z| \leq k) = P(Z \leq k) - P(Z < -k) = \Phi(k) - \Phi(-k) = \Phi(k) - [1 - \Phi(k)] = 2\Phi(k) - 1$$
5. $P(|Z| \leq 3) \cong 0,99$. Possiamo quindi approssimare ad 1 la probabilità che Z assuma valori in modulo maggiori di 3, e proprio per questo motivo le tavole non riportano i valori della funzione di ripartizione corrispondenti a punti superiori a 3.

Standardizzazione

Data una generica variabile aleatoria $X \sim \mathcal{N}(\mu, \sigma^2)$, una particolare trasformazione affine, che assume un ruolo fondamentale, è quella nota come *standardizzazione*, ovvero:

$$\frac{X - \mu}{\sigma}$$

Si osserva facilmente che la variabile così ottenuta avrà ancora una distribuzione di probabilità dello stesso tipo della distribuzione di probabilità di X , ma con media nulla e varianza unitaria:

$$E\left(\frac{X - \mu}{\sigma}\right) = 0 \qquad \text{VAR}\left(\frac{X - \mu}{\sigma}\right) = 1$$

Nel caso particolare in cui X sia una variabile aleatoria gaussiana:

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

La variabile aleatoria:

$$\frac{X - \mu}{\sigma}$$

È una normale standard. In questo caso, la standardizzazione risulta particolarmente utile perché è possibile ricavare dalle tavole i valori della funzione di ripartizione di una normale standard, ma non di una generica gaussiana. Possiamo poi ottenere i valori della funzione di ripartizione di X a partire da quelli di Z nel modo seguente:

$$F_X(x) = P(X \leq x) = P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = P\left(Z \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

Quantili

In molti casi però non è richiesto di calcolare qual è la probabilità che una variabile aleatoria normale X assuma un valore non superiore ad una certa soglia, ma si richiede di individuare, data una certa probabilità p , qual è la soglia q_p tale che la probabilità che X sia non superiore a q_p risulti essere uguale a p . Il valore di q_p così individuato viene detto *quantile di ordine p di X* :

$$q_p = \text{"quantile di ordine } p\text{"} \quad \leftrightarrow \quad P(X \leq q_p) = p$$

L'operazione di individuazione del quantile può essere eseguita per variabili aleatorie di qualunque tipo, sia discrete che continue (nel caso discreto si hanno però alcuni problemi aggiuntivi, legati al fatto che la relativa funzione di distribuzione è continua a tratti). Noi però ci soffermiamo solo sul caso continuo.

Consideriamo in particolare la situazione in cui si voglia calcolare il quantile di ordine p di una variabile aleatoria normale X . Allora:

1. Per prima cosa, calcoliamo il quantile di ordine p della variabile aleatoria normale standard, Z , andando a cercare sulla tabella qual è il valore per il quale la funzione di ripartizione assume valore p . Indichiamo tale grandezza con z_p .
2. A questo punto, possiamo eseguire i seguenti calcoli:

$$p = F_X(q_p) = \Phi\left(\frac{q_p - \mu}{\sigma}\right) \quad \rightarrow \quad z_p = \frac{q_p - \mu}{\sigma} \quad \rightarrow \quad q_p = \sigma z_p + \mu$$

Proprietà: distribuzione della somma di variabili normali i.i.d.

A questo punto, possiamo anche riprendere un altro importante teorema del Calcolo delle Probabilità: date n variabili aleatorie indipendenti identicamente distribuite $X_1, X_2, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$, la loro somma è a sua volta una variabile aleatoria con distribuzione normale.

Inoltre, ricordando le semplici proprietà di media e varianza:

$$\left(\sum_{j=1}^n X_j\right) \sim \mathcal{N}(n\mu, n\sigma^2)$$

Se poi dividiamo la variabile aleatoria così ottenuta per n , naturalmente otterremo un'altra variabile aleatoria, che (come avremo modo di approfondire) è la media campionaria, e che non potrà che essere una normale:

$$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Consideriamo un esempio: sia μ l'esatto valore di misura di una certa grandezza. Se effettuiamo una certa misurazione X_1 della grandezza μ , utilizzando uno strumento con una certa precisione nota, che è definita come σ^{-1} , ci aspettiamo che il risultato non sia esattamente uguale a μ , ma si tratterà di una variabile aleatoria, che può essere vista come:

$$X_1 = \mu + \varepsilon$$

Dove:

$$\varepsilon \sim \mathcal{N}(0, \sigma^2)$$

Overo, si ha un errore casuale che ha una distribuzione normale di media nulla (siamo cioè nell'ipotesi di assenza di errori sistematici). Questo modello è tipico della situazione analizzata. Avremo inoltre:

$$X_1 \sim \mathcal{N}(\mu, \sigma^2)$$

Dove μ è un'incognita, mentre σ^2 è un parametro costruttivo dello strumento, e supponiamo che sia noto perché fornito dal produttore.

Il teorema centrale del limite

Enunciato del teorema

Sia X_1, X_2, \dots, X_n una sequenza di variabili aleatorie indipendenti identicamente distribuite con media μ e varianza $\sigma^2 > 0$. Allora:

$$\lim_{n \rightarrow +\infty} P\left(\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} < z\right) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(z)$$

Dove $\Phi(z)$ è la funzione di ripartizione della distribuzione normale standard:

$$\Phi(z) = \mathcal{N}(0,1)$$

Significato

In sostanza, stiamo affermando che, indipendentemente dalla distribuzione di probabilità delle variabili aleatorie X_1, X_2, \dots, X_n , purché esse siano indipendenti e identicamente distribuite, la loro somma ha una distribuzione che può essere approssimata come una variabile aleatoria normale di media μ e varianza $\frac{\sigma^2}{n}$:

$$\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim_{\text{approx}} \mathcal{N}(0,1) \qquad \bar{X} \sim_{\text{approx}} \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Funzione generatrice di momenti

Definizione della funzione generatrice dei momenti (f.g.m.)

Data una variabile aleatoria X , possiamo calcolare la media della variabile aleatoria e^{tX} :

$$E(e^{tX})$$

Se esiste un intorno del punto zero tale che per ogni t appartenente a tale intorno si abbia:

$$E(e^{tX}) < +\infty$$

Allora definiamo la funzione generatrice dei momenti di X come:

$$M_X(t) = E(e^{tX})$$

Quindi:

1. Se X è una variabile aleatoria discreta, avremo:

$$M_X(t) = \sum_x [f(x) \cdot e^{tx}]$$

2. Se X è una variabile aleatoria continua, avremo:

$$M_X(t) = \int_{-\infty}^{+\infty} f(x) \cdot e^{tx} dx$$

Calcolo della distribuzione a partire dalla funzione generatrice dei momenti

Il motivo fondamentale per il quale abbiamo introdotto la funzione generatrice dei momenti è che, calcolandone l'antitrasformata, si ottiene la funzione di densità della variabile di partenza. In altri termini, esiste una corrispondenza biunivoca tra la funzione di densità e la funzione generatrice dei momenti.

In particolare, per calcolare l'antitrasformata si utilizzano in realtà le opportune tabelle. In altri termini X e Y hanno la stessa funzione di densità se e solo se hanno la stessa funzione generatrice dei momenti.

$$X \sim F$$

$$Y \sim G$$

$$F = G \Leftrightarrow M_X(t) = M_Y(t)$$

Funzione generatrice dei momenti della somma di variabili aleatorie indipendenti

La funzione generatrice dei momenti della somma di variabili aleatorie indipendenti è la produttoria delle funzioni generatrici dei momenti delle singole variabili:

$$M_{\sum_{i=1}^n X_i}(t) = E(e^{t \cdot \sum_{i=1}^n X_i}) = E\left(\prod_{i=1}^n e^{tX_i}\right) = \prod_{i=1}^n E(e^{tX_i}) = \prod_{i=1}^n M_{X_i}(t)$$

Proprietà

La ragione per la quale la funzione generatrice dei momenti di una variabile aleatoria qualsiasi X viene indicata con questo nome è che la sua derivata k -esima rispetto alla variabile t assume in $t = 0$ il valore della media di X^k , ovvero del momento k -esimo di X :

$$\left. \frac{dM_X(t)}{dt} \right|_{t=0} = E[X^k]$$

Esempio di utilizzo

Si considerino n variabili aleatorie esponenziali i.i.d., ovvero:

$$X_1, X_2, \dots, X_n \text{ i.i.d. } \sim \mathcal{E}(\beta)$$

Supponiamo di voler determinare la legge di distribuzione della somma delle variabili appena descritte.

Uno dei modi possibili è quello di utilizzare la funzione generatrice di momenti della variabile aleatoria X .

Iniziamo allora calcolando la funzione generatrice dei momenti di una variabile aleatoria esponenziale:

$$X \sim \mathcal{E}(\beta)$$

Avremo:

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{+\infty} f(x) \cdot e^{tx} dx = \int_0^{+\infty} \frac{1}{\beta} e^{-\frac{x}{\beta}} \cdot e^{tx} dx = \frac{1}{\beta} \int_0^{+\infty} e^{-\left(\frac{1}{\beta}-t\right)x} dx = \\ &= \frac{1}{\beta} \left(\frac{1}{\beta}-t\right)^{-1} \int_0^{+\infty} \left(\frac{1}{\beta}-t\right) e^{-\left(\frac{1}{\beta}-t\right)x} dx = \frac{1}{\beta} \left(\frac{1}{\beta}-t\right)^{-1} = \frac{1}{1-\beta t} \end{aligned}$$

Tale calcolo vale però solo a patto che $\frac{1}{\beta} - t > 0$, ovvero:

$$t < \beta^{-1}$$

Siccome $\beta > 0$ per ipotesi, abbiamo certamente individuato un intorno dell'origine nel quale la funzione $M_X(t)$ appena calcolata non va all'infinito, perciò $M_X(t)$ è effettivamente la funzione generatrice dei momenti di X .

A questo punto, avremo:

$$M_{\sum_{i=0}^n X_j}(t) = \left(\frac{1}{1-\beta t}\right)^n \quad \text{se } t < \frac{1}{\beta}$$

Il vantaggio di eseguire tale calcolo è che, come si nota, sono stati sufficienti in realtà pochi calcoli per ottenere tale risultato e poi, utilizzando le tabelle, si ricava facilmente che la funzione di densità corrispondente (ovvero l'antitrasformata della funzione così calcolata) è la funzione di densità gamma.

$$\Gamma(n, \beta)$$

In particolare, si tratta di una distribuzione di Earlang, che è un caso particolare di distribuzione Γ e che viene utilizzata per modellare l'istante di arrivo dell' n -esimo guasto in un sistema.

Molto più complesso sarebbe stato calcolare la distribuzione di probabilità utilizzando l'integrale di convoluzione.

La distribuzione gamma

Funzione di densità

A questo punto, possiamo introdurre una distribuzione di probabilità continua che finora non abbiamo studiato, e che si rivelerà di fondamentale importanza all'interno del nostro corso: la funzione gamma. Se una variabile aleatoria X ha distribuzione di probabilità gamma con i parametri α e β , scriviamo:

$$X \sim \Gamma(\alpha, \beta)$$

La funzione di densità di X sarà allora:

$$f_X(x, \alpha, \beta) = \frac{x^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha} e^{-\frac{x}{\beta}} \cdot \mathbb{I}_{(0;+\infty)}(x), \quad \alpha, \beta \in \mathbb{R}$$

Dove:

- α è il *parametro di regolarità*;
- β prende il nome di *parametro di scala*;
- $\Gamma(\alpha)$ è la costante che serve per far in modo che si abbia:

$$\int_0^{+\infty} f_X(x, \alpha, \beta) dx = 1$$

Tale condizione è ovviamente indispensabile affinché $f_X(x, \alpha, \beta)$ sia effettivamente una densità di probabilità. Si può notare che tale integrale non dipende in realtà da β , perciò possiamo arbitrariamente fissare β , ad esempio, ad 1, e definire $\Gamma(\alpha)$ come:

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} \cdot e^{-x} dx$$

In particolare, valgono le seguenti relazioni, che risultano utili per calcolare il valore di $\Gamma(\alpha)$ nei casi pratici di interesse:

$$\begin{cases} \Gamma(1) = 1 \\ \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} \\ \Gamma(\alpha + 1) = \alpha \cdot \Gamma(\alpha), \quad \alpha > 0 \end{cases}$$

Da queste regole si ricava in maniera molto semplice ed intuitiva che, se n è un numero intero non negativo, allora abbiamo:

$$\Gamma(n + 1) = n!$$

Mentre, per definizione:

$$\Gamma(0) = 1$$

Funzione generatrice dei momenti

La funzione generatrice dei momenti di una variabile aleatoria gamma è data dall'espressione:

$$M_{\Gamma(\alpha, \beta)} = \left(\frac{1}{1 - \beta t}\right)^\alpha \quad \text{se } t < \frac{1}{\beta}$$

Media e varianza

1. Per calcolare la media, possiamo sfruttare la proprietà della funzione generatrice dei momenti per la quale $M'_X(0)$ è la media di X :

$$E[X] = \left. \frac{dM_X(t)}{dt} \right|_{t=0} = \left[\alpha \left(\frac{1}{1 - \beta t}\right)^{\alpha-1} \frac{\beta}{(1 - \beta t)^2} \right]_{t=0} = \left[\alpha \beta \left(\frac{1}{1 - \beta t}\right)^{\alpha+1} \right]_{t=0} = \alpha \beta$$

2. Per calcolare la varianza, calcoliamo dapprima il momento secondo:

$$E[X^2] = \left. \frac{d^2 M_X(t)}{dt^2} \right|_{t=0} = \alpha \beta \left[(\alpha + 1) \left(\frac{1}{1 - \beta t}\right)^\alpha \frac{\beta}{(1 - \beta t)^2} \right]_{t=0} = \alpha \beta^2 (\alpha + 1) \left[\left(\frac{1}{1 - \beta t}\right)^{\alpha+2} \right]_{t=0} = \alpha \beta^2 (\alpha + 1)$$

Quindi:

$$\text{VAR}[X] = E[X^2] - E^2[X] = \alpha^2 \beta^2 + \alpha \beta^2 - \alpha^2 \beta^2 = \alpha \beta^2$$

Proprietà n. 1: prodotto tra una variabile gamma ed una costante

Si noti che, se consideriamo la variabile aleatoria gamma: $X \sim \Gamma(\alpha, \beta)$

Allora, la variabile aleatoria: $Y = c \cdot X, \quad c \in \mathbb{R}$

Avrà distribuzione gamma con gli stessi parametri di X , a meno della costante moltiplicativa c che compare nel secondo dei due parametri:

$$Y \sim \Gamma(\alpha, c \cdot \beta)$$

Infatti, avremo:

$$M_Y(t) = M_{cX}(t) = E[e^{tcX}] = E[e^{(tc)X}] = \left(\frac{1}{1 - \beta ct}\right)^\alpha \quad \text{se } ct < \frac{1}{\beta}, \text{ ovvero } t < \frac{1}{c\beta}$$

E quella appena ottenuta non è altro che la funzione generatrice di una distribuzione

$$\Gamma(\alpha, c\beta)$$

Questa è la ragione per la quale tale parametro è noto come *parametro di scala*.

Proprietà n. 2: somma di variabili gamma con lo stesso parametro di scala

Siano date due variabili aleatorie X e Y , indipendenti tra loro ed entrambe con distribuzione gamma, aventi lo stesso parametro di scala:

$$X \sim \Gamma(\alpha_1, \beta) \qquad Y \sim \Gamma(\alpha_2, \beta)$$

Allora, se si considera la variabile aleatoria ottenuta come somma tra le due:

$$W = X + Y$$

Tale variabile aleatoria è una variabile gamma con parametro di regolarità $\alpha_1 + \alpha_2$ e parametro di scala β :

$$W \sim \Gamma(\alpha_1 + \alpha_2, \beta)$$

Possiamo facilmente verificarlo considerando la funzione generatrice dei momenti:

$$M_W(t) = E[e^{t(X+Y)}] = E[e^{tX} e^{tY}]$$

Sfruttando l'indipendenza tra le variabili aleatorie date, otteniamo:

$$M_W(t) = E[e^{tX}] \cdot E[e^{tY}] = \left(\frac{1}{1 - \beta t}\right)^{\alpha_1} \left(\frac{1}{1 - \beta t}\right)^{\alpha_2} = \left(\frac{1}{1 - \beta t}\right)^{\alpha_1 + \alpha_2} = \Gamma(\alpha_1 + \alpha_2, \beta)$$

Possiamo inoltre affermare che, se sono date due variabili aleatorie X e Y indipendenti, tali che:

$$X \sim \Gamma(\alpha_1, \beta) \qquad (X + Y) \sim \Gamma(\delta, \beta) \qquad \delta > \alpha_1$$

Possiamo dire con certezza che:

$$Y \sim \Gamma(\delta - \alpha_1, \beta)$$

Infatti:

$$M_{X+Y}(t) = \left(\frac{1}{1 - \beta t}\right)^\delta = M_X(t)M_Y(t) = \left(\frac{1}{1 - \beta t}\right)^{\alpha_1} M_Y(t) \rightarrow M_Y(t) = \left(\frac{1}{1 - \beta t}\right)^{\delta - \alpha_1} = \Gamma(\delta - \alpha_1, \beta)$$

Caso particolare: l'esponenziale

Se $\alpha = 1$, allora la gamma coincide con l'esponenziale: $\Gamma(1, \beta) = \mathcal{E}(\beta)$

La distribuzione chi-quadro

Chi-quadro ad un grado di libertà

Un altro caso particolare della distribuzione Γ è quello che si ha quando i parametri sono 0.5 e 2. La variabile ottenuta è detta **chi-quadro con 1 grado di libertà** χ_1^2 :

$$\Gamma\left(\frac{1}{2}, 2\right) = \chi_1^2$$

Questa distribuzione di probabilità è la stessa che contraddistingue una variabile aleatoria X ottenuta come quadrato di una variabile aleatoria Z normale standard:

$$X = Z^2, \quad Z \sim \mathcal{N}(0,1) \quad \rightarrow \quad X \sim \chi_1^2$$

ed è proprio per questa ragione che viene detta *chi-quadro*.

Chi-quadro a n gradi di libertà

La distribuzione Γ con parametri $0.5 \cdot n$ e 2 è detta **chi-quadro con n gradi di libertà** χ_n^2 :

$$\Gamma\left(\frac{n}{2}, 2\right) = \chi_n^2$$

Se consideriamo le variabili aleatorie normali:

$$X_1, X_2, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2) \quad i. i. d.$$

Possiamo eseguire la standardizzazione di ciascuna di tali variabili, ottenendo:

$$\frac{X_1 - \mu}{\sigma}, \frac{X_2 - \mu}{\sigma}, \dots, \frac{X_n - \mu}{\sigma} \sim \mathcal{N}(0,1) \quad i. i. d.$$

Di conseguenza, in base a quanto visto al punto precedente, avremo:

$$\left(\frac{X_1 - \mu}{\sigma}\right)^2, \left(\frac{X_2 - \mu}{\sigma}\right)^2, \dots, \left(\frac{X_n - \mu}{\sigma}\right)^2 \sim \chi_1^2 \quad i. i. d.$$

Se a questo punto consideriamo la variabile aleatoria somma di tutte quelle così ottenute:

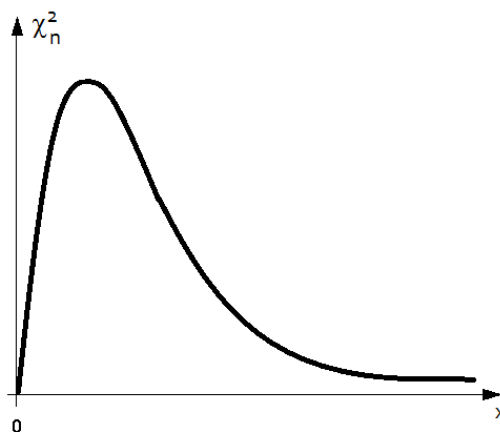
$$W = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2$$

Quella che otteniamo è una variabile aleatoria con funzione generatrice dei momenti:

$$M_W(t) = \left[\left(\frac{1}{1-2t}\right)^{\frac{1}{2}} \right]^n = \left(\frac{1}{1-2t}\right)^{\frac{n}{2}}$$

Che corrisponde proprio alla funzione generatrice dei momenti di una variabile chi-quadro a n gradi di libertà. Il numero di gradi di libertà corrisponde allora al numero di variabili aleatorie normali standard che è necessario elevare al quadrato e sommare per ottenere la distribuzione chi-quadro corrispondente.

Grafico della funzione di densità



Proprietà

Come conseguenza della proprietà n. 2 relativa alle variabili gamma, possiamo affermare che, se è data una variabile aleatoria:

$$W = Y_1 + Y_2$$

Dove Y_1 e Y_2 sono indipendenti e dove si ha:

$$Y_1 \sim \chi_1^2$$

$$W \sim \chi_n^2$$

$$n \geq 2$$

Allora, abbiamo:

$$Y_2 \sim \Gamma\left(\frac{n}{2} - \frac{1}{2}, 2\right) = \chi_{n-1}^2$$

Media e varianza

Ricordando che stiamo semplicemente analizzando un caso particolare di distribuzione gamma, abbiamo:

- Media: $E[\chi_n^2] = \frac{n}{2} \cdot 2 = n$
- Varianza: $VAR[\chi_n^2] = \frac{n}{2} \cdot 2^2 = 2n$

Approssimazione

Naturalmente, per una variabile aleatoria X con distribuzione chi-quadro, per valori grandi di n , tendenti cioè ad infinito, possiamo utilizzare il teorema centrale del limite, ed approssimare una variabile aleatoria con distribuzione chi-quadro utilizzando una variabile aleatoria normale:

$$\mathcal{N}(n, 2n)$$

Esistono in realtà delle approssimazioni migliori per la distribuzione di chi-quadro; tuttavia, per i nostri scopi questa approssimazione sarà sufficiente.

La densità T-student

Definizione

Una variabile aleatoria con densità T-student è una variabile aleatoria che può sempre essere pensata come ottenuta a partire dalle due variabili aleatorie Z e W così definite:

$$Z \sim \mathcal{N}(0,1)$$

$$W \sim \chi_a^2$$

Calcolando:

$$T = \frac{Z}{\sqrt{W}} \cdot \sqrt{a}$$

Allora, la funzione di densità che si ottiene è:

$$f_T(t) = \frac{\Gamma\left(\frac{a+1}{2}\right)}{\Gamma\left(\frac{a}{2}\right)\sqrt{\pi a}} \left(1 + \frac{t^2}{a}\right)^{-\frac{a+1}{2}}, \quad t \in \mathbb{R}, a = 1, 2, \dots$$

Grafico della funzione di densità

Il grafico della funzione di densità è molto simile a quello della normale standard, con l'unica differenza che le code che si ottengono sono più grosse rispetto a quelle che si hanno con una distribuzione gaussiana.

Stima di media e varianza di distribuzioni di probabilità

Supponiamo ora di non conoscere con esattezza un certo parametro di una data distribuzione di probabilità, e di volerne ottenere una *statistica*. Come possiamo procedere?

Consideriamo come situazione iniziale il caso in cui si voglia stimare la media di una variabile casuale. Ad esempio, nel caso della distribuzione esponenziale ciò equivale a stimare il parametro β .

Terminologia

Dato

Le stime verranno eseguite sempre sulla base di un insieme di dati sperimentali, rilevati cioè dalla pratica, mediante delle misurazioni. I dati sono delle variabili aleatorie indipendenti e identicamente distribuite:

$$X_1, X_2, \dots, X_n \text{ i. i. d.}$$

Statistica

Chiamiamo statistica una qualsiasi funzione dei dati, ovvero un qualsiasi valore che viene calcolato sulla base dei dati stessi.

Stimatore

Uno stimatore è una particolare statistica che viene utilizzata per *campionare* (o *stimare*) un certo parametro o una certa caratteristica. Lo stimatore sarà a sua volta una variabile aleatoria.

Stima

La stima è il valore osservato dello stimatore.

Mean Square Error (MSE)

Chiamiamo *Mean Square Error* (MSE) la media dei quadrati degli scarti tra lo stimatore T_n e il parametro stimato ϑ :

$$MSE(T_n) = E[(T_n - \vartheta)^2]$$

Il MSE è perciò definibile anche come il momento secondo della variabile aleatoria $T_n - \vartheta$.

Valutazione di uno stimatore

Naturalmente, il nostro obiettivo è quello di introdurre degli stimatori che siano "di buona qualità" ovvero, intuitivamente, che approssimino bene il parametro incognito.

Per valutare la qualità dello stimatore, siccome quest'ultimo è a sua volta una variabile aleatoria, possiamo pensare di prendere in analisi la sua media e la sua varianza. Le osservazioni seguenti valgono sia per il caso continuo che per quello discreto.

Stimatore non distorto

Detto ϑ il parametro che stiamo stimando e detta $T = g(X_1, X_2, \dots, X_n)$ la statistica utilizzata come stimatore, diciamo che lo stimatore T non è distorto se la sua media è uguale a ϑ :

$$E(T) = \vartheta$$

Stimatore consistente

Detto ϑ il parametro che stiamo stimando e detta $T = g(X_1, X_2, \dots, X_n)$ la statistica utilizzata come stimatore, diciamo che T è uno stimatore consistente di ϑ se:

$$\lim_{n \rightarrow +\infty} E[(T_n - \vartheta)^2] = 0$$

Stimatore asintoticamente non distorto

Detto ϑ il parametro che stiamo stimando e detta $T = g(X_1, X_2, \dots, X_n)$ la statistica utilizzata come stimatore, diciamo che T è uno stimatore asintoticamente non distorto di ϑ se:

$$\lim_{n \rightarrow +\infty} E(T_n) = \vartheta$$

Stimatore consistente in media quadratica

Detto ϑ il parametro che stiamo stimando e detta $T = g(X_1, X_2, \dots, X_n)$ la statistica utilizzata come stimatore, diciamo che T è uno stimatore consistente in media quadratica di ϑ se:

$$\lim_{n \rightarrow +\infty} E(T_n) = \vartheta \quad e \quad \lim_{n \rightarrow +\infty} VAR(T_n) = 0$$

Osservazioni

1. Naturalmente, la condizione migliore di uno stimatore è quella in cui esso risulti essere non distorto e consistente. La condizione di consistenza in media quadratica è invece un po' meno stringente.
2. Nel caso in cui lo stimatore non sia distorto, il *Mean Square Error* coincide con la varianza dello stimatore (ciò lo si ricava facilmente, perché in tal caso ϑ è la media di T_n):

$$MSE(T_n) = VAR(T_n)$$

Di conseguenza, la condizione di consistenza è in tal caso equivalente a:

$$\lim_{n \rightarrow +\infty} VAR(T_n) = 0$$

3. Se invece lo stimatore è distorto, allora ϑ non coinciderà con la media di T_n . Possiamo comunque osservare che, come già affermato, $MSE(T_n)$ è il momento secondo di $T_n - \vartheta$, perciò, dalla formula pratica della varianza:

$$VAR(X) = E(X^2) - E^2(X) \quad \rightarrow \quad E(X^2) = VAR(X) + E^2(X)$$

Possiamo ricavare:

$$MSE(T_n) = E[(T_n - \vartheta)^2] = VAR(T_n - \vartheta) + E^2(T_n - \vartheta)$$

Ricordando poi le proprietà di media e varianza:

$$MSE(T_n) = VAR(T_n) + [E(T_n) - \vartheta]^2$$

Se lo stimatore è consistente in media quadratica, il limite del MSE tende a zero per $n \rightarrow +\infty$.

La media campionaria

Media campionaria (o empirica)

La media campionaria, talvolta detta anche media empirica, è la media dei dati. Si tratta perciò di una particolare statistica, che viene calcolata come:

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{1}{n} \cdot \sum_{i=1}^n X_i$$

Non distorsione della media campionaria

Possiamo facilmente verificare che \bar{X} è uno stimatore non distorto. Infatti:

$$E(\bar{X}) = E\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \frac{1}{n} [E(X_1) + E(X_2) + \dots + E(X_n)] = \frac{n \cdot \mu}{n} = \mu$$

Dove μ è la media stimata.

Consistenza della media campionaria

Analogamente, possiamo verificare che \bar{X} è uno stimatore consistente. Essendo non distorto, possiamo verificarlo semplicemente per mezzo del calcolo della varianza:

$$VAR(\bar{X}) = VAR\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \frac{1}{n^2} VAR(X_1 + X_2 + \dots + X_n)$$

Siccome per ipotesi le variabili che rappresentano i dati sono i.i.d., possiamo sfruttare la loro indipendenza ed ottenere:

$$VAR(\bar{X}) = \frac{1}{n^2} [VAR(X_1) + VAR(X_2) + \dots + VAR(X_n)] = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Avremo perciò:

$$\lim_{n \rightarrow +\infty} MSE(T_n) = \lim_{n \rightarrow +\infty} VAR(T_n) = \lim_{n \rightarrow +\infty} \frac{\sigma^2}{n} = 0$$

La varianza campionaria

Varianza campionaria

La varianza campionaria S^2 è uno stimatore che viene utilizzato per stimare la varianza incognita di una certa variabile aleatoria. A tale scopo, si considera un campione di n dati:

$$X_1, X_2, \dots, X_n \text{ i. i. d., } n \geq 2$$

Definiamo allora:

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

Si noti che in alcuni testi la varianza campionaria viene definita indicando come denominatore il valore n ; la definizione che abbiamo dato noi però impedisce di calcolare un indice di dispersione nel caso in cui si abbia un solo campione, scelta che risulterebbe evidentemente del tutto irragionevole. Esiste inoltre una ragione più profonda per la quale la definizione adottata è quella appena riportata, ma analizzeremo tale ragione solo in seguito.

Proprietà della varianza campionaria: non distorsione

La varianza campionaria è uno stimatore non distorto di σ^2 . Possiamo infatti verificarlo. Innanzitutto, ricordiamo che questo significa che:

$$E[S^2] = \sigma^2$$

Abbiamo:

$$\begin{aligned} \sum_{i=1}^n (X_i - \bar{X})^2 &= \sum_{i=1}^n (X_i + \mu - \mu - \bar{X})^2 = \sum_{i=1}^n [(X_i - \mu) - (\bar{X} - \mu)]^2 = \\ &= \sum_{i=1}^n [(X_i - \mu)^2 + (\bar{X} - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu)] = \sum_{i=1}^n (X_i - \mu)^2 + n(\bar{X} - \mu)^2 - 2(\bar{X} - \mu) \sum_{i=1}^n (X_i - \mu) \end{aligned}$$

È facile osservare che abbiamo:

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

Perciò, sostituendo, otteniamo:

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

Avremo dunque:

$$E[S^2] = \frac{1}{n-1} E \left[\sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2 \right] = \frac{1}{n-1} \left\{ E \left[\sum_{i=1}^n (X_i - \mu)^2 \right] - nE[(\bar{X} - \mu)^2] \right\}$$

A questo punto, è si può notare che, per definizione di varianza:

$$E[(\bar{X} - \mu)^2] = \text{VAR}(\bar{X})$$

Da cui:

$$\begin{aligned} E[S^2] &= \frac{1}{n-1} E \left[\sum_{i=1}^n (X_i - \mu)^2 \right] - \frac{n}{n-1} \text{VAR}(\bar{X}) = \frac{1}{n-1} \sum_{i=1}^n E[(X_i - \mu)^2] - \frac{n}{n-1} \text{VAR}(\bar{X}) = \\ &= \frac{1}{n-1} \left[\sum_{i=1}^n \text{VAR}(X_i) - n\text{VAR}(\bar{X}) \right] = \frac{1}{n-1} \left[\sum_{i=1}^n \sigma^2 - n \frac{\sigma^2}{n} \right] = \frac{1}{n-1} [n\sigma^2 - \sigma^2] = \sigma^2 \end{aligned}$$

Proprietà della varianza campionaria: consistenza in media quadratica

Analogamente, ma con un lungo procedimento algebrico (che omettiamo), si ricava l'espressione della varianza di S^2 . Da tale espressione si nota che $\text{VAR}(S^2)$ tende a 0 per $n \rightarrow +\infty$, con una velocità circa uguale a quella di $\frac{1}{n}$. Possiamo così concludere che S^2 è uno stimatore consistente in media quadratica.

Media e varianza campionaria nel caso gaussiano

Proprietà dei campioni normali: indipendenza tra media campionaria e varianza campionaria

Se è dato un campione normale, ovvero:

$$X_1, X_2, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2) \text{ i.i.d.}, n \geq 2$$

Allora si può dimostrare che la media campionaria \bar{X} e la varianza campionaria S^2 sono indipendenti.

Distribuzione di probabilità della media campionaria nel caso di campioni normali

Nel caso in analisi, sulla base di tutte le osservazioni finora svolte:

$$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Distribuzione di probabilità di S^2 nel caso di campioni normali

Supponiamo ora di avere un campione di dati:

$$X_1, X_2, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2) \text{ i.i.d.}, n \geq 2$$

Con μ, σ^2 incognite. Consideriamo inoltre la variabile aleatoria:

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

Siccome in precedenza abbiamo già dimostrato che vale l'uguaglianza:

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

Otterremo:

$$X = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2 - n\left(\frac{\bar{X} - \mu}{\sigma}\right)^2$$

Chiamiamo per praticità:

$$Y_1 = n\left(\frac{\bar{X} - \mu}{\sigma}\right)^2 = \left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}\right)^2$$

Naturalmente, $\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ è la standardizzata di \bar{X} , e siccome \bar{X} è una variabile normale, quella così ottenuta è una normale standard. Di conseguenza, Y_1 è il quadrato di una normale standard, perciò:

$$Y_1 \sim \chi_1^2$$

Inoltre, possiamo indicare:

$$W_1 = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2$$

E, siccome W_1 è la somma di n quadrati di variabili che sono le standardizzazioni di altrettante variabili aleatorie normali, è chiaro in base alle proprietà precedentemente enunciate che si avrà:

$$W_1 \sim \chi_n^2$$

Perciò possiamo concludere che, avendo:

$$X = W_1 - Y_1$$

Avremo necessariamente:

$$X \sim \chi_{n-1}^2$$

Ora, siccome vale la relazione:

$$S^2 = \frac{\sigma^2}{n-1} X$$

Avremo:

$$S^2 \sim \Gamma\left(\frac{n}{2}, 2\frac{\sigma^2}{n-1}\right)$$

Intervalli di confidenza

Precisione e accuratezza

A questo punto, è interessante valutare qual è la probabilità che lo stimatore si discosti dal valore vero del parametro di un certo valore massimo. Vogliamo in altri termini che si abbia:

$$P(|\hat{\vartheta} - \vartheta| < \delta) = \gamma$$

Con $\delta \rightarrow 0$ e $\gamma \rightarrow 1$, dove δ è la **precisione della stima**, mentre γ viene chiamato **accuratezza della stima**. Possiamo notare che tali grandezze sono strettamente correlate: se si fissa l'accuratezza γ , a patto di non intervenire su altre grandezze, si determina con certezza il valore della precisione δ , e viceversa.

Il concetto di intervallo di confidenza

In alcuni casi, anziché fornire una stima puntuale di un certo parametro, può essere opportuno fornire un certo intervallo, detto intervallo di confidenza, entro il quale il valore vero del parametro cadrà con una certa probabilità nota. Tale probabilità rappresenta il concetto di accuratezza della stima, mentre l'ampiezza dell'intervallo è strettamente legata alla precisione della stima.

Consideriamo ad esempio la media campionaria: come è ormai noto, la media di una variabile aleatoria può essere stimata con la media campionaria; tuttavia, in una variabile aleatoria continua X , la probabilità che X assuma esattamente il valore uguale alla media campionaria \bar{X} è nulla (perché si tratta di un singolo valore reale costante), perciò possiamo pensare di fornire una misura probabilistica dell'errore che si commette sostituendo al valore del parametro μ il valore assunto dallo stimatore \bar{X} .

Intervallo di confidenza per la media nel caso di campione gaussiano

- Caso 1: varianza nota

Sia x_1, \dots, x_n una realizzazione del campione casuale X_1, \dots, X_n della popolazione $\mathcal{N}(\mu, \sigma^2)$. Fissato $\gamma \in (0,1)$, se la varianza σ^2 è nota, cercare di definire un intervallo di confidenza per la media si traduce nel definire l'accuratezza e la precisione della stima:

$$\gamma = P\left(|\bar{X} - \mu| < z_{\frac{\gamma+1}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right) = P\left(\bar{X} - z_{\frac{\gamma+1}{2}} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{\gamma+1}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

Abbiamo in tal modo ottenuto un certo intervallo di valori, detto *intervallo di confidenza* (IC) per la media μ di *livello di confidenza* γ :

$$\left(\bar{X} - z_{\frac{\gamma+1}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\frac{\gamma+1}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

Proviamo adesso a vedere come determinare la precisione δ a partire dall'accuratezza γ :

$$\gamma = P(|\bar{X} - \mu| < \delta) = P\left(\frac{|\bar{X} - \mu|}{\frac{\sigma}{\sqrt{n}}} < \frac{\delta}{\frac{\sigma}{\sqrt{n}}}\right) = P\left(\left|\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}\right| < \frac{\delta}{\frac{\sigma}{\sqrt{n}}}\right) = P\left(|Z| < \frac{\delta}{\frac{\sigma}{\sqrt{n}}}\right) = 2\Phi\left(\frac{\delta}{\frac{\sigma}{\sqrt{n}}}\right) - 1$$

Da questo si ricava:

$$\Phi\left(\frac{\delta}{\frac{\sigma}{\sqrt{n}}}\right) = \frac{\gamma + 1}{2} \quad \rightarrow \quad \frac{\delta}{\frac{\sigma}{\sqrt{n}}} = z_{\frac{\gamma+1}{2}} \quad \rightarrow \quad \delta = z_{\frac{\gamma+1}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

Possiamo allora scrivere:

$$\gamma = P\left(|\bar{X} - \mu| < z_{\frac{\gamma+1}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

Questa relazione mette in mostra che, più si vuole una stima accurata, meno la stima sarà precisa, e viceversa, a patto che si considerino fissi i valori di σ e di n . Tra questi parametri è chiaro che l'unico sul quale si può intervenire è n : per rendere sia più precisa che più accurata una stima, l'unica cosa che possiamo fare è aumentare il numero di osservazioni sulla base delle quali la stima stessa si basa.

• Caso 2: varianza incognita

Supponiamo ora che anche la varianza σ^2 sia incognita: in questo caso, è necessario stimare anche la varianza stessa, utilizzando la *varianza campionaria* S^2 .

Se il valore di n è sufficientemente grande, allora possiamo assumere che, approssimativamente, la variabile aleatoria così definita:

$$\frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}}$$

Abbia una distribuzione normale standard (cioè anche se si utilizza la varianza stimata s^2 anziché quella reale σ^2 continua a valere il concetto espresso dal teorema centrale del limite). Si noti che con s è stata indicata una realizzazione di S .

Allora l'intervallo di confidenza che si ottiene, e che viene detto *asintotico* di livello approssimato γ , è:

$$\left(\bar{x} - z_{\frac{\gamma+1}{2}} \cdot \frac{s}{\sqrt{n}}, \bar{x} + z_{\frac{\gamma+1}{2}} \cdot \frac{s}{\sqrt{n}} \right)$$

Se però il valore di n è piccolo, la variabile aleatoria precedentemente introdotta non ha distribuzione normale standard, e la variabilità aumenta di molto. La distribuzione di probabilità di tale variabile è infatti una *T di student* con $n - 1$ gradi di libertà:

$$\frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}} \sim t_{n-1}$$

Infatti $Z = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim \mathcal{N}(0,1)$, mentre $W = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$ e, siccome la variabile sopra riportata è:

$$\frac{Z}{\sqrt{W}} \cdot \sqrt{n-1}$$

Possiamo semplicemente applicare la definizione di *T di student*. Si noti che, nel caso in cui si abbia $n \geq 60$, la densità *T di student* converge alla densità normale standard, perciò l'intervallo di confidenza assume la stessa forma rispetto a quello che si ha nel caso in cui la varianza sia nota.

Nel caso generale, occorre utilizzare quantili della *T di student*. Abbiamo infatti:

$$P\left(-q < \frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}} < q\right) = \gamma$$

Con:

$$P\left(\frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}} < q\right) = \gamma + \frac{1 - \gamma}{2} = \frac{1 + \gamma}{2} \rightarrow q = t_{n-1}\left(\frac{1 + \gamma}{2}\right)$$

Da cui:

$$P\left(-t_{n-1}\left(\frac{1 + \gamma}{2}\right) < \frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}} < t_{n-1}\left(\frac{1 + \gamma}{2}\right)\right) = \gamma$$

E quindi:

$$P\left(\bar{X} - \sqrt{\frac{S^2}{n}} \cdot t_{n-1}\left(\frac{1 + \gamma}{2}\right) < \mu < \bar{X} + \sqrt{\frac{S^2}{n}} \cdot t_{n-1}\left(\frac{1 + \gamma}{2}\right)\right) = \gamma$$

Intervallo di confidenza per la varianza del caso di campione gaussiano

Dato un certo campione gaussiano:

$$X_1, X_2, \dots, X_n \text{ i. i. d. } \sim \mathcal{N}(\mu, \sigma^2)$$

Vogliamo individuare una "forbice" di valori tale che si abbia un certo livello di sicurezza che il valore vero di σ^2 appartenga a tale intervallo bilatero. In altri termini, l'obiettivo è quello di individuare due valori:

$$T_1(x_1, x_2, \dots, x_n)$$

$$T_2(x_1, x_2, \dots, x_n)$$

Tali che:

$$P(T_1 < \sigma^2 < T_2) = \gamma, \quad \gamma \in (0,1)$$

L'intervallo che otterremo in questo modo sarà detto *intervallo di confidenza con livello di confidenza γ* , il più possibile vicino ad 1.

• Caso 1: intervallo bilatero nel caso di media incognita

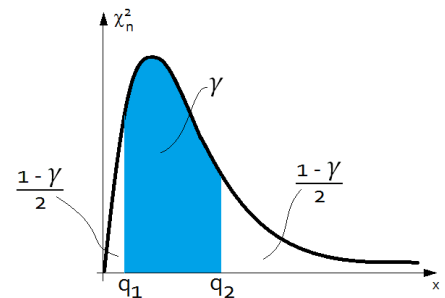
Ipotizziamo che anche la media μ sia incognita. In questo caso, possiamo partire ricordando che:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$

La situazione sarà allora quella rappresentata nella figura a lato.

Vogliamo dunque trovare i valori q_1 e q_2 tali che:

$$P\left(q_1 < \frac{(n-1)S^2}{\sigma^2} < q_2\right) = \gamma$$



E che:

$$P\left(\frac{(n-1)S^2}{\sigma^2} \leq q_1\right) = P\left(\frac{(n-1)S^2}{\sigma^2} \geq q_2\right) = \frac{(1-\gamma)}{2}$$

Di conseguenza, q_1 è il quantile della distribuzione chi-quadro a $n - 1$ gradi di libertà nel valore $\frac{1-\gamma}{2}$, mentre q_2 è il quantile valutato in $\frac{1-\gamma}{2} + \gamma = \frac{1+\gamma}{2}$:

$$q_1 = \chi_{n-1}^2\left(\frac{1-\gamma}{2}\right)$$

$$q_2 = \chi_{n-1}^2\left(\frac{1+\gamma}{2}\right)$$

L'intervallo di confidenza che otteniamo è perciò:

$$P\left(\chi_{n-1}^2\left(\frac{1-\gamma}{2}\right) < \frac{(n-1)S^2}{\sigma^2} < \chi_{n-1}^2\left(\frac{1+\gamma}{2}\right)\right) = \gamma$$

Da cui:

$$P\left(\frac{S^2(n-1)}{\chi_{n-1}^2\left(\frac{1+\gamma}{2}\right)} < \sigma^2 < \frac{S^2(n-1)}{\chi_{n-1}^2\left(\frac{1-\gamma}{2}\right)}\right) = \gamma$$

Indicata con s^2 la realizzazione di S^2 , l'intervallo di confidenza con livello di confidenza γ sarà allora:

$$\left(\frac{s^2(n-1)}{\chi_{n-1}^2\left(\frac{1+\gamma}{2}\right)}; \frac{s^2(n-1)}{\chi_{n-1}^2\left(\frac{1-\gamma}{2}\right)}\right)$$

• Caso 2: intervallo unilatero nel caso di media incognita

Talvolta si è interessati a calcolare un intervallo unilatero anziché bilatero, ovvero si vuole individuare una statistica del tipo:

a) Nel caso in cui si cerchi un intervallo del tipo *lower bound*:

$$P(T_L < \sigma^2) = \gamma$$

E si avrà così:

$$P\left(\frac{(n-1)S^2}{\sigma^2} < q\right) = \gamma$$

Perciò:

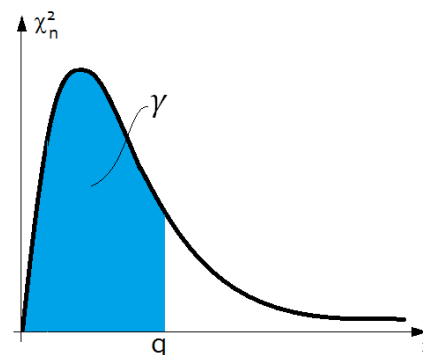
$$q = \chi_{n-1}^2(\gamma)$$

E quindi:

$$P\left(\frac{(n-1)S^2}{\sigma^2} < \chi_{n-1}^2(\gamma)\right) = \gamma \rightarrow P\left(\sigma^2 > \frac{(n-1)S^2}{\chi_{n-1}^2(\gamma)}\right)$$

E concludiamo così:

$$T_L = \frac{(n-1)S^2}{\chi_{n-1}^2(\gamma)}$$



b) Nel caso in cui si cerchi un intervallo del tipo *upper bound*:

$$P(T_U > \sigma^2) = \gamma$$

Avremo così:

$$P\left(\frac{(n-1)S^2}{\sigma^2} > q\right) = \gamma$$

E perciò:

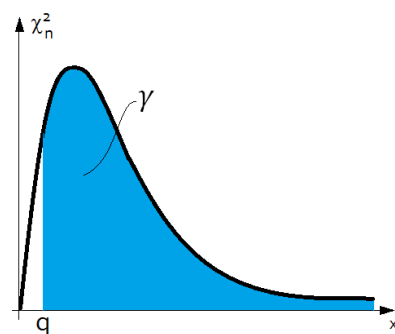
$$q = \chi_{n-1}^2(1 - \gamma)$$

E quindi:

$$P\left(\frac{(n-1)S^2}{\sigma^2} > \chi_{n-1}^2(1 - \gamma)\right) = \gamma \rightarrow P\left(\sigma^2 < \frac{(n-1)S^2}{\chi_{n-1}^2(1 - \gamma)}\right)$$

E concludiamo così:

$$T_U = \frac{(n-1)S^2}{\chi_{n-1}^2(1 - \gamma)}$$



• Caso 3: intervallo bilatero nel caso di media nota

Consideriamo ora il caso in cui la media è nota e si ha $\mu = \mu_0$. In questo caso, come risulta ovvio, conviene utilizzare una diversa statistica, che è così definita:

$$S_0^2 = \frac{\sum_{i=1}^n (X_j - \mu_0)^2}{n}$$

Si avrà così:

$$\frac{nS_0^2}{\sigma^2} \sim \chi_n^2$$

Allora dobbiamo avere:

$$P\left(q_1 < \frac{nS_0^2}{\sigma^2} < q_2\right) = \gamma$$

Imponendo inoltre:

$$P\left(\frac{nS_0^2}{\sigma^2} \leq q_1\right) = P\left(\frac{nS_0^2}{\sigma^2} \geq q_2\right) = \frac{(1-\gamma)}{2}$$

Di conseguenza:

$$q_1 = \chi_n^2 \left(\frac{1-\gamma}{2}\right) \qquad q_2 = \chi_n^2 \left(\frac{1+\gamma}{2}\right)$$

L'intervallo di confidenza che otteniamo è perciò:

$$P\left(\chi_n^2 \left(\frac{1-\gamma}{2}\right) < \frac{nS_0^2}{\sigma^2} < \chi_n^2 \left(\frac{1+\gamma}{2}\right)\right) = \gamma$$

Da cui:

$$P\left(\frac{S_0^2 n}{\chi_n^2 \left(\frac{1+\gamma}{2}\right)} < \sigma^2 < \frac{S_0^2 n}{\chi_n^2 \left(\frac{1-\gamma}{2}\right)}\right) = \gamma$$

Indicata con s_0^2 la realizzazione di S_0^2 , l'intervallo di confidenza con livello di confidenza γ sarà allora:

$$\left(\frac{ns_0^2}{\chi_{n-1}^2 \left(\frac{1+\gamma}{2}\right)}; \frac{ns_0^2}{\chi_{n-1}^2 \left(\frac{1-\gamma}{2}\right)}\right)$$

Metodi per la stima dei parametri: stima puntuale

Metodologie di stima dei parametri

Il metodo per analogia

Nei precedenti paragrafi abbiamo definito gli stimatori puntuali *media campionaria* e *varianza campionaria*; tali stimatori sono stati introdotti di fatto con il cosiddetto **metodo per analogia**: i parametri sono infatti stati stimati per mezzo delle grandezze empiriche corrispondenti alle caratteristiche stesse che volevamo stimare.

Spesso però i parametri da stimare non sono media e varianza, perciò non è possibile ricorrere semplicemente al loro significato ed utilizzare il metodo per analogia. È perciò necessario introdurre delle metodologie diverse per eseguire la stima dei parametri.

Le metodologie di stima di parametri o caratteristiche della popolazione

Sia dato un campione casuale di $n \geq 1$ osservazioni, aventi una certa densità f :

$$X_1, X_2, \dots, X_n \text{ i.i.d. } \sim f(x, \vartheta_1, \vartheta_2, \dots, \vartheta_m), \quad n, m \geq 1$$

Dove almeno uno degli $m \geq 1$ parametri è incognito. Si supponga di voler stimare i parametri incogniti (o il parametro incognito) ϑ_i con uno stimatore opportuno $\hat{\vartheta}_i$, oppure di voler stimare una sintesi di tale parametro, detta *caratteristica della popolazione* k , ovvero una funzione dipendente solo dai parametri in questione:

$$k = k(\vartheta_1, \vartheta_2, \dots, \vartheta_m)$$

Mediante uno stimatore che indichiamo con \hat{k} .

Le metodologie possibili per eseguire tali operazioni sono diverse, ma tutte tengono conto sia di informazioni teoriche (come il tipo di densità in analisi), sia dei dati effettivamente raccolti. In particolare, le metodologie che studieremo sono 2:

1. Il *metodo dei momenti*, introdotto da Karl Pearson alla fine dell'Ottocento
2. Il *metodo di massima verosimiglianza*, introdotto negli anni Venti del secolo scorso da Ronald Fisher.

Il metodo dei momenti

Indichiamo i momenti relativi al campione casuale dato mediante la seguente simbologia:

$$\begin{aligned}\mu_1(\vartheta_1, \vartheta_2, \dots, \vartheta_m) &= E[X] \\ \mu_2(\vartheta_1, \vartheta_2, \dots, \vartheta_m) &= E[X^2] \\ &\dots \\ \mu_m(\vartheta_1, \vartheta_2, \dots, \vartheta_m) &= E[X^m]\end{aligned}$$

Come messo in evidenza, sono di interesse solamente i primi m momenti, dove m è il numero di parametri da stimare (si è ipotizzato che si debbano stimare tutti i parametri della distribuzione di probabilità).

L'idea di base del metodo dei momenti è quella di:

1. Stimare la media della distribuzione con il momento primo campionario (cioè la media campionaria):

$$\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j := M_1$$

2. Stimare il generico momento r -esimo della distribuzione con il momento campionario r -esimo:

$$\frac{1}{n} \sum_{j=1}^n X_j^r := M_r$$

Si usano cioè le versioni empiriche dei momenti, dette appunto **momenti campionari**, per stimare i momenti reali della distribuzione. Tali stimatori (anche se non lo dimostriamo) sono non distorti e consistenti.

Una volta eseguita la stima di tutti i primi m momenti della distribuzione, è possibile costruire un sistema di m equazioni in m incognite, del tipo:

$$\begin{cases} \mu_1(\vartheta_1, \vartheta_2, \dots, \vartheta_m) = M_1 \\ \dots \\ \mu_m(\vartheta_1, \vartheta_2, \dots, \vartheta_m) = M_m \end{cases}$$

Dove, naturalmente, le incognite sono i parametri $\vartheta_1, \vartheta_2, \dots, \vartheta_m$. Allora, se il sistema ammette soluzione, le soluzioni del sistema sono delle funzioni dei momenti campionari, i quali sono per definizione delle statistiche, perciò è chiaro che le soluzioni ottenute sono a loro volta delle statistiche, e di conseguenza possono essere usate come stimatori dei parametri ignoti.

In altri termini, la soluzione del sistema (se esiste) $\hat{\vartheta}_1, \hat{\vartheta}_2, \dots, \hat{\vartheta}_m$ è costituita da statistiche, che prendono il nome di *stimatori di $\vartheta_1, \vartheta_2, \dots, \vartheta_m$ ottenuti con il metodo stimatore dei momenti*.

Osservazioni

Si noti che le uniche informazioni che vengono usate in questo caso sono i momenti; può tuttavia darsi che delle distribuzioni molto diverse tra loro abbiano gli stessi momenti, e questo ci fa facilmente capire che il metodo dei momenti, pur essendo un metodo molto semplice, è un moto scarsamente "preciso".

Esempio di applicazione n. 1: distribuzione gamma

Si supponga di disporre di un campione di n misurazioni, e si ipotizzi che le misurazioni siano tutte modellabili con variabili aleatorie indipendenti con distribuzione di probabilità $\Gamma(\alpha, \beta)$. Si ipotizzi inoltre che α e β siano parametri incogniti e che si desideri fornirne delle stime $\hat{\alpha}$ e $\hat{\beta}$.

Allora, seguendo il procedimento descritto poco fa, calcoliamo il momento primo \bar{X} e il momento secondo M_2 (che saranno semplicemente dei numeri ottenuti a partire dai dati sperimentali), e costruiamo il sistema:

$$\begin{cases} \mu_1(\alpha, \beta) = \bar{X} \\ \mu_2(\alpha, \beta) = M_2 \end{cases}$$

Ricordando che in una distribuzione gamma la media è data dal prodotto tra i due parametri:

$$\mu_1(\alpha, \beta) = \alpha\beta$$

E che vale la relazione:

$$\mu_2(\alpha, \beta) - [\mu_1(\alpha, \beta)]^2 = Var(X) = \sigma^2 \rightarrow \mu_2(\alpha, \beta) = \sigma^2 + [\mu_1(\alpha, \beta)]^2 = \sigma^2 + (E[X])^2$$

Dove, nel caso di distribuzione gamma:

$$\sigma^2 = \alpha\beta^2$$

Otteniamo:

$$\begin{cases} \alpha\beta = \bar{X} \\ \alpha\beta^2 + (E[X])^2 = M_2 \end{cases} \rightarrow \begin{cases} \alpha\beta = \bar{X} \\ \alpha\beta^2 = M_2 - (E[X])^2 \end{cases} \rightarrow \begin{cases} \alpha\beta = \bar{X} \\ \alpha\beta^2 = M_2 - \bar{X}^2 \end{cases}$$

Notando che:

$$\sum_{j=1}^n (X_j - \bar{X})^2 = \sum_{j=1}^n X_j^2 - 2\bar{X} \sum_{j=1}^n X_j + n\bar{X}^2 = nM_2 - 2n\bar{X}^2 + n\bar{X}^2 = nM_2 - n\bar{X}^2$$

Possiamo riscrivere il sistema come segue:

$$\begin{cases} \alpha\beta = \bar{X} \\ \alpha\beta^2 = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})^2 \end{cases}$$

Ricordando poi che lo stimatore varianza campionaria è definito come:

$$S^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X})^2$$

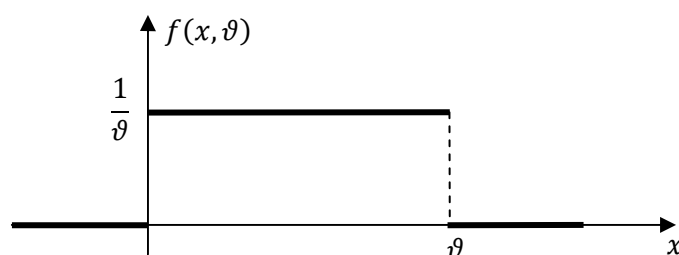
Otteniamo facilmente:

$$\begin{cases} \alpha\beta = \bar{X} \\ \alpha\beta^2 = \frac{n-1}{n} S^2 \end{cases} \rightarrow \begin{cases} \frac{n-1}{n} \frac{S^2}{\beta} = \bar{X} \\ \alpha = \frac{n-1}{n} \frac{S^2}{\beta^2} \end{cases} \rightarrow \begin{cases} \hat{\beta}_{mom} = \frac{S^2}{\bar{X}} \frac{n-1}{n} \\ \hat{\alpha}_{mom} = \frac{\bar{X}}{\beta} = \frac{n-1}{n} \frac{\bar{X}^2}{S^2} \end{cases}$$

Esempio di applicazione n. 2: distribuzione uniforme tra 0 e ϑ

Un altro esempio possibile riguarda il caso in cui il campione dato sia costituito da variabili uniformemente distribuite nell'intervallo $[0, \vartheta]$, $\vartheta > 0$:

$$X_1, X_2, \dots, X_n \text{ i. i. d. } \sim \mathcal{U}(0, \vartheta)$$



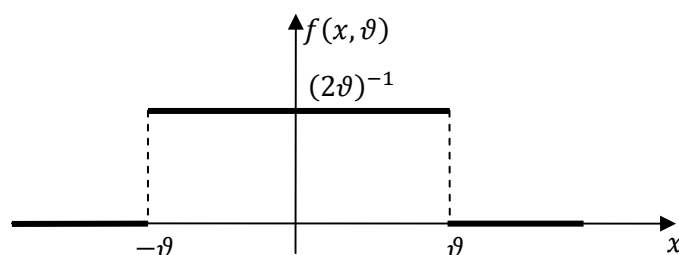
Se utilizziamo il metodo dei momenti, allora, avendo un solo parametro, calcoliamo solamente il momento primo, ovvero la media campionaria \bar{X} , e poniamo:

$$\bar{X} = \frac{\vartheta}{2} \rightarrow \hat{\vartheta}_{mom} = 2\bar{X}$$

Esempio di applicazione n. 3: distribuzione uniforme tra $-\vartheta$ e ϑ

Se invece il campione è costituito da variabili uniformemente distribuite tra $-\vartheta$ e ϑ , allora:

$$X_1, X_2, \dots, X_n \text{ i. i. d. } \sim \mathcal{U}(-\vartheta, \vartheta)$$



Allora, in questo caso abbiamo necessariamente:

$$\mu_1(\vartheta) = 0$$

Di conseguenza, viene a mancare la dipendenza del momento primo dal parametro ϑ stesso, e quindi non lo si può usare per il calcolo dello stimatore di ϑ . In questo caso, si osserva inoltre che, detta X una variabile aleatoria con la distribuzione del tipo in analisi, si ha:

$$E[X^{2k-1}] = 0, \quad \forall k \in \mathbb{N}$$

Possiamo però considerare un momento di ordine r , con un r pari. In particolare, il metodo prevede che si scelga sempre il momento con ordine inferiore, perciò in questo caso:

$$\mu_2(\vartheta) = \text{VAR}[X] + \mu_1(\vartheta)^2 = \frac{(2\vartheta)^2}{12} = \frac{\vartheta^2}{3}$$

Di conseguenza, calcoleremo dai dati sperimentali il momento secondo M_2 , e useremo lo stimatore:

$$\hat{\vartheta}_{mom} = \sqrt{3M_2}$$

Questo esempio in ogni caso ci permette di dedurre che si ha una certa arbitrarietà, in quanto sarebbe teoricamente possibile utilizzare dei momenti qualsiasi per eseguire il calcolo degli stimatori, purché siano in numero uguale al numero dei parametri da stimare.

Il metodo di massima verosimiglianza

Introduzione

Lo stimatore di massima verosimiglianza, spesso indicato con l'acronimo MLE (*Maximum Likelihood Estimator*), è uno stimatore creato sulla base di un maggior numero di informazioni teoriche rispetto a quelle utilizzate dal metodo dei momenti, in quanto si tiene conto del tipo di densità dei campioni.

Spiegazione del metodo

Si consideri il campione:

$$X_1, X_2, \dots, X_n \text{ i. i. d. } \sim f(x, \boldsymbol{\vartheta})$$

Dove con $\boldsymbol{\vartheta}$ indichiamo un vettore di n elementi, contenente i parametri della distribuzione:

$$\boldsymbol{\vartheta} = [\vartheta_1, \vartheta_2, \dots, \vartheta_n]$$

Allora, la densità congiunta delle variabili aleatorie X_1, X_2, \dots, X_n è:

$$f(x_1, x_2, \dots, x_n, \boldsymbol{\vartheta}) = \prod_{j=1}^n f(x_j, \boldsymbol{\vartheta})$$

Si può allora pensare a tale funzione non come ad una funzione di x_1, x_2, \dots, x_n , bensì come ad una funzione dei parametri incogniti, ovvero di $\boldsymbol{\vartheta}$. Tale funzione prende il nome di *funzione di verosimiglianza del campione*, e viene indicata con:

$$L_{\boldsymbol{\vartheta}}(x_1, x_2, \dots, x_n) = \prod_{j=1}^n f(x_j, \boldsymbol{\vartheta})$$

Si supponga che la distribuzione di probabilità f sia una distribuzione discreta (questa scelta è legata solo al fatto che ciò comporta maggiore semplicità espositiva, ma quanto diremo vale anche in caso di distribuzioni continue). In tal caso, si ha:

$$L_{\boldsymbol{\vartheta}}(x_1, x_2, \dots, x_n) = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

Risulta quindi chiaro che è opportuno stimare $\boldsymbol{\vartheta}$ in modo tale che la funzione di verosimiglianza sia massima, in quanto ciò significa massimizzare la probabilità che le rilevazioni raccolte (x_1, x_2, \dots, x_n) siano state ottenute dalla distribuzione di probabilità f con parametri $\boldsymbol{\vartheta}$.

Si individua così $\boldsymbol{\vartheta} \in \Theta$, dove Θ è lo *spazio parametrico*, ovvero l'insieme dei possibili valori di $\boldsymbol{\vartheta}$, tale che:

$$\max_{\boldsymbol{\vartheta} \in \Theta} L_{\boldsymbol{\vartheta}}(x_1, x_2, \dots, x_n)$$

Definizione

Data la realizzazione campionaria x_1, x_2, \dots, x_n di X_1, X_2, \dots, X_n , sia

$$L_{\boldsymbol{\vartheta}}(x_1, x_2, \dots, x_n) = \prod_{j=1}^n f(x_j, \boldsymbol{\vartheta})$$

La *funzione di verosimiglianza*; allora, se esiste $\boldsymbol{T} = g(x_1, x_2, \dots, x_n)$ tale che:

$$L_{\boldsymbol{T}}(x_1, x_2, \dots, x_n) = \max_{\boldsymbol{\vartheta} \in \Theta} L_{\boldsymbol{\vartheta}}(x_1, x_2, \dots, x_n)$$

\boldsymbol{T} viene detto *stimatore di massima verosimiglianza di $\boldsymbol{\vartheta}$* e viene indicato con $\hat{\boldsymbol{\vartheta}}_{MLE}$ (dove MLE sta per *Maximum Likelihood Estimator*) o $\hat{\boldsymbol{\vartheta}}_{ML}$ (*Maximum Likelihood*). Se inoltre si desidera stimare una caratteristica $k = k(\boldsymbol{\vartheta})$ della distribuzione, allora il suo stimatore di massima verosimiglianza è semplicemente dato da:

$$\hat{k}_{MLE} = k(\hat{\boldsymbol{\vartheta}}_{MLE})$$

Osservazioni

1. Lo stimatore si trova necessariamente nello spazio parametrico, quindi tutti i parametri assumeranno sempre valori appartenenti al loro dominio. Questo non era invece garantito con il metodo dei momenti.
2. Se si ottengono le stesse osservazioni ma con un diverso ordine (ad esempio, i campioni $\langle 1, 2, 3 \rangle$ e $\langle 3, 1, 2 \rangle$) allora la stima ottenuta è la stessa. Questa proprietà, valida anche per il metodo dei momenti, viene detta **proprietà di simmetria** e ci indica che a tutte le osservazioni viene attribuita uguale importanza.
3. Il massimo della funzione di verosimiglianza potrebbe non esistere o non essere unico, perciò lo stimatore di massima verosimiglianza in tali situazioni non esiste o non è unico.
4. Per individuare il massimo della funzione di verosimiglianza, è utile tenere conto che:
 - a) Nel caso più semplice, nel quale la funzione di verosimiglianza è derivabile e dipende da una sola variabile, basta calcolare la derivata, uguagliarla a zero e, utilizzando le semplici regole dell'Analisi Matematica, verificare se si tratta di un punto di massimo, di minimo o di un flesso.
 - b) Siccome la funzione di verosimiglianza è sempre positiva, è talvolta utile cercare il massimo della funzione:

$$\log L_{\vartheta}(x_1, x_2, \dots, x_n)$$

Anziché quello di $L_{\vartheta}(x_1, x_2, \dots, x_n)$ stessa: essendo il logaritmo un operatore monotono, i punti di massimo saranno necessariamente gli stessi. Questo aiuta soprattutto perché la funzione di verosimiglianza viene ottenuta dal prodotto di molte funzioni.

- c) In alcuni casi, la funzione non è derivabile, e quindi è opportuno appoggiarsi ad un grafico della funzione di verosimiglianza.
- d) Se la funzione di verosimiglianza dipende da più parametri, allora l'individuazione del massimo è più complessa, in quanto è necessario valutare i punti nei quali si annullano le derivate parziali, e quindi valutare l'Hessiano; tuttavia, nelle nostre applicazioni non verrà mai richiesta la valutazione dell'Hessiano.

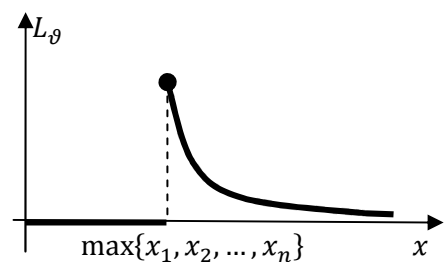
Esempio di applicazione n. 1: distribuzione uniforme tra 0 e ϑ

Prendiamo come esempio un campione costituito da variabili uniformemente distribuite nell'intervallo $[0, \vartheta]$, $\vartheta > 0$ (lo stesso esempio che abbiamo considerato parlando del metodo dei momenti):

$$X_1, X_2, \dots, X_n \text{ i. i. d. } \sim \mathcal{U}(0, \vartheta)$$

In questo caso, abbiamo:

$$\begin{aligned} L_{\vartheta}(x_1, x_2, \dots, x_n) &= \prod_{j=1}^n f(x_j, \vartheta) = \\ &= \frac{1}{\vartheta} \mathbb{I}_{[0, \vartheta]}(x_1) \cdot \frac{1}{\vartheta} \mathbb{I}_{[0, \vartheta]}(x_2) \cdot \dots \cdot \frac{1}{\vartheta} \mathbb{I}_{[0, \vartheta]}(x_n) = \frac{1}{\vartheta^n} \mathbb{I}_{[0, \vartheta]} \left(\max_{j=1, \dots, n} x_j \right) = \\ &= \frac{1}{\vartheta^n} \mathbb{I}_{(\max\{x_1, x_2, \dots, x_n\}; +\infty)}(\vartheta) \end{aligned}$$



Come si osserva dal grafico della funzione di verosimiglianza, il punto di massimo è proprio coincidente con il massimo delle rilevazioni effettuate, quindi avremo:

$$\hat{\vartheta}_{MLE} = \max\{x_1, x_2, \dots, x_n\}$$

Confronto tra i due metodi

Come si osserva dall'esempio relativo alla distribuzione uniforme tra 0 e ϑ , gli stimatori che si ottengono con i due metodi risultano talvolta completamente diversi tra loro. È quindi opportuno domandarsi quale dei due metodi risulti essere migliore.

Criterio di valutazione dell'accuratezza

Un possibile criterio da utilizzare è quello che consiste nel valutare l'accuratezza. Si vuole cioè verificare se si ha:

$$\forall \delta P(|\hat{\vartheta}_{mom} - \vartheta| < \delta) \leq P(|\hat{\vartheta}_{MLE} - \vartheta| < \delta)$$

Allora, significa che $\hat{\vartheta}_{MLE}$ fornisce una stima meno accurata di quella fornita da $\hat{\vartheta}_{mom}$, perciò è opportuno utilizzare lo stimatore $\hat{\vartheta}_{MLE}$.

Tuttavia, questo criterio è di difficile applicazione nella pratica.

Confronto degli errori quadratici medi

Un criterio diverso è quello che consiste nel confrontare tra loro gli errori quadratici medi:

$$\begin{aligned} MSE[\hat{\vartheta}_{mom}] &= E[(\hat{\vartheta}_{mom} - \vartheta)^2] \\ MSE[\hat{\vartheta}_{MLE}] &= E[(\hat{\vartheta}_{MLE} - \vartheta)^2] \end{aligned}$$

Se si ha poi:

$$MSE[\hat{\vartheta}_{mom}] > MSE[\hat{\vartheta}_{MLE}]$$

Allora è preferibile utilizzare lo stimatore $\hat{\vartheta}_{MLE}$.

Un caso concreto

Tornando all'esempio di partenza, si può dimostrare che si ha:

$$\begin{aligned} MSE[\hat{\vartheta}_{mom}] &= E[(2\bar{X} - \vartheta)^2] = VAR[2\bar{X} - \vartheta] + E^2[2\bar{X} - \vartheta] = VAR[2\bar{X}] + E^2[2\bar{X} - \vartheta] = \\ &= 4VAR[\bar{X}] = 4 \frac{1}{n} \frac{\vartheta^2}{12} = \frac{\vartheta^2}{3n} \end{aligned}$$

Si dimostra inoltre che si ha:

$$\begin{aligned} VAR[\hat{\vartheta}_{MLE}] &= \frac{n\vartheta^2}{(n+1)^2(n+2)} \\ E[\hat{\vartheta}_{MLE}] &= \frac{n\vartheta}{n+1} \end{aligned}$$

E quindi, eseguendo i conti:

$$MSE[\hat{\vartheta}_{MLE}] = VAR[\hat{\vartheta}_{MLE} - \vartheta] + E^2[\hat{\vartheta}_{MLE} - \vartheta] = \frac{2\vartheta^2}{(n+1)(n+2)}$$

Osserviamo allora che:

1. $\hat{\vartheta}_{MLE}$ è distorto, perché la sua media non è ϑ . È però asintoticamente non distorto (la sua media tende a ϑ , per n tendente all'infinito).
2. $\hat{\vartheta}_{MLE}$ è consistente. Tuttavia, non è detto che uno stimatore di massima verosimiglianza lo sia.
3. $\hat{\vartheta}_{MLE}$ è da preferirsi a $\hat{\vartheta}_{mom}$ perché il suo MSE tende a zero più rapidamente di quello di $\hat{\vartheta}_{mom}$.

Ricerca dello stimatore ottimo

La ricerca dello stimatore con il minor MSE

Come abbiamo visto, il confronto tra due stimatori risulta talvolta complesso, e talvolta risulta del tutto impossibile, in quanto l'errore quadratico medio potrebbe dipendere dai parametri incogniti.

Fatte queste considerazioni, risulta allora utile chiedersi se, considerando la classe di tutti i possibili stimatori di un certo parametro o di una certa caratteristica, è possibile individuare lo stimatore che ha l'errore quadratico medio più piccolo uniformemente rispetto al valore del parametro incognito (cioè per ogni valore ammissibile del parametro incognito stesso). In altri termini, dato il campione:

$$X_1, X_2, \dots, X_n \text{ i.i.d. } \sim f(x, \vartheta)$$

E data la caratteristica:

$$k = k(\vartheta)$$

Definiamo:

$$\mathcal{T} = \{T: T \text{ è uno stimatore di } k \text{ che ammette MSE}\}$$

Vogliamo individuare, se esiste, lo stimatore:

$$T^* \in \mathcal{T} : \text{MSE}(T^*) \leq \text{MSE}(T) \quad \forall T \in \mathcal{T}, \forall \vartheta$$

Si può però dimostrare che questo problema non ammette soluzione.

Esempio

Consideriamo come esempio il caso in cui la caratteristica in analisi sia semplicemente la media:

$$k = \mu$$

Allora, come noto, possiamo scegliere di utilizzare come stimatore la media campionaria $T_1 = \bar{X}$. L'errore quadratico medio che si ottiene in tal caso, come già dimostrato, è pari a:

$$\text{MSE}(T_1) = \frac{\sigma^2}{n} > 0$$

Tuttavia, è possibile anche scegliere di utilizzare degli stimatori banali, come ad esempio lo stimatore:

$$T_2 = 10.23$$

Questo stimatore nella pratica non ha chiaramente alcuna utilità, perché si tratta di una costante scelta senza neppure basarsi sui dati raccolti, ma rimane pur sempre, a livello formale, un possibile stimatore di μ . L'errore quadratico medio in tal caso è così calcolabile:

$$\text{MSE}(T_2) = \text{Var}(T_2) + E[(T_2 - \mu)^2] = \text{Var}(10.23) + E[(E[10.23] - \mu)^2] = E[(E[10.23] - \mu)^2]$$

Nell'ipotesi particolarmente fortunata in cui la media sia effettivamente 10.23, l'errore quadratico medio di T_2 risulta quindi nullo, mentre è abbastanza ovvio capire che se la media si discosta molto di tale valore, l'errore quadratico medio di T_1 è inferiore a quello di T_2 . In sostanza quindi non esiste alcuno stimatore che abbia un errore quadratico medio inferiore a quello di T_2 per ogni possibile valore di μ , ma lo stesso stimatore di μ ha un errore quadratico superiore a quello di altri stimatori per diversi valori di μ .

Restrizione del problema

Per rendere risolvibile il problema è allora necessario restringere la classe considerata, eliminando tutti gli stimatori banali. Siccome sappiamo che, per un generico stimatore vale la relazione:

$$\text{MSE}(T) = \text{Var}(T) + [E(T) - \vartheta]^2$$

E che, se lo stimatore non è distorto:

$$E(T) = \vartheta \quad \rightarrow \quad \text{MSE}(T) = \text{VAR}(T)$$

Consideriamo solamente la classe \mathcal{T}_U degli stimatori non distorti:

$$\mathcal{T}_U = \{T: T \text{ è stimatore di } k \text{ e } E(T) = k \quad \forall \vartheta\}$$

in modo che individuare lo stimatore ottimo si riconduca a ricercare lo stimatore con la minima varianza, ovvero, trovare quel T^* tale che:

$$\text{VAR}(T^*) \leq \text{VAR}(T) \quad \forall T \in \mathcal{T}_U, \forall \vartheta$$

Questo problema, a differenza del precedente, ammette soluzione.

Stimatore UMVUE

Dato un campione

$$X_1, X_2, \dots, X_n \text{ i. i. d. } \sim f(\cdot, \vartheta), \quad \vartheta \in \Theta$$

e dato lo stimatore T^* della caratteristica $k(\vartheta)$, diciamo che T è lo stimatore non distorto a varianza uniformemente minima (UMVUE, *Uniform Minimum Variance Unbiased Estimator*) se soddisfa le due condizioni seguenti:

1. T^* non è distorto per $k(\vartheta)$
2. $VAR(T^*) \leq VAR(T) \quad \forall T \in \mathcal{T}_U, \forall \vartheta$

Dove \mathcal{T}_U è l'insieme di tutti e soli gli stimatori non distorti di $k(\vartheta)$ a varianza finita:

$$\mathcal{T}_U = \{T: T \text{ è stimatore di } k \text{ e } E(T) = k \forall \vartheta\}$$

Esistenza ed unicità dello stimatore non distorto

- Per alcune caratteristiche, è possibile che non esista alcuno stimatore non distorto. Ciò naturalmente non significa che non sia possibile stimare tale caratteristica. Consideriamo ad esempio un campione costituito da un solo elemento:

$$X_1 \sim Bin(5, \vartheta), \quad 0 < \vartheta < 1$$

E ipotizziamo di voler stimare la caratteristica:

$$k = \frac{1}{\vartheta}$$

La cosa più sensata da fare sarebbe quella di stimare:

$$\vartheta = \frac{1}{x_1}$$

E quindi stimeremo:

$$k = x_1$$

Essendo però x_1 un numero intero, è ovvio che non potrà essere vero che, per ogni valore di ϑ , la media dello stimatore è uguale a ϑ , e quindi lo stimatore è distorto.

- Se uno stimatore non distorto esiste, è possibile che ne esistano anche altri (possono anche essere infiniti). Consideriamo ad esempio il seguente campione:

$$X_1, X_2, \dots, X_n \sim Poi(\vartheta), \quad \vartheta > 0$$

Come noto, il parametro ϑ di una distribuzione di Poisson corrisponde sia alla media della variabile casuale, sia alla sua varianza. Possiamo allora stimare ϑ sia con la media campionaria, sia con la varianza campionaria, e sappiamo dalle osservazioni precedenti che in entrambi i casi otteniamo degli stimatori non distorti. Inoltre, possiamo considerare uno stimatore costruito come media pesata dei due precedenti, ovvero:

$$T = a\bar{X} + (1 - a)S^2, \quad 0 \leq a \leq 1$$

E si verifica facilmente che la media di T risulta essere proprio uguale a ϑ . Di conseguenza, concludiamo che se per una caratteristica esistono almeno due stimatori non distorti, allora è possibile costruire infiniti stimatori non distorti di quella stessa caratteristica.

Disuguaglianza di Fréchet-Cramer-Rao

Esempio introduttivo

Si prenda in analisi un campione con distribuzione di Poisson, costituito da una sola variabile aleatoria:

$$X_1 \sim \mathcal{P}(\vartheta)$$

Dove possiamo ad esempio ipotizzare che il campione rappresenti il numero di telefonate arrivate il primo giorno. Possiamo poi chiamare X_2 e X_3 il numero di telefonate arrivate, rispettivamente, il secondo ed il terzo giorno (che però non sono state rilevate, quindi non fanno parte del campione). Siamo quindi interessati a stimare la seguente caratteristica:

$$k = P(\text{nei 2 giorni successivi non arrivi alcuna telefonata}) = P(X_2 = 0, X_3 = 0)$$

Siccome siamo sotto l'ipotesi di indipendenza tra il numero di telefonate arrivate in ogni singola giornata, X_2 e X_3 sono variabili aleatorie indipendenti e quindi la probabilità congiunta è il prodotto delle singole probabilità:

$$k = P(X_2 = 0) \cdot P(X_3 = 0) = f_{\vartheta}(0) \cdot f_{\vartheta}(0) = e^{-\vartheta} e^{-\vartheta} = e^{-2\vartheta}$$

Per stimare k dovremo usare un opportuno stimatore, che indichiamo con:

$$T = g(x_1)$$

Dove x_1 è il valore realmente assunto dal numero di telefonate arrivate il primo giorno. Come abbiamo visto, è auspicabile che lo stimatore sia non distorto, perciò possiamo imporre questa caratteristica:

$$E(T) = e^{-2\vartheta}$$

Calcoliamo allora $E(T)$:

$$E(T) = \sum_{k=0}^{+\infty} g(k) f_{\vartheta}(k) = \sum_{k=0}^{+\infty} \left[g(k) \frac{e^{-\vartheta} \vartheta^k}{k!} \right]$$

Sostituendo questo valore nella precedente uguaglianza:

$$\sum_{k=0}^{+\infty} \left[\frac{e^{-\vartheta} \vartheta^k}{k!} g(k) \right] = e^{-2\vartheta}$$

Ora, dividendo entrambi i membri per $e^{-\vartheta}$, otteniamo:

$$\sum_{k=0}^{+\infty} \left[\frac{\vartheta^k}{k!} g(k) \right] = e^{-\vartheta}$$

Ricordando adesso la scomposizione della serie di Taylor di una funzione esponenziale:

$$e^{-\vartheta} = \sum_{k=0}^{+\infty} \left[(-1)^k \frac{\vartheta^k}{k!} \right]$$

La precedente uguaglianza diventa:

$$\sum_{k=0}^{+\infty} \left[\frac{\vartheta^k}{k!} g(k) \right] = \sum_{k=0}^{+\infty} \left[(-1)^k \frac{\vartheta^k}{k!} \right]$$

Da tale espressione ricaviamo in modo ovvio che lo stimatore T deve essere (quando si ha un unico dato):

$$T = g(x_1) = (-1)^{x_1}$$

Lo stimatore così ottenuto è però uno stimatore completamente privo di senso, in quanto stima una caratteristica k che, per sua natura, appartiene all'intervallo $(0,1)$, con un valore che è necessariamente -1 oppure 1 . Nonostante questo, lo stimatore risulta essere formalmente ottimale, perché non è distorto. Di conseguenza, non risulta particolarmente interessante ai nostri scopi stabilire qual è lo stimatore ottimale. Ci concentriamo piuttosto su come sia possibile stabilire la varianza minima di uno stimatore per la data caratteristica: in tal modo, potremo confrontare la varianza dello stimatore individuato con quella minima per gli stimatori della caratteristica in analisi, e se le due varianze sono coincidenti o comunque molto vicine, questo è indice della bontà dello stimatore considerato.

La disuguaglianza

Sia dato un campione:

$$X_1, X_2, \dots, X_n \text{ i.i.d. } \sim f(x, \vartheta), \quad \vartheta \in \Theta \subseteq \mathbb{R}$$

E si voglia stimare una caratteristica k della distribuzione dei dati:

$$k = k(\vartheta)$$

Sia inoltre T uno stimatore non distorto di k :

$$E(T) = k$$

Dotato di varianza finita:

$$VAR(T) < +\infty$$

- Ipotesi

1. Θ è un intervallo aperto.
2. f è una funzione di densità il cui supporto (ovvero l'insieme di valori di x nei quali la densità non è nulla) è indipendente dai parametri della distribuzione stessa:

$$S = \{x : f(x, \vartheta) > 0\} \text{ è indipendente da } \vartheta$$

3. La funzione $\theta \mapsto f(x, \theta)$, cioè la f letta come funzione di θ , è derivabile rispetto al parametro θ su tutto l'insieme Θ , per ogni valore di $x \in S$.
4. La media della derivata logaritmica di f è costantemente nulla (ovvero, per ogni ϑ):

$$E_{\vartheta} \left[\frac{d}{d\vartheta} \log f(X_1, \vartheta) \right] = 0 \quad \forall \vartheta \in \Theta$$

5. Si abbia inoltre:

$$0 < E_{\vartheta} \left[\left(\frac{d}{d\vartheta} \log f(X_1, \vartheta) \right)^2 \right] < +\infty$$

6. $k: \Theta \rightarrow \mathbb{R}$ è una funzione derivabile su tutto l'insieme Θ si abbia:

$$k'(\vartheta) = E_{\vartheta} \left(T \frac{d}{d\vartheta} \log L_{\vartheta}(X_1, \dots, X_n) \right) \quad \forall \vartheta \in \Theta$$

- Tesi

Allora, si ha:

$$Var(T) \geq \frac{(k'(\vartheta))^2}{n \cdot I(\vartheta)} \quad \forall \vartheta \in \Theta$$

Dove $I(\vartheta)$, detta anche **informazione di Fisher**, è definita come:

$$I(\vartheta) = E \left[\left(\frac{d}{d\vartheta} \log f(X_1, \vartheta) \right)^2 \right]$$

Inoltre, la precedente disuguaglianza diventa un'uguaglianza, del tipo:

$$Var(T) = \frac{(k'(\vartheta))^2}{n \cdot I(\vartheta)} \quad \forall \vartheta \in \Theta$$

Se e solo se esiste una funzione $a(n, \vartheta)$ tale che:

$$P_{\vartheta} \left(\frac{d}{d\vartheta} \log L_{\vartheta}(X_1, \dots, X_n) = a(n, \vartheta) \cdot (T - k(\vartheta)) \right) = 1 \quad \forall \vartheta \in \Theta$$

Osservazione

Da questa disuguaglianza si ricava di fatto che il tasso ottimale di decrescita della varianza in funzione della dimensione n del campione è n^{-1} . Naturalmente, ciò però vale solo a patto che valgano tutte le ipotesi precedentemente enunciate. Ad esempio, se la distribuzione fosse uniforme, tali ipotesi non risulterebbero verificate, e infatti in questa situazione si può ottenere anche un tasso di decrescita del tipo n^{-2} . Si nota inoltre che:

$$n I(\vartheta) = I_n(\vartheta) = E \left[\left(\frac{d}{d\vartheta} \log L_{\vartheta}(x_1, x_2, \dots, x_n, \vartheta) \right)^2 \right]$$

L'importanza delle ipotesi

Tutte le ipotesi che abbiamo introdotto sono delle ipotesi di regolarità che riguardano la famiglia di tutte le densità $f(x, \vartheta)$ con $\vartheta \in \Theta$.

- Ipotesi 1

La prima delle nostre ipotesi risulta essere importante per la derivabilità della funzione: su un intervallo chiuso infatti la funzione non risulta derivabile nell'estremo dell'intervallo stesso.

- Ipotesi 2

a) Consideriamo come esempio una distribuzione esponenziale:

$$f(x, \vartheta) = \mathcal{E}(\vartheta)$$

Allora in questo caso il supporto della funzione è:

$$S = \{x : f(x, \vartheta) > 0\} = (0, +\infty)$$

Quindi, in questo caso, l'ipotesi risulta essere verificata, perché l'intervallo è indipendente da ϑ .

b) Prendiamo ora in analisi una distribuzione di Poisson:

$$f(x, \vartheta) = \mathcal{P}(\vartheta)$$

In tal caso, il supporto è:

$$S = \{0, 1, 2, 3, \dots\} = \mathbb{N}$$

E quindi anche in tale situazione l'ipotesi è verificata.

c) Se consideriamo invece una distribuzione uniforme:

$$f(x, \vartheta) = \mathcal{U}(0, \vartheta)$$

Abbiamo allora:

$$S = [0, \vartheta]$$

E perciò l'ipotesi non è soddisfatta (come avevamo già accennato, per il modello uniforme non possiamo applicare le conclusioni che derivano da questo teorema).

- Ipotesi 3 e 4

Supponiamo che la variabile aleatoria in analisi sia assolutamente continua (cioè che diremo vale però anche per variabili discrete). Come noto, abbiamo:

$$\int_{-\infty}^{+\infty} f(x, \vartheta) dx = \int_S f(x, \vartheta) dx = 1$$

Derivando entrambi i membri dell'uguaglianza:

$$\frac{d}{d\vartheta} \int_S f(x, \vartheta) dx = 0$$

Siccome, in virtù dell'ipotesi 2, l'intervallo S è indipendente da ϑ , possiamo spostare il segno di derivazione all'interno del segno di integrale:

$$\int_S \frac{d}{d\vartheta} f(x, \vartheta) dx = 0$$

Stiamo inoltre considerando solo valori all'interno del supporto di f , perciò possiamo dividere e moltiplicare per $f(x, \vartheta)$, che è certamente un quantità non nulla:

$$\int_S \frac{\frac{d}{d\vartheta} f(x, \vartheta)}{f(x, \vartheta)} f(x, \vartheta) dx = 0$$

Ricordando le proprietà delle derivate, si ricava facilmente che la precedente uguaglianza equivale a:

$$\int_S \left(\frac{d}{d\vartheta} \log f(x, \vartheta) \right) f(x, \vartheta) dx = 0$$

Da cui:

$$E \left(\frac{d}{d\vartheta} \log f(X_1, \vartheta) \right) = 0$$

Stimatore efficiente

Uno stimatore non distorto si dice efficiente se la sua varianza è uguale al confine di Fréchet-Cramer-Rao.

Esempio

Testo

Sia dato il seguente campione:

$$X_1, X_2, \dots, X_n \text{ i. i. d. } \sim \mathcal{P}(\vartheta)$$

1. Si determini lo stimatore di ϑ con il metodo dei momenti, ovvero $\hat{\vartheta}_M$.
2. Se esiste, trovare uno stimatore T_1 di ϑ non distorto con varianza uguale al confine di Fréchet-Cramer-Rao.

Soluzione

• Punto 1

Per determinare lo stimatore $\hat{\vartheta}_M$ di ϑ con il metodo dei momenti, dobbiamo semplicemente porre:

$$\hat{\vartheta}_M = M_1 = \bar{X}$$

• Punto 2

Verifichiamo che le ipotesi del teorema di Fréchet-Cramer-Rao siano soddisfatte dalla distribuzione di probabilità di Poisson, ovvero:

$$f(x, \vartheta) = \begin{cases} \frac{e^{-\vartheta} \vartheta^x}{x!} & \text{se } x \in \mathbb{N} \\ 0 & \text{altrimenti} \end{cases} \quad \vartheta > 0$$

1. Il parametro ϑ è definito sull'intervallo $\Theta = (0, +\infty)$, che è un intervallo aperto, perciò la prima ipotesi risulta essere verificata.
2. Il supporto di $f(x, \vartheta)$ è costituito dall'insieme dei numero naturale \mathbb{N} , perciò anche la seconda ipotesi è verificata.
3. Supponiamo di aver fissato un certo valore di x appartenente al supporto S . Allora, risulta chiaro che la funzione di distribuzione, considerata come una variabile di ϑ , sarà:

$$f(x, \vartheta) = \frac{e^{-\vartheta} \vartheta^x}{x!}$$

Che è chiaramente una funzione infinitamente derivabile rispetto a ϑ , indipendentemente da quale valore di x appartenente ad S è stato scelto.

4. Calcoliamo ora:

$$E_{\vartheta} \left[\frac{d}{d\vartheta} \log f(X_1, \vartheta) \right]$$

Prima di tutto, valutiamo:

$$\begin{aligned} \frac{d}{d\vartheta} \log f(x_1, \vartheta) &= \frac{d}{d\vartheta} \log f_{X_1}(\vartheta) = \frac{d}{d\vartheta} \log \left(\frac{e^{-\vartheta} \vartheta^{x_1}}{x_1!} \right) = \\ \frac{d}{d\vartheta} [\log(e^{-\vartheta}) + \log \vartheta^{x_1} - \log x_1!] &= \frac{d}{d\vartheta} (-\vartheta) + x_1 \frac{d}{d\vartheta} \log \vartheta = -1 + \frac{1}{\vartheta} x_1 = \frac{x_1 - \vartheta}{\vartheta} \end{aligned}$$

Quindi possiamo calcolare:

$$E_{\vartheta} \left[\frac{d}{d\vartheta} \log f(X_1, \vartheta) \right] = E \left[\frac{X_1 - \vartheta}{\vartheta} \right] = \frac{1}{\vartheta} E[X_1 - \vartheta] = \frac{1}{\vartheta} [E(X_1) - \vartheta] = 0$$

In quanto, essendo X_1 una distribuzione di Poisson con parametro ϑ , è ovvio che la sua media si ϑ .

5. Calcoliamo ora:

$$I(\vartheta) = E_{\vartheta} \left[\left(\frac{d}{d\vartheta} \log f(X_1, \vartheta) \right)^2 \right] = E \left[\left(\frac{X_1 - \vartheta}{\vartheta} \right)^2 \right] = \frac{1}{\vartheta^2} E[(X_1 - \vartheta)^2] = \frac{1}{\vartheta^2} \text{VAR}(X_1) = \frac{1}{\vartheta^2} \vartheta = \frac{1}{\vartheta}$$

Perciò la quantità calcolata è necessariamente maggiore di zero e finita.

Proviamo ora a verificare se lo stimatore \bar{X} individuato al punto 1 soddisfa l'ipotesi numero 6:

$$k_1(\vartheta) = \vartheta \quad \rightarrow \quad k'_1(\vartheta) = 1$$

Calcoliamo ora la funzione di massima verosimiglianza:

$$L_\vartheta(X_1, \dots, X_n) = \prod_{j=1}^n \left(\frac{e^{-\vartheta} \vartheta^{x_j}}{x_j!} \right) = e^{-n\vartheta} \prod_{j=1}^n \left(\frac{\vartheta^{x_j}}{x_j!} \right) = e^{-n\vartheta} \frac{\vartheta^{\sum_{j=1}^n x_j}}{\prod_{j=1}^n (x_j!)} = e^{-n\vartheta} \frac{\vartheta^{n\bar{X}}}{\prod_{j=1}^n (x_j!)}$$

Quindi:

$$\log L_\vartheta(X_1, \dots, X_n) = \log \left(e^{-n\vartheta} \frac{\vartheta^{n\bar{X}}}{\prod_{j=1}^n (x_j!)} \right) = -n\vartheta + n\bar{X} \log \vartheta - \log \prod_{j=1}^n (x_j!)$$

Da cui:

$$\frac{d}{d\vartheta} \log L_\vartheta(X_1, \dots, X_n) = -n + \frac{n\bar{X}}{\vartheta}$$

Quindi:

$$E \left(T \frac{d}{d\vartheta} \log L_\vartheta(X_1, \dots, X_n) \right) = E \left(\bar{X} \left(-n + \frac{n\bar{X}}{\vartheta} \right) \right) = E \left(-n\bar{X} + \frac{n\bar{X}^2}{\vartheta} \right) = -n E(\bar{X}) + \frac{n}{\vartheta} E(\bar{X}^2)$$

Se ricordiamo che:

$$VAR(\bar{X}) = E(\bar{X}^2) - E^2(\bar{X}) \quad \rightarrow \quad E(\bar{X}^2) = VAR(\bar{X}) + E^2(\bar{X}) = \frac{\vartheta}{n} + \vartheta^2$$

Otteniamo:

$$E \left(T \frac{d}{d\vartheta} \log L_\vartheta(X_1, \dots, X_n) \right) = -n\vartheta + \frac{n}{\vartheta} \left(\frac{\vartheta}{n} + \vartheta^2 \right) = -n\vartheta + 1 + n\vartheta = 1$$

E quindi anche l'ipotesi 6 è effettivamente verificata.

A questo punto, possiamo osservare che, nel caso in cui si cerchi lo stimatore di ϑ , il confine di Fréchet-Cramer-Rao (FCR) è calcolabile come:

$$\frac{[k'(\vartheta)]^2}{n \cdot I(\vartheta)} = \frac{\vartheta}{n}$$

Che coincide con la varianza di \bar{X} :

$$VAR(\bar{X}) = \frac{\vartheta}{n}$$

Perciò la media campionaria è uno stimatore efficiente della media reale (ovvero ha varianza coincidente con il confine FCR) ed è proprio lo stimatore cercato

Proprietà degli stimatori di massima verosimiglianza

Proprietà n. 1: stimatori MLE ed efficienza

Enunciato

Dato un certo campione:

$$X_1, X_2, \dots, X_n \text{ i. i. d.}$$

E data la caratteristica k da stimare, se tutte le ipotesi del teorema di Fréchet-Cramer-Rao sono soddisfatte, si può affermare che se esiste uno stimatore efficiente \hat{k}_{MLE} di k , allora tale stimatore è lo stimatore di verosimiglianza.

Tuttavia, non è detto che uno stimatore di massima verosimiglianza sia anche efficiente.

Procedimento pratico che ne deriva

Se si cerca lo stimatore ottimo efficiente, si può quindi individuare lo stimatore di massima verosimiglianza della caratteristica in analisi e verificare se effettivamente lo stimatore così trovato è efficiente. In caso contrario, significa che non esiste alcuno stimatore efficiente della caratteristica in analisi.

Dimostrazione

Consideriamo il campione

$$X_1, X_2, \dots, X_n \text{ i. i. d.}$$

E chiamiamo k la caratteristica da stimare. Ipotizziamo che siano verificate tutte le ipotesi della disuguaglianza di FCR, e chiamiamo l_{FCR} il limite di Fréchet-Cramer-Rao. Sia inoltre \hat{k}_{MLE} lo stimatore di k ottenuto con il metodo di massima verosimiglianza.

1. Supponiamo ora che esista un stimatore efficiente per k . Allora, sfruttando la validità delle ipotesi della disuguaglianza di FCR:

$$P_{\vartheta} \left(\frac{d}{d\vartheta} \log L_{\vartheta}(x_1, \dots, x_n) = a(n, \vartheta) \cdot (T - \vartheta) \right) = 1 \quad \forall \vartheta \in \Theta$$

Dove T è uno stimatore non distorto di ϑ e $a(n, \vartheta)$ è un'opportuna funzione.

2. Dire che \hat{k}_{MLE} è lo stimatore di massima verosimiglianza di k equivale ad affermare che:

$$\hat{k}_{MLE} = k(\hat{\vartheta}_{MLE})$$

Sfruttando poi la regolarità derivante dalla validità delle ipotesi della disuguaglianza di FCR, possiamo affermare che ciò implica anche che:

$$\left[\frac{\partial}{\partial \vartheta} \log L_{\vartheta}(x_1, \dots, x_n) \right]_{\vartheta = \hat{\vartheta}_{MLE}} = 0$$

3. Siccome la formula enunciata al punto 1 deve valere per ogni valore di ϑ , vale certamente anche nel caso $\vartheta = \hat{\vartheta}_{MLE}$, perciò possiamo sostituire $a(n, \vartheta) \cdot (T - \vartheta)$ all'interno dell'equazione appena scritta:

$$[a(n, \vartheta) \cdot (T - \vartheta)]_{\vartheta = \hat{\vartheta}_{MLE}} = 0$$

Oververo:

$$a(n, \hat{\vartheta}_{MLE}) \cdot (T - \hat{\vartheta}_{MLE}) = 0$$

4. Possiamo dimostrare che $a(n, \vartheta) \neq 0$ per ogni valore di ϑ . Infatti, se chiamiamo:

$$nI(\vartheta) = I_n(\vartheta)$$

Abbiamo:

$$\frac{\partial}{\partial \vartheta} \log L_{\vartheta}(x_1, \dots, x_n) = \frac{\partial}{\partial \vartheta} \left[\log \prod_{i=1}^n f(X_i, \vartheta) \right] = \frac{\partial}{\partial \vartheta} \left[\sum_{i=1}^n \log f(X_i, \vartheta) \right] = \sum_{i=1}^n \frac{\partial}{\partial \vartheta} \log f(X_i, \vartheta)$$

Quindi:

$$\begin{aligned} VAR \left(\frac{\partial}{\partial \vartheta} \log L_{\vartheta}(x_1, \dots, x_n) \right) &= VAR \left(\sum_{i=1}^n \frac{\partial}{\partial \vartheta} \log f(X_i, \vartheta) \right) = n VAR \left(\frac{\partial}{\partial \vartheta} \log f(X_1, \vartheta) \right) = \\ &= n E \left[\left(\frac{\partial}{\partial \vartheta} \log f(X_1, \vartheta) \right)^2 \right] - n \left[E \left[\frac{\partial}{\partial \vartheta} \log f(X_1, \vartheta) \right] \right]^2 \end{aligned}$$

Ma, per l'ipotesi numero 4 della disuguaglianza di FCR (che abbiamo assunto essere valida:

$$E \left[\frac{\partial}{\partial \vartheta} \log f(X_1, \vartheta) \right] = 0$$

Perciò:

$$VAR \left(\frac{\partial}{\partial \vartheta} \log L_{\vartheta}(x_1, \dots, x_n) \right) = n E \left[\left(\frac{\partial}{\partial \vartheta} \log f(X_1, \vartheta) \right)^2 \right] = n I(\vartheta) = I_n(\vartheta)$$

Come conseguenza dell'equazione riportata al punto 1:

$$I_n(\vartheta) = \text{VAR} \left(\frac{\partial}{\partial \vartheta} \log L_{\vartheta}(x_1, \dots, x_n) \right) = \text{VAR}(a(n, \vartheta) \cdot (T - \vartheta)) = a^2(n, \vartheta) \cdot \text{VAR}(T)$$

Ma, per l'ipotesi numero 4 della disuguaglianza di FCR:

$$I(\vartheta) = E \left[\left(\frac{\partial}{\partial \vartheta} \log f(X_1, \vartheta) \right)^2 \right] > 0$$

E perciò:

$$I_n(\vartheta) = nI(\vartheta) > 0 \rightarrow a^2(n, \vartheta) \cdot \text{VAR}(T) > 0$$

E da questa disuguaglianza, risulta ovvio che:

$$a(n, \vartheta) > 0$$

5. Da quanto dimostrato ai punti 4 e 5 otteniamo:

$$T - \hat{\vartheta}_{MLE} = 0$$

Ovvero:

$$T = \hat{\vartheta}_{MLE}$$

Esempio n. 1

Consideriamo ad esempio:

$$X_1, X_2, \dots, X_n \sim \text{Pois}(\vartheta)$$

In questo caso, lo stimatore di massima verosimiglianza è la media campionaria:

$$\hat{\vartheta}_{ML} = \bar{X}$$

Che sappiamo essere uno stimatore non distorto e consistente per ϑ :

$$E(\bar{X}) = \vartheta \qquad \text{VAR}(\bar{X}) = \frac{\sigma^2}{n} = \frac{\vartheta^2}{n}$$

Inoltre:

$$\begin{aligned} I(\vartheta) &= E \left[\left(\frac{\partial}{\partial \vartheta} \log f(X_1, \vartheta) \right)^2 \right] = E \left[\left(\frac{\partial}{\partial \vartheta} \log \frac{e^{-\vartheta} \vartheta^{X_1}}{X_1!} \right)^2 \right] = E \left[\left[\frac{\partial}{\partial \vartheta} [-\vartheta + X_1 \log \vartheta - \log X_1!] \right]^2 \right] = \\ &= E \left[\left(\frac{X_1}{\vartheta} - 1 \right)^2 \right] = E \left[\left(\frac{X_1}{\vartheta} \right)^2 - \frac{2X_1}{\vartheta} + 1 \right] = \frac{1}{\vartheta^2} E[X_1^2] - 2 + 1 = \frac{1}{\vartheta^2} [\text{VAR}[X_1] - E^2[X_1]] + 1 = \\ &= \frac{1}{\vartheta^2} [\text{VAR}[X_1] - E^2[X_1]] + 1 = \frac{1}{\vartheta^2} - 1 + 1 = \frac{1}{\vartheta^2} \end{aligned}$$

E quindi:

$$l_{FCR} = \frac{k'(\vartheta)}{n I(\vartheta)} = \frac{1}{n \frac{1}{\vartheta^2}} = \frac{\vartheta}{n}$$

Perciò:

$$\text{VAR}(\bar{X}) = l_{FCR} \rightarrow \bar{X} \text{ efficiente}$$

Esempio n. 2

Con riferimento al campione utilizzato nel precedente esempio, consideriamo ora:

$$k_2 = 1 - e^{-\vartheta}$$

Allora, avremo:

$$\hat{k}_{2ML} = 1 - e^{-\bar{X}}$$

La sua media è:

$$E[\hat{k}_{2ML}] = E[1 - e^{-\bar{X}}] = E[1] - E[e^{-\bar{X}}] = 1 - E[e^{-\bar{X}}] = 1 - E\left[e^{-\frac{1}{n}\sum_{i=1}^n X_i}\right]$$

Ricordando che la funzione generatrice dei momenti è definita come:

$$M_X(t) = E[e^{tX}]$$

Otteniamo:

$$E[\hat{k}_{2ML}] = 1 - E\left[e^{-\frac{1}{n}\sum_{i=1}^n X_i}\right] = 1 - M_{\sum_{i=1}^n X_i}\left(-\frac{1}{n}\right)$$

Siccome le variabili X_i sono tutte Poisson con media ϑ , la loro somma sarà una Poisson con media $n\vartheta$:

$$\sum_{i=1}^n X_i \sim Pois(n\vartheta)$$

Perciò la sua funzione generatrice dei momenti è:

$$M_{\sum_{i=1}^n X_i}(t) = [e^{\vartheta(e^t-1)}]^n = e^{n\vartheta(e^t-1)}$$

Perciò:

$$E[\hat{k}_{2ML}] = 1 - e^{n\vartheta\left(e^{-\frac{1}{n}}-1\right)}$$

Abbiamo così ricavato che questo stimatore è distorto. Di conseguenza, lo stimatore non potrà essere efficiente; siccome non è efficiente lo stimatore di verosimiglianza, possiamo concludere che la caratteristica in analisi non ammette alcuno stimatore efficiente.

Proprietà n. 2

Enunciato

Sia X_1, \dots, X_n, \dots una successione di variabili aleatorie indipendenti e identicamente distribuite con comune funzione di densità $f(x, \vartheta), \vartheta \in \Theta$. Sia inoltre $\{T_n\}_n$ la successione degli stimatori di massima verosimiglianza di $k(\vartheta)$. Se le seguenti condizioni sono vere:

1. $f(x, \vartheta)$ soddisfa le condizioni di regolarità imposte come ipotesi della disuguaglianza di FCR.
2. $f(x, \vartheta)$ è derivabile tre volte rispetto a ϑ e le derivate prima, seconda e terza sono continue e limitate.

Allora:

o. Per ogni n , lo stimatore di massima verosimiglianza $\{T_n\}_n$ di $k(\vartheta)$ esiste ed è unico.

1. La successione $\{T_n\}_n$ è asintoticamente non distorta per $k(\vartheta)$. Quindi:

$$\lim_{n \rightarrow +\infty} E(T_n) = k$$

2. La successione $\{T_n\}_n$ è consistente in media quadratica per $k(\vartheta)$:

$$\lim_{n \rightarrow +\infty} VAR(T_n) = 0$$

3. La successione $\{T_n\}_n$ è asintoticamente gaussiana con media asintotica $k(\vartheta)$ e varianza asintotica

$$l_{FCR} = \frac{[k'(\vartheta)]^2}{n \cdot I(\vartheta)}, \text{ cioè:}$$

$$\lim_{n \rightarrow +\infty} P\left(\frac{T_n - k(\vartheta)}{\sqrt{\frac{[k'(\vartheta)]^2}{n \cdot I(\vartheta)}}} \leq z\right) = \Phi(z), \quad \forall z \in \mathbb{R}$$

Applicazioni pratiche

1. Nelle applicazioni pratiche, il punto numero 3 che abbiamo appena indicato ha una particolare importanza. Infatti, l'asintotica normalità implica che, per n "abbastanza grande", si possa considerare:

$$T_n \sim \mathcal{N}\left(k(\vartheta), \frac{[k'(\vartheta)]^2}{n \cdot I(\vartheta)}\right)$$

Dove la varianza tende a zero per $n \rightarrow +\infty$. Inoltre, è evidente dalla proposizione precedente che, siccome lo stimatore tende ad avere media $k(\vartheta)$ e varianza $\frac{[k'(\vartheta)]^2}{n \cdot I(\vartheta)}$, gli stimatori di massima verosimiglianza sono anche asintoticamente efficienti.

2. Una particolare applicazione che se ne ricava è quella relativa all'individuazione degli intervalli di confidenza. Supponiamo ad esempio di voler individuare un intervallo di confidenza per k con livello di confidenza γ : come noto, questo significa individuare un valore di q per il quale (nel caso bilatero):

$$P\left(-q < \frac{\hat{k} - k}{\sqrt{\frac{[k'(\vartheta)]^2}{n \cdot I(\vartheta)}}} < q\right) \approx \gamma$$

Ma, in virtù dell'asintotica normalità, ciò sarà verificato se e solo se:

$$q = \frac{z_{1+\gamma/2}}{2}$$

Si noti bene però che tutto questo vale solo nel caso di grandi campioni.

Esempio di applicazione

Consideriamo ad esempio il campione:

$$X_1, X_2, \dots, X_n \text{ i.i.d. } \sim \text{Pois}(\vartheta)$$

Supponiamo inoltre di voler trovare un intervallo di confidenza γ per la caratteristica:

$$k = \vartheta$$

Sappiamo già che lo stimatore di massima verosimiglianza è la media campionaria e conosciamo il limite di FCR:

$$\hat{k}_{ML} = \bar{X} \qquad l_{FCR} = \frac{\vartheta}{n}$$

E che la media campionaria è efficiente, perciò la sua varianza è data proprio da l_{FCR} . Allora:

$$P\left(-q < \frac{\bar{X} - \vartheta}{\sqrt{\frac{\vartheta}{n}}} < q\right) \approx \gamma$$

Quindi, per l'asintotica normalità:

$$P\left(-z_{\frac{1+\gamma}{2}} < \frac{\bar{X} - \vartheta}{\sqrt{\frac{\vartheta}{n}}} < z_{\frac{1+\gamma}{2}}\right) \approx \gamma \quad \rightarrow \quad P\left(\bar{X} - z_{\frac{1+\gamma}{2}} \sqrt{\frac{\vartheta}{n}} < \vartheta < \bar{X} + z_{\frac{1+\gamma}{2}} \sqrt{\frac{\vartheta}{n}}\right) \approx \gamma$$

Si noti però che:

$$\frac{[k'(\vartheta)]^2}{n \cdot I(\vartheta)} = \frac{\vartheta}{n}$$

È un'espressione nella quale compare ancora il valore ϑ . Di conseguenza, gli estremi dell'intervallo dipendono ancora dallo stesso ϑ . Per risolvere questo problema, si calcola il limite di FCR semplicemente sostituendo nella sua espressione il valore del parametro stimato, e si ottiene così:

$$P\left(\bar{X} - z_{\frac{1+\gamma}{2}} \sqrt{\frac{\bar{X}}{n}} < \vartheta < \bar{X} + z_{\frac{1+\gamma}{2}} \sqrt{\frac{\bar{X}}{n}}\right) \approx \gamma$$

Il metodo della quantità pivotale

Introduzione

Gli intervalli di confidenza

Sia dato un campione:

$$X_1, X_2, \dots, X_n \text{ i. i. d. } \sim f(x, \vartheta)$$

E sia data una caratteristica da stimare:

$$k = k(\vartheta)$$

Si supponga inoltre di disporre di una realizzazione x_1, x_2, \dots, x_n del campione dato. Come noto, un intervallo di confidenza bilatero per k è un intervallo in cui cade k con una certa probabilità (detta confidenza e indicata con γ), i cui estremi sono dipendenti solo dalla realizzazione del campione dato:

$$t_1(x_1, x_2, \dots, x_n) < k < t_2(x_1, x_2, \dots, x_n)$$

In altri termini:

$$P(T_1(X_1, X_2, \dots, X_n) < k < T_2(X_1, X_2, \dots, X_n)) \geq \gamma \quad \forall \vartheta$$

Dove T_1 e T_2 sono due opportune statistiche.

Finora, nel calcolo degli intervalli di confidenza abbiamo sempre sostituito la disuguaglianza debole presente nella precedente relazione con un simbolo di uguaglianza. Se le osservazioni dovessero essere discrete, sarebbe indispensabile utilizzare una disuguaglianza; tuttavia, nelle applicazioni che analizzeremo, ci troveremo sempre (salvo rare eccezioni che analizzeremo di volta in volta) nel caso in cui anche le distribuzioni discrete possono essere approssimate con una normale, perciò possiamo continuare ad utilizzare il simbolo di uguaglianza.

Il metodo

Come abbiamo già visto, l'intervallo può anche essere unilatero, semplicemente sostituendo ad una delle due statistiche un valore illimitato oppure un valore finito (ad esempio, $T_2 = 0$ oppure $T_2 = 1, \dots$).

Il metodo della quantità pivotale Q è un metodo che può essere utilizzato per la costruzione degli intervalli di confidenza. In particolare, la quantità pivotale è una funzione che dipende solamente dai dati: la sua legge (funzione di distribuzione) non dipende dai parametri incogniti:

$$Q(\vartheta, X_1, X_2, \dots, X_n)$$

I passi da compiere

Vediamo ora quali sono i passi che occorre compiere per mettere in pratica il metodo:

1. Si impone che si abbia:

$$P(q_1 < Q(\vartheta, X_1, X_2, \dots, X_n) < q_2) = \gamma$$

E si determinano i valori di q_1 e q_2 che rispettino tale uguaglianza.

2. Si "inverte" la relazione individuata, ovvero si passa da una relazione del tipo:

$$q_1 < Q(\vartheta, X_1, X_2, \dots, X_n) < q_2$$

Ad una relazione del tipo:

$$t_1 < k < t_2$$

In altri termini, si cerca di ottenere dall'intervallo relativo alla quantità pivotale Q , mediante una serie di opportune trasformazioni algebriche, un intervallo relativo alla caratteristica di interesse.

Esempio

Sia dato un campione di n osservazioni con distribuzione esponenziale:

$$X_1, X_2, \dots, X_n \text{ i. i. d. } \sim \mathcal{E}(\vartheta), \quad \vartheta > 0 \text{ incognito}$$

Lo stimatore di massima verosimiglianza per ϑ è la media campionaria (per brevità omettiamo tutto il procedimento che ci consente di verificarlo):

$$\hat{\vartheta}_{ML} = \bar{X}$$

Vogliamo ora trovare un intervallo bilatero per ϑ con livello di confidenza γ . Per prima cosa, costruiamo la quantità pivotale. Per farlo, dobbiamo sfruttare le informazioni relative alla statistica. Nel nostro caso, abbiamo:

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

Sulla base delle proprietà relative alle distribuzioni esponenziali e gamma:

$$X_1, X_2, \dots, X_n \sim \mathcal{E}(\vartheta) = \Gamma(1, \vartheta) \quad \rightarrow \quad \left(\sum_{i=1}^n X_i \right) \sim \Gamma(n, \vartheta) \quad \rightarrow \quad \bar{X} \sim \Gamma\left(n, \frac{\vartheta}{n}\right)$$

Allora, possiamo considerare come quantità pivotale la statistica definita come:

$$\frac{\bar{X}}{\vartheta} \sim \Gamma\left(n, \frac{1}{n}\right)$$

La distribuzione di tale statistica infatti non dipende dal parametro ϑ . Tuttavia, non sono note le tabelle relative alla distribuzione così ottenuta; cerchiamo per tale ragione di ricondurci ad una variabile aleatoria con distribuzione chi-quadro. Per farlo, è sufficiente moltiplicare per $2n$ la quantità pivotale precedente:

$$Q = 2n \frac{\bar{X}}{\vartheta} \sim \Gamma(n, 2) = \chi_{2n}^2$$

Possiamo allora scrivere:

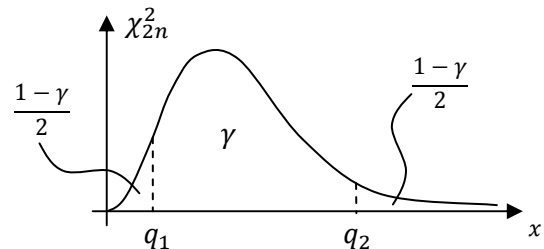
$$\gamma = P(q_1 < Q < q_2)$$

E quindi:

$$q_1 = \chi_{2n}^2 \left(\frac{1-\gamma}{2} \right) \qquad q_2 = \chi_{2n}^2 \left(\frac{1+\gamma}{2} \right)$$

E quindi:

$$\gamma = P\left(\chi_{2n}^2 \left(\frac{1-\gamma}{2} \right) < 2n \frac{\bar{X}}{\vartheta} < \chi_{2n}^2 \left(\frac{1+\gamma}{2} \right) \right)$$



A questo punto, non ci resta che eseguire il secondo passo, ovvero dobbiamo "invertire" l'intervallo:

$$\frac{1}{2n\bar{X}} \chi_{2n}^2 \left(\frac{1-\gamma}{2} \right) < \frac{1}{\vartheta} < \frac{1}{2n\bar{X}} \chi_{2n}^2 \left(\frac{1+\gamma}{2} \right)$$

Ovvero:

$$\frac{2n\bar{X}}{\chi_{2n}^2 \left(\frac{1+\gamma}{2} \right)} < \vartheta < \frac{2n\bar{X}}{\chi_{2n}^2 \left(\frac{1-\gamma}{2} \right)}$$

Si noti che la struttura dell'intervallo così ottenuto è di fatto analoga a quella dell'intervallo per la varianza di una popolazione gaussiana.

La verifica di ipotesi

Introduzione

Introduzione alla verifica d'ipotesi

A questo punto, vogliamo passare ad un argomento profondamente diverso da quelli finora trattati, ma che richiede come prerequisito i concetti precedentemente introdotti: la verifica d'ipotesi.

La verifica di ipotesi è in sintesi un problema di tipo ipotetico: non siamo in questo caso interessati a conoscere esattamente una caratteristica, oppure l'intervallo nel quale tale caratteristica si troverà con una certa probabilità nota a priori; l'obiettivo che vogliamo raggiungere in questo caso è quello di verificare se una certa congettura risulta essere soddisfatta, con un certo grado di probabilità, nonostante la popolazione sia parzialmente incognita.

Esempio

Per comprendere meglio il problema che vogliamo affrontare, partiamo da un esempio. Un'azienda che produce delle cinghie di trasmissione ha brevettato un nuovo metodo di produzione che, sulla base di quanto dichiarato dai laboratori che lo hanno ideato, dovrebbe aumentare la vita media delle cinghie stesse, portandola da 50.000 km a 56.000 km. Prima di avviare la produzione secondo il nuovo brevetto, l'azienda vuole però verificare se effettivamente si ha il miglioramento sperato oppure no, in quanto tale modifica comporta chiaramente dei costi di transizione.

La situazione appena illustrata è un caso tipico nel quale si rende necessario eseguire un test d'ipotesi. A tale scopo, l'azienda testa il nuovo metodo di produzione, realizzando le cinghie che vengono montate su un certo numero di automobili (ipotizziamo ad esempio che siano 35). Si ottiene così un certo campione di misurazioni della vita media delle cinghie. L'azienda deciderà poi se avviare oppure no la nuova modalità di produzione: in particolare, sulla base di opportune politiche, l'azienda considererà veritiere le dichiarazioni dei laboratori che hanno brevettato il nuovo metodo di produzione se e solo se la durata media rilevata sul campione di 35 automobili risulterà essere almeno pari a 57.000 km.

I concetti fondamentali della verifica d'ipotesi

Verifica di ipotesi

Eseguire la verifica d'ipotesi significa verificare una certa congettura, relativa ad un parametro o ad una caratteristica di una popolazione, o relativa all'intera distribuzione della popolazione.

Le ipotesi

Il primo elemento della verifica d'ipotesi è rappresentato dalle ipotesi stesse. Si deve infatti definire una certa "ipotesi statistica", ovvero un'affermazione sul parametro, che traduca di fatto la congettura iniziale. Nell'esempio, la congettura è "la preoccupazione dell'azienda è fondata", ovvero "la durata media delle cinghie non è superiore a 56.000 km", che si traduce in $\mu \leq 56.000 \text{ km}$.

Tuttavia, per essere più precisi, il test d'ipotesi richiede l'esistenza di due diverse ipotesi:

1. L'ipotesi nulla

L'ipotesi nulla può essere definita in modo informale come quell'ipotesi che è vera fino a prova contraria, e che si vorrebbe fosse falsa. Viene indicata con H_0 .

2. L'ipotesi alternativa

L'ipotesi alternativa è l'ipotesi che si vuole verificare. Viene indicata con H_1 .

Nell'esempio:

$$H_0: \mu \leq 56.000 \text{ km}$$

$$H_1: \mu > 56.000 \text{ km}$$

Si noti che l'ipotesi nulla e l'ipotesi alternativa non devono necessariamente essere complementari. Ad esempio, sarebbe stato accettabile anche formulare le ipotesi nel modo:

$$H_0: \mu = 56.000 \text{ km}$$

$$H_1: \mu > 56.000 \text{ km}$$

Oppure:

$$H_0: \mu < 56.000 \text{ km}$$

$$H_1: \mu > 57.000 \text{ km}$$

Si noti inoltre che l'ipotesi nulla e l'ipotesi alternativa non sono simmetriche: non è infatti possibile scambiarle, perché si otterrebbe altrimenti una soluzione diversa del problema.

Le ipotesi possono inoltre essere:

- Ipotesi semplici

L'ipotesi si dice semplice se specifica completamente la distribuzione di probabilità incognita. Ad esempio, un'ipotesi del tipo:

$$\mu = 56.000 \text{ km}$$

È un'ipotesi semplice

- Ipotesi composte

L'ipotesi si dice composta se non è semplice, ovvero se non specifica completamente la distribuzione di probabilità incognita. Ad esempio, ipotesi del tipo:

$$\mu \neq 56.000 \text{ km}$$

$$\mu > 56.000 \text{ km}$$

Sono ipotesi composte.

Tipicamente, le ipotesi vengono formulate allo scopo di scegliere tra una nuova metodologia e una vecchia metodologia, e l'ipotesi nulla è del tipo "la nuova teoria non funziona meglio della vecchia" e l'ipotesi alternativa è del tipo "la nuova teoria funziona meglio della vecchia".

Una statistica si dice non parametrica (o *distribution free*) nel caso in cui non si conosca la distribuzione di probabilità sottostante.

I dati

Un altro degli ingredienti fondamentali della verifica d'ipotesi sono i dati. Per eseguire la verifica d'ipotesi è infatti necessario raccogliere un certo campione di dati:

$$X_1, X_2, \dots, X_n \text{ i. i. d.}$$

Del quali si otterrà una certa realizzazione x_1, x_2, \dots, x_n .

La regola di decisione: regione critica

Il terzo elemento fondamentale del test d'ipotesi è la regola di decisione. Infatti, data l'incertezza dei dati (si considera inevitabilmente solo un campione di dimensione finita), non è possibile ottenere un risultato assolutamente certo, così come accade in genere nelle procedure di tipo induttivo.

Bisogna perciò stabilire una regola di decisione, ovvero un criterio in base al quale stabilire quale decisione prendere. Riprendendo l'esempio iniziale, una possibile regola di decisione (quella indicata nell'esempio di partenza) è:

$$\bar{X} \geq 57.000 \text{ km}$$

Tale scelta è sensata, perché considerare falsa l'ipotesi nulla nel caso $\bar{X} \geq 56.000 \text{ km}$ significherebbe prendere una decisione poco affidabile, in quanto la media reale potrebbe essere diversa da quella campionaria.

In generale, la regola di decisione è un insieme, che chiamiamo *regione critica* e indichiamo con \mathcal{G} , e che definiamo come l'insieme di tutti i risultati sperimentali per i quali rifiutiamo l'ipotesi nulla:

$$\mathcal{G} = \{(x_1, \dots, x_n) \text{ per le quali rifiuto } H_0\}$$

Statistica test

Chiamiamo statistica test quella statistica sulla base della quale prendiamo le decisioni relative al test d'ipotesi.

Errori

Una volta presa la decisione, non è detto che quest'ultima sia corretta. È possibile perciò calcolare la probabilità di errore. In particolare, i possibili esiti del test sono indicati in tabella:

	Decisione	
Realtà \	Non rifiuto H_0	Rifiuto H_0
H_0 è vera	Decisione giusta	Errore di I tipo
H_0 è falsa	Errore di II tipo	Decisione giusta

Gli errori possibili sono perciò di due diverse tipologie:

- **Errore di I tipo**
Si verifica quando l'ipotesi nulla è vera, ma H_0 viene rifiutata. In sostanza quindi i dati appartengono alla regione critica, ma l'ipotesi nulla è vera.
- **Errore di II tipo**
Si verifica quando l'ipotesi nulla è falsa, ma H_0 viene accettata. In sostanza quindi i dati non appartengono alla regione critica, ma l'ipotesi nulla è falsa.

Probabilità di errore di I tipo

Sulla base di quanto abbiamo detto la probabilità di errore di primo tipo, che viene indicata con α , è la probabilità che, sapendo che H_0 è vera, i dati appartengano alla regione critica; in simboli:

$$P_{H_0}((x_1, \dots, x_n) \in \mathcal{G})$$

Quindi, dato il campione:

$$X_1, X_2, \dots, X_n \sim f(x, \vartheta)$$

E date le ipotesi:

$$H_0: \vartheta \in \Theta_0$$

$$H_1: \vartheta \in \Theta_1$$

Abbiamo:

$$\alpha(\vartheta) = P_{\vartheta \in \Theta_0}(\mathcal{G})$$

Si osserva perciò che la probabilità di errore dipende anche dal valore vero (e incognito) del parametro ϑ relativamente al quale l'ipotesi viene formulata.

Probabilità di errore di II tipo

Analogamente, la probabilità di errore di secondo tipo viene indicata con β ed è la probabilità che, sapendo che H_0 è falsa, i dati non appartengano alla regione critica. In simboli:

$$\beta(\vartheta) = P_{\vartheta \in \Theta_1}(\mathcal{G}^c)$$

Si noti però che NON vale l'uguaglianza:

$$\beta(\vartheta) = 1 - \alpha(\vartheta)$$

Perché le due probabilità vengono calcolate sulla base di valori di ϑ appartenenti ad insiemi diversi.

Potenza del test

Chiamiamo funzione di potenza del test la funzione $\pi(\vartheta)$, che rappresenta la probabilità di rifiutare l'ipotesi nulla nel caso in cui sia falsa:

$$\pi(\vartheta) = P_{\vartheta \in \Theta_1}(\mathcal{G})$$

In questo caso vale allora la relazione:

$$\pi(\vartheta) = 1 - \beta(\vartheta), \quad \vartheta \in \Theta_1$$

Possiamo allora affermare che la potenza del test rappresenta la probabilità di prendere la corretta decisione di falsificare H_0 .

Osservazione

Se la regione critica \mathcal{G} viene scelta in modo tale da minimizzare l'errore di I tipo, si avrà inevitabilmente un aumento dell'errore di secondo tipo, e viceversa. Si tratta quindi del tipico "problema della coperta troppo corta".

Di conseguenza, per confrontare tra loro due diversi test, si utilizza in genere la potenza, preferendo i test con potenza superiore, perché si considera più importante minimizzare la probabilità di errore di primo tipo.

Significatività

La significatività (*level of significance*, o *ampiezza* o *dimensione*) della regione critica è l'estremo superiore della probabilità di errore di I tipo:

$$\alpha = \sup_{\vartheta \in \Theta_0} \alpha(\vartheta)$$

Test d'ipotesi non distorto

Un test d'ipotesi si dice non distorto se la sua funzione di potenza $\pi(\vartheta)$ è sempre maggiore della sua significatività α , per ogni $\vartheta \in \Theta_0$.

Il p-value

Definizione

Nella verifica d'ipotesi, il *p-value* è una statistica che rappresenta il più piccolo valore del livello di significatività α che porta a rifiutare l'ipotesi nulla H_0 sulla base dei dati raccolti. Il p-value è un valore appartenente all'intervallo $[0,1]$.

Calcolo pratico del p-value

Nella pratica, per valutare il p-value si procede in questo modo:

1. Si calcola il valore della statistica test T con i dati raccolti x_1, x_2, \dots, x_n :

$$t = T(x_1, x_2, \dots, x_n)$$

2. Si calcola:

$$pvalue = P_{H_0}(T \in \mathcal{G})$$

Distribuzione del p-value

Si dimostra che, se i dati sono continui (cioè la densità di X è assolutamente continua), allora, la distribuzione del p-value nel caso in cui l'ipotesi nulla sia vera è:

$$F_{\bar{P}_{H_0}}(p) = \begin{cases} 0 & \text{se } p \leq 0 \\ p & \text{se } 0 < p < 1 \\ 1 & \text{se } p \geq 1 \end{cases} \quad f_{\bar{P}_{H_0}} = \begin{cases} 1 & \text{se } 0 < p < 1 \\ 0 & \text{altrimenti} \end{cases}$$

Ovvero:

$$\bar{P}_{H_0} \sim \mathcal{U}(0,1)$$

Uso del p-value

Una volta che il p-value è stato calcolato, può essere confrontato con il livello di confidenza α prefissato:

1. Se $pvalue > \alpha$, l'ipotesi nulla viene accettata;
2. Se $pvalue \leq \alpha$, l'ipotesi nulla viene rifiutata.

In linea generale diciamo che:

- Se il p-value è minore dell'1%, si ha *forte evidenza contro* H_0 .
- Se il p-value è compreso tra il 2,5% e il 5%, si dice che si ha *debole evidenza contro* H_0 e la decisione presa dipende quindi dal livello di confidenza.
- Se il p-value è maggiore o uguale al 10%, allora si dice che *dai dati non emerge contrarietà ad* H_0 .

In ogni caso, la decisione viene presa sulla base del confronto col livello di confidenza.

Esempio

Consideriamo il caso seguente:

$$H_0: \vartheta \in \Theta_0$$

$$H_1: \vartheta \in \Theta_1$$

E ipotizziamo di avere raccolto i dati:

$$x_1, x_2, \dots, x_n$$

Ipotizziamo inoltre che il test in analisi abbia una regione critica descritta da una frase del tipo: "rifiuto H_0 se $T \geq k$ ", dove T è la statistica test. Potremo allora calcolare il p-value andando ad individuare il valore della statistica test in corrispondenza dei dati raccolti:

$$t = T(x_1, x_2, \dots, x_n)$$

Dopodiché avremo:

$$p - value = P_{H_0}(T \geq t)$$

Lemma di Neyman-Pearson

Introduzione

Ricapitolando quanto abbiamo finora visto, possiamo affermare che un buon test deve cercare di minimizzare sia la probabilità di errore di I tipo, sia la probabilità di errore di II tipo, ma ciò è di fatto impossibile, perciò è necessario individuare un compromesso soddisfacente.

Allora, per trovare la soluzione ottimale, si prevede che si fissi un limite massimo alla probabilità di errore di I tipo, e che si scelga tra tutti i test con probabilità di errore di primo tipo non superiore a tale limite, quel test che abbia la più bassa probabilità di errore di secondo tipo. Il limite alla probabilità di errore di primo tipo è dato dall'ampiezza del test.

Trovare un test che soddisfi queste ipotesi significa individuare il test che sia **uniformemente più potente** tra i test di ampiezza α , dove l'espressione **uniformemente più potente** sta ad indicare che ciò vale per ogni $\vartheta \in \Theta_1$ (si noti che l'ipotesi potrebbe essere sul modello e non sul parametro, ma tutto ciò che diremo vale in ogni caso).

Il lemma di Neyman-Pearson

Quando si usa

Il lemma di Neyman-Pearson ci consente di risolvere il problema appena illustrato in un caso particolare, ovvero nel caso in cui, dato il campione casuale:

$$X_1, X_2, \dots, X_n \sim f(x, \vartheta)$$

E data la funzione di verosimiglianza $L_{\vartheta}(X_1, X_2, \dots, X_n)$, si abbia un test con le seguenti ipotesi:

$$H_0: \vartheta = \vartheta_0$$

$$H_1: \vartheta = \vartheta_1$$

Ovvero, sia l'ipotesi nulla che l'ipotesi alternativa sono semplici. Si noti inoltre che in questo caso il termine **uniformemente** risulta essere di fatto inutile, in quanto Θ_1 è un insieme contenente solo l'elemento ϑ_1 .

Enunciato

Sotto le ipotesi appena descritte, il lemma di Neyman-Pearson afferma che il test più potente per verificare le ipotesi date è il test avente come regione critica:

$$\mathcal{G} = \left\{ (x_1, x_2, \dots, x_n) : \frac{L_{\vartheta_0}(x_1, \dots, x_n)}{L_{\vartheta_1}(x_1, \dots, x_n)} \leq \delta \right\}$$

Osservazione

La regione critica così definita impone in sostanza che sia più probabile che i dati siano stati generati da una distribuzione con il parametro ϑ_1 anziché con il parametro ϑ_0 : se ciò è vero, infatti, la funzione di verosimiglianza valutata in ϑ_1 sarà superiore di quella valutata in ϑ_0 , e quindi il rapporto sarà piccolo, perciò i campioni probabilmente apparterranno alla regione critica, e l'ipotesi nulla verrà rifiutata.

Dimostrazione

1. Partiamo da un'osservazione. Dato un sottoinsieme A di \mathcal{G} :

$$A \subseteq \mathcal{G}$$

Se \mathcal{G} è definita come sopra indicato, in ogni punto di A (proprio perché tale punto appartiene anche a \mathcal{G}) abbiamo:

$$L_{\vartheta_0}(x_1, \dots, x_n) \leq \delta L_{\vartheta_1}(x_1, \dots, x_n)$$

Allora, se siamo nel continuo:

$$P_{\vartheta_0}(A) = P_{\vartheta_0}((X_1, X_2, \dots, X_n) \in A) = \int_A L_{\vartheta_0}(x_1, x_2, \dots, x_n, \vartheta_0) dx_1 \dots dx_n$$

Sfruttando la precedente disuguaglianza:

$$P_{\vartheta_0}(A) = \int_A L_{\vartheta_0}(x_1, x_2, \dots, x_n, \vartheta_0) dx_1 \dots dx_n \leq \delta \int_A L_{\vartheta_1}(x_1, x_2, \dots, x_n, \vartheta_0) dx_1 \dots dx_n = \delta P_{\vartheta_1}(A)$$

Quindi:

$$P_{\vartheta_0}(A) \leq \delta P_{\vartheta_1}(A)$$

In maniera del tutto analoga, possiamo verificare che, considerando $B \subseteq \mathcal{G}^c$, otteniamo:

$$P_{\vartheta_0}(B) \geq \delta P_{\vartheta_1}(B)$$

2. Vogliamo ora considerare una regione critica \mathcal{F} di ampiezza minore o uguale all'ampiezza di \mathcal{G} , ovvero tale che:

$$P_{\vartheta_0}(\mathcal{F}) \leq P_{\vartheta_0}(\mathcal{G})$$

Possiamo considerare che:

$$\mathcal{F} = (\mathcal{F} \cap \mathcal{G}) \cup (\mathcal{F} \cap \mathcal{G}^c) \qquad \mathcal{G} = (\mathcal{F} \cap \mathcal{G}) \cup (\mathcal{F}^c \cap \mathcal{G})$$

Perciò la precedente disuguaglianza sarà:

$$P_{\vartheta_0}(\mathcal{F} \cap \mathcal{G}) + P_{\vartheta_0}(\mathcal{F} \cap \mathcal{G}^c) \leq P_{\vartheta_0}(\mathcal{F} \cap \mathcal{G}) + P_{\vartheta_0}(\mathcal{F}^c \cap \mathcal{G})$$

Da cui:

$$P_{\vartheta_0}(\mathcal{F} \cap \mathcal{G}^c) \leq P_{\vartheta_0}(\mathcal{F}^c \cap \mathcal{G})$$

L'obiettivo è quello di verificare la tesi, secondo la quale la potenza di \mathcal{G} è maggiore o uguale alla potenza di \mathcal{F} :

$$\pi_{\mathcal{F}} \leq \pi_{\mathcal{G}}$$

Cioè:

$$P_{\vartheta_1}(\mathcal{F}) \leq P_{\vartheta_1}(\mathcal{G})$$

Che equivale anche a dire:

$$P_{\vartheta_1}(\mathcal{F} \cap \mathcal{G}^c) \leq P_{\vartheta_1}(\mathcal{F}^c \cap \mathcal{G})$$

Siccome:

$$\mathcal{F} \cap \mathcal{G}^c \subseteq \mathcal{G}^c$$

Possiamo dire che:

$$P_{\vartheta_0}(\mathcal{F} \cap \mathcal{G}^c) \geq \delta P_{\vartheta_1}(\mathcal{F} \cap \mathcal{G}^c) \rightarrow P_{\vartheta_1}(\mathcal{F} \cap \mathcal{G}^c) \leq \frac{1}{\delta} P_{\vartheta_0}(\mathcal{F} \cap \mathcal{G}^c)$$

Siccome abbiamo visto che dall'ipotesi consegue:

$$P_{\vartheta_0}(\mathcal{F} \cap \mathcal{G}^c) \leq P_{\vartheta_0}(\mathcal{F}^c \cap \mathcal{G})$$

Possiamo scrivere anche:

$$P_{\vartheta_1}(\mathcal{F} \cap \mathcal{G}^c) \leq \frac{1}{\delta} P_{\vartheta_0}(\mathcal{F}^c \cap \mathcal{G})$$

Ma, dal momento che $\mathcal{F}^c \cap \mathcal{G} \subseteq \mathcal{G}$:

$$P_{\vartheta_0}(\mathcal{F}^c \cap \mathcal{G}) \leq \delta P_{\vartheta_1}(\mathcal{F}^c \cap \mathcal{G})$$

Da cui otteniamo, sostituendo banalmente nella disequazione precedente:

$$P_{\vartheta_1}(\mathcal{F} \cap \mathcal{G}^c) \leq \frac{1}{\delta} \delta P_{\vartheta_1}(\mathcal{F}^c \cap \mathcal{G})$$

Ovvero:

$$P_{\vartheta_1}(\mathcal{F} \cap \mathcal{G}^c) \leq P_{\vartheta_1}(\mathcal{F}^c \cap \mathcal{G})$$

Che è proprio la tesi che volevamo dimostrare.

Verifica d'ipotesi su popolazione gaussiana: gli Z-test e i T-test

Sia dato il campione:

$$X_1, X_2, \dots, X_n \text{ i. i. d. } \sim \mathcal{N}(\mu, \sigma^2)$$

Allora, la funzione di verosimiglianza sarà data da:

$$L_{\mu, \sigma^2}(x_1, \dots, x_n) = \prod_{i=1}^n f(x_i, \mu, \sigma^2) = \prod_{i=1}^n \left[\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} \right] = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2} =$$

Già in precedenza abbiamo dimostrato che:

$$\sum_{i=1}^n (x_i - \mu)^2 = \sum_{i=1}^n (x_i - \bar{X})^2 + n(\bar{X} - \mu)^2 = (n-1)S^2 + n(\bar{X} - \mu)^2$$

Da cui:

$$L_{\mu, \sigma^2}(x_1, \dots, x_n) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{-\frac{1}{2\sigma^2} [(n-1)S^2 + n(\bar{X} - \mu)^2]}$$

Questo significa che ci serve conoscere solo la media campionaria e la varianza campionaria per essere in grado di calcolare la funzione di verosimiglianza.

Verifica di ipotesi sulla media con varianza nota: ipotesi nulla e alternativa semplici

Supponiamo di voler eseguire una verifica di ipotesi relativa alla media, conoscendo con esattezza la varianza della distribuzione. Ipotizziamo anche di avere un test del tipo "ipotesi semplice contro ipotesi semplice", ovvero:

$$H_0: \mu = \mu_0$$

$$H_1: \mu = \mu_1$$

Allora, le ipotesi del lemma di Neyman-Pearson sono verificate, perciò il test più potente sarà quello la cui regione critica è data da:

$$\begin{aligned} \frac{L_{\mu_0, \sigma^2}(x_1, \dots, x_n)}{L_{\mu_1, \sigma^2}(x_1, \dots, x_n)} &= \frac{\left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{-\frac{1}{2\sigma^2} [(n-1)S^2 + n(\bar{X} - \mu_0)^2]}}{\left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{-\frac{1}{2\sigma^2} [(n-1)S^2 + n(\bar{X} - \mu_1)^2]}} = e^{-\frac{1}{2\sigma^2} [n(\bar{X} - \mu_0)^2 - n(\bar{X} - \mu_1)^2]} = \\ &= e^{-\frac{1}{2\sigma^2} [n\mu_0^2 - n\mu_1^2]} \cdot e^{-\frac{1}{2\sigma^2} (-2n\bar{X}\mu_0 + 2n\bar{X}\mu_1)} = e^{-\frac{1}{2\sigma^2} [n\mu_0^2 - n\mu_1^2]} \cdot e^{\frac{n\bar{X}}{\sigma^2} (\mu_0 - \mu_1)} \end{aligned}$$

Dove il primo fattore è di fatto un valore costante, che non dipende dai dati:

$$c = e^{-\frac{1}{2\sigma^2} [n\mu_0^2 - n\mu_1^2]}$$

Se ipotizziamo ad esempio di avere $\mu_0 > \mu_1$, allora la funzione così ottenuta è una funzione monotona crescente in \bar{X} . Allora vale:

$$\frac{L_{\mu_0, \sigma^2}(x_1, \dots, x_n)}{L_{\mu_1, \sigma^2}(x_1, \dots, x_n)} \leq \delta$$

Se e solo se si ha:

$$\bar{X} \leq k$$

Dove k è un opportuno valore ricavato da δ e che dipende da α :

$$\alpha = \alpha(\mu_0) = P_{\mu_0}(\text{rifiuto } H_0) = P_{\mu_0}(\bar{X} \leq k)$$

Quindi:

$$\alpha = \Phi\left(\frac{k - \mu_0}{\frac{\sigma}{\sqrt{n}}}\right) \rightarrow \frac{k - \mu_0}{\frac{\sigma}{\sqrt{n}}} = z_\alpha$$

Siccome però α è una probabilità di errore, il suo valore sarà in genere "piccolo" (comunque minore di un mezzo), perciò è più pratico per ricercare i valori sulle tavole riscrivere la relazione precedente come:

$$\frac{k - \mu_0}{\frac{\sigma}{\sqrt{n}}} = -z_{1-\alpha}$$

Dalla quale ricaviamo:

$$k = \mu_0 - \frac{\sigma}{\sqrt{n}} z_{1-\alpha}$$

Abbiamo così concluso che il test più potente di livello α del tipo "ipotesi semplice contro ipotesi semplice", entrambe relative alla media di una gaussiana, nel caso $\mu_1 < \mu_0$, è il test che ha regione critica:

$$\mathcal{G} = \left\{ (x_1, x_2, \dots, x_n) : \bar{x} \leq \mu_0 - \frac{\sigma}{\sqrt{n}} z_{1-\alpha} \right\}$$

Con un procedimento del tutto analogo si ricava invece che, nel caso $\mu_1 > \mu_0$, si ha invece:

$$\mathcal{G} = \left\{ (x_1, x_2, \dots, x_n) : \bar{x} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha} \right\}$$

Si può notare che la regione critica non dipende esplicitamente da μ_1 , ma solo da μ_0 (salvo il fatto che la regione critica cambia se μ_1 diventa inferiore o superiore di μ_0). Di conseguenza, la regione critica:

$$\mathcal{G} = \left\{ (x_1, x_2, \dots, x_n) : \bar{x} \leq \mu_0 - \frac{\sigma}{\sqrt{n}} z_{1-\alpha} \right\}$$

Identifica il test ottimo per ogni test d'ipotesi del tipo:

$$H_0: \mu = \mu_0$$

$$H_1: \mu = \mu_1$$

Con $\mu_0 > \mu_1$; analogamente, la regione critica:

$$\mathcal{G} = \left\{ (x_1, x_2, \dots, x_n) : \bar{x} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha} \right\}$$

Identifica il test ottimo per ogni ipotesi nulla dello stesso tipo appena illustrato, ma con $\mu_0 < \mu_1$. I test di questo tipo, siccome utilizzano i quantili della normale standard, vengono detti anche **z-test**.

Nota

Si usano gli Z-test per la media anche nel caso in cui si abbiano grandi campioni (in genere $n \geq 30$) ma non di tipo gaussiano, perché si sfrutta l'ipotesi di asintotica gaussianità della media campionaria. In questo caso ovviamente si tratterà però non di un test esatto, bensì di un test *asintotico*.

Verifica di ipotesi sulla media con varianza nota: ipotesi nulla e alternativa composte

- Consideriamo ora il caso:

$$H_0: \mu = \mu_0 \qquad H_1: \mu < \mu_0$$

In questo caso il test ottimo è ancora quello con regione critica:

$$\mathcal{G} = \left\{ (x_1, x_2, \dots, x_n) : \bar{x} \leq \mu_0 - \frac{\sigma}{\sqrt{n}} z_{1-\alpha} \right\}$$

- Analogamente, se consideriamo ora il caso:

$$H_0: \mu = \mu_0 \qquad H_1: \mu > \mu_0$$

In questo caso il test ottimo è ancora quello con regione critica:

$$\mathcal{G} = \left\{ (x_1, x_2, \dots, x_n) : \bar{x} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha} \right\}$$

- Se consideriamo invece il caso:

$$H_0: \mu \geq \mu_0 \qquad H_1: \mu < \mu_0$$

Si può dimostrare che in questa situazione non esiste un test più potente in assoluto. Allora, si cerca il test ottimo limitando la ricerca ai soli test non distorti. In tale situazione, si verifica che la regione critica che identifica il test ottimo è ancora una volta:

$$\mathcal{G} = \left\{ (x_1, x_2, \dots, x_n) : \bar{x} \leq \mu_0 - \frac{\sigma}{\sqrt{n}} z_{1-\alpha} \right\}$$

Possiamo quindi affermare che tale regione critica identifica il test più potente per il problema con le ipotesi nulla e alternativa sopra riportate, limitatamente ai test di ampiezza α e non distorti.

- In maniera analoga, considerando:

$$H_0: \mu \leq \mu_0 \qquad H_1: \mu > \mu_0$$

Si può dimostrare che non esiste un test più potente in assoluto, perciò si limita il campo di ricerca ai soli test non distorti e si verifica che la regione critica che identifica il test ottimo è:

$$\mathcal{G} = \left\{ (x_1, x_2, \dots, x_n) : \bar{x} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha} \right\}$$

Possiamo quindi affermare che tale regione critica identifica il test più potente per il problema con le ipotesi nulla e alternativa sopra riportate, limitatamente ai test di ampiezza α e non distorti.

Anche i test di queste tipologie sono detti *z-test*, perché, come i precedenti, si basano sull'uso della distribuzione normale standard e dei suoi quantili.

Test del rapporto di verosimiglianza

Tutti i risultati finora ottenuti sono legati direttamente o indirettamente al lemma di Neyman-Pearson. Tuttavia, esistono anche delle situazioni nelle quali tale lemma non è applicabile.

Consideriamo ad esempio il problema (relativo sempre ad una popolazione gaussiana) con ipotesi:

$$H_0: \vartheta \in \Theta_0 \qquad H_1: \vartheta \in \Theta_1$$

Possiamo ora considerare:

$$\Lambda = \frac{\sup_{\vartheta \in \Theta_0} L_{\vartheta}(x_1, \dots, x_n)}{\sup_{\vartheta \in \Theta_1} L_{\vartheta}(x_1, \dots, x_n)}$$

E utilizzare una regione critica tale per cui si rifiuti H_0 se $\Lambda \leq \delta$:

$$\mathcal{G} = \{ (x_1, x_2, \dots, x_n) : \Lambda \geq \delta \}$$

Dove δ viene fissato in modo tale che:

$$\alpha \geq \sup_{\vartheta \in \Theta_0} P_{\vartheta}(\Lambda \leq \delta)$$

Questo test però potrebbe non essere il test più potente per la risoluzione del problema di verifica di ipotesi dato.

Un caso particolare del problema in analisi è quello in cui si ha:

$$H_0: \vartheta = \vartheta_0$$

$$H_1: \vartheta \neq \vartheta_0$$

In tale situazione:

$$\Lambda = \frac{L_{\vartheta_0}(x_1, \dots, x_n)}{\sup_{\vartheta \in \Theta_1} L_{\vartheta}(x_1, \dots, x_n)} = \frac{L_{\vartheta_0}(x_1, \dots, x_n)}{L_{\hat{\vartheta}_{ML}}(x_1, \dots, x_n)}$$

Nel caso particolare in cui il parametro ϑ sia la media:

$$H_0: \mu = \mu_0$$

$$H_1: \mu \neq \mu_0$$

Otteniamo:

$$\Lambda = \frac{L_{\vartheta}(x_1, \dots, x_n)}{L_{\hat{\vartheta}_{ML}}(x_1, \dots, x_n)} = \frac{\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}[(n-1)S^2 + n(\bar{X} - \mu_0)^2]}}{\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}[(n-1)S^2 + n(\bar{X} - \bar{X})^2]}} = e^{-\frac{n(\bar{X} - \mu_0)^2}{2\sigma^2}}$$

Quindi la regione critica sarà:

$$\mathcal{G} = \left\{ (x_1, x_2, \dots, x_n) : e^{-\frac{n(\bar{X} - \mu_0)^2}{2\sigma^2}} \leq \delta \right\}$$

La disequazione appena scritta equivale a:

$$-\frac{n(\bar{X} - \mu_0)^2}{2\sigma^2} \leq \ln \delta \quad \rightarrow \quad \frac{n(\bar{X} - \mu_0)^2}{\sigma^2} \geq 2 \ln \delta \quad \rightarrow \quad \frac{(\bar{X} - \mu_0)^2}{\frac{\sigma^2}{n}} \geq 2 \ln \delta \quad \rightarrow \quad \left(\frac{\bar{X} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} \right)^2 \geq 2 \ln \delta$$

Da cui ricaviamo infine:

$$\frac{|\bar{X} - \mu_0|}{\sqrt{\frac{\sigma^2}{n}}} \geq \sqrt{2 \ln \delta}$$

Il valore di δ in realtà non ha alcuna importanza, perciò potremmo sostituire il secondo membro della precedente disuguaglianza con un generico q . In sostanza, rifiutiamo H_0 se:

$$\frac{|\bar{X} - \mu_0|}{\sqrt{\frac{\sigma^2}{n}}} \geq q$$

Dove q deve essere tale che:

$$\alpha = \sup_{\mu = \mu_0} P \left(\frac{|\bar{X} - \mu_0|}{\sqrt{\frac{\sigma^2}{n}}} \geq q \right) = P_{\mu_0} \left(\frac{|\bar{X} - \mu_0|}{\sqrt{\frac{\sigma^2}{n}}} \geq q \right)$$

Ma, sotto l'ipotesi H_0 (che per il calcolo di questa probabilità deve essere considerata vera), la variabile aleatoria che appare al primo membro della disequazione è una normale standard, perciò:

$$q = z_{1-\frac{\alpha}{2}}$$

Il caso a varianza incognita

Se ipotizzassimo invece che la varianza fosse incognita, tutti i risultati finora individuati rimangono sostanzialmente invariati, salvo il fatto che anziché utilizzare la varianza effettiva si utilizzerà la varianza campionaria e, anziché utilizzare i quantili della distribuzione normale standard, si utilizzeranno quelli della T-student. Proprio per questa ragione, i test di questo tipo vengono chiamati **T-test**.

Verifica d'ipotesi su popolazione gaussiana: test chi-quadro sulla varianza (metodo degli IC) – caso a media incognita

Il tipo di test

Sia dato il campione gaussiano:

$$X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$$

E si voglia eseguire un test del tipo:

$$H_0: \sigma^2 = \sigma_0^2 \qquad H_1: \sigma^2 \neq \sigma_0^2$$

La media μ può essere nota oppure incognita. Iniziamo considerandola incognita. Si noti che quanto diremo ora vale solo nel caso in cui il campione sia *esattamente gaussiano*, e non nel caso in cui sia *approssimativamente* gaussiano.

Idea di base

L'idea di base del test è quella di partire dall'intervallo di confidenza bilatero della varianza:

$$\frac{s^2(n-1)}{\chi_{n-1}^2\left(\frac{1+\gamma}{2}\right)} < \sigma^2 < \frac{s^2(n-1)}{\chi_{n-1}^2\left(\frac{1-\gamma}{2}\right)}$$

Se σ^2 non appartiene all'intervallo di confidenza, possiamo considerare questo dato come indicatore del fatto che l'ipotesi nulla è falsa.

Si usa allora la regola:

$$\text{rifiuto } H_0: \sigma^2 = \sigma_0^2 \text{ a favore di qualunque altro valore (ovvero } H_1: \sigma^2 \neq \sigma_0^2 \text{ se } \sigma_0^2 \notin IC(\sigma^2))$$

Questa regola equivale a imporre la regione critica:

$$\begin{aligned} \mathcal{G} &= \left\{ (x_1, \dots, x_n): \sigma_0^2 \leq \frac{s^2(n-1)}{\chi_{n-1}^2\left(\frac{1+\gamma}{2}\right)} \text{ oppure } \sigma_0^2 \geq \frac{s^2(n-1)}{\chi_{n-1}^2\left(\frac{1-\gamma}{2}\right)} \right\} = \\ &= \left\{ (x_1, \dots, x_n): \chi_{n-1}^2\left(\frac{1+\gamma}{2}\right) \leq \frac{s^2(n-1)}{\sigma_0^2} \text{ oppure } \chi_{n-1}^2\left(\frac{1-\gamma}{2}\right) \geq \frac{s^2(n-1)}{\sigma_0^2} \right\} \end{aligned}$$

Statistica test

Si può quindi usare come statistica test la statistica:

$$\frac{S^2(n-1)}{\sigma_0^2}$$

Dove si ricorda che S^2 è la varianza campionaria e σ_0^2 il valore che stiamo "testando".

Generalizzazione

Generalizzando, possiamo considerare un test su una certa caratteristica k di una distribuzione di probabilità, dove il test è del tipo:

$$H_0: k = k_0 \qquad H_1: k \neq k_0$$

Si costruisce quindi un intervallo di confidenza bilatero per k , ovvero $IC(k)$, con livello di confidenza γ :

$$T_1 < k < T_2$$

Il test corrispondente prevede che si rifiuti H_0 se $k_0 \leq T_1$ oppure $k_0 \geq T_2$, e il suo livello è $\alpha = 1 - \gamma$.

Test simili

Il modo di procedere finora illustrato si usa non solo nel caso in cui le ipotesi siano del tipo "uguale" contro "diverso":

$$\begin{array}{ll} H_0: \sigma^2 \leq \sigma_0^2 & H_1: \sigma^2 > \sigma_0^2 \\ H_0: \sigma^2 = \sigma_0^2 & H_1: \sigma^2 > \sigma_0^2 \\ H_0: \sigma^2 = \sigma_0^2 & H_1: \sigma^2 = \sigma_1^2, \quad \sigma_1^2 \neq \sigma_0^2 \end{array}$$

Costruiamo in particolare il test relativo al primo di questi casi; supponiamo di volere un test a livello α . Costruiamo allora un IC per la varianza, $IC(\sigma^2)$, del tipo $(c, +\infty)$ con:

$$\gamma = 1 - \alpha$$

Oververo:

$$\left(\frac{S^2(n-1)}{\chi_{n-1}^2(\gamma)}; +\infty \right)$$

In sostanza, quanto finora scritto significa che siamo certi al 95% che il valore vero della varianza è superiore a $\frac{S^2(n-1)}{\chi^2(\gamma)}$. Quindi, se $\sigma_0^2 \notin IC(\sigma^2)$, significa che i valori di σ^2 che sono indicati in H_0 , ovvero quelli inferiori a σ_0^2 , sono incompatibili con i dati. Si usa quindi la regola di rifiuto:

$$\text{rifiuto } H_0 \text{ se } \sigma_0^2 \leq \frac{S^2(n-1)}{\chi_{n-1}^2(\gamma)}$$

Possiamo verificare che questo test è effettivamente a livello:

$$\alpha = 1 - \gamma$$

Infatti:

$$\begin{aligned} \alpha &= \sup_{\sigma^2 \leq \sigma_0^2} P\left(\sigma_0^2 \leq \frac{S^2(n-1)}{\chi_{n-1}^2(\gamma)}\right) = \sup_{\sigma^2 \leq \sigma_0^2} P\left(S^2(n-1) \geq \sigma_0^2 \chi_{n-1}^2(\gamma)\right) = \\ &= \sup_{\sigma^2 \leq \sigma_0^2} P\left(\frac{S^2(n-1)}{\sigma^2} \geq \frac{\sigma_0^2 \chi_{n-1}^2(\gamma)}{\sigma^2}\right) \end{aligned}$$

Si nota poi che $\frac{S^2(n-1)}{\sigma^2}$ ha distribuzione χ_{n-1}^2 . È chiaro che la probabilità che una chi-quadro sia superiore ad un numero $\frac{\sigma_0^2 \chi_{n-1}^2(\gamma)}{\sigma^2}$ aumenta all'aumentare di σ^2 (pensando al grafico della funzione di ripartizione, "spostiamo più a sinistra la soglia", perciò "l'area a destra diventa più grande"). È chiaro allora che il limite superiore si ha quando $\sigma^2 = \sigma_0^2$; quindi:

$$\alpha = P\left(\frac{S^2(n-1)}{\sigma_0^2} \geq \frac{\sigma_0^2 \chi_{n-1}^2(\gamma)}{\sigma_0^2}\right) = P\left(\frac{S^2(n-1)}{\sigma_0^2} \geq \chi_{n-1}^2(\gamma)\right) = P\left(\frac{S^2(n-1)}{\sigma_0^2} \geq \chi_{n-1}^2(1 - \alpha)\right)$$

Nota

A differenza degli Z-test sulla media, i chi-quadro test sulla varianza non possono essere utilizzati se il campione fornito non ha distribuzione normale: in assenza di un campione gaussiano, anche di grandi dimensioni, questi test non valgono.

Verifica d'ipotesi su popolazione gaussiana: test chi-quadro sulla varianza (metodo degli IC) – caso a media nota

Il tipo di test

Sia dato il campione gaussiano:

$$X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$$

E si voglia eseguire un test del tipo:

$$H_0: \sigma^2 \in A_0$$

$$H_1: \sigma^2 \in A_1$$

Dove A_0 ed A_1 sono due insiemi disgiunti:

$$A_0 \cap A_1 = \emptyset$$

Si ipotizzi ora che la media μ sia nota, e uguale ad un certo valore μ_0 .

Statistica di test

La statistica di test è un questo caso:

$$\frac{S_0^2 n}{\sigma_0^2}$$

Dove si ricorda che:

$$S_0^2 = \frac{1}{n} \sum_{j=1}^n (x_j - \mu_0)^2$$

I quantili che si usano in questo caso sono quelli del tipo:

$$\chi_n^2(a)$$

Dove a è ad esempio $1 - \alpha$ se il test è del tipo " \leq " contro " \geq ", e così via, seguendo per il resto le stesse regole che abbiamo visto nel caso a media incognita.

Osservazione

Si osserva allora che si ha un rapporto biunivoco tra il concetto di stima intervallare e quello di test d'ipotesi. Trovare un intervallo di confidenza di un certo parametro e trovare una regione critica per eseguire un test su quel parametro sono due problemi duali, ma non sono uno la negazione dell'altro, perché "vivono in mondi diversi".

Test sui dati accoppiati – test di omogeneità sulle medie

I campioni accoppiati

I test sui dati accoppiati sono test nei quali si individua una popolazione e, per ogni individuo che vi appartiene, si misurano due diverse grandezze X ed Y . Si ottengono così dei campioni, detti **accoppiati**, del tipo:

$$(X_1, Y_1), \dots, (X_n, Y_n)$$

Ogni coppia è indipendente dalle altre, ovvero:

$$(X_1, Y_1), \dots, (X_n, Y_n) \text{ i. i. d. } \sim f(x, y, \vartheta)$$

Si noti però che non si ha necessariamente indipendenza tra i due elementi di ciascuna coppia.

Dati di questo tipo possono essere raccolti per diversi scopi:

1. Confrontare X e Y e scoprire così se seguono lo stesso modello unidimensionale. Questo significa verificare se le rilevazioni riguardanti X sono omogenee rispetto a quelle di Y , ovvero se le funzioni di ripartizione sono uguali.
2. Per verificare se X ed Y sono indipendenti, ovvero se:

$$f(x, y, \vartheta) = f_x(x, \vartheta) \cdot f_y(y, \vartheta)$$

Il test di omogeneità sui dati accoppiati

Il test di omogeneità è il primo dei due test prima descritti. Abbiamo vari casi:

1. Le ipotesi sono:

$$H_0: F = G$$

$$H_1: F \neq G$$

2. Un altro caso è quello in cui le ipotesi sono:

$$H_0: F(w) \leq G(w) \text{ per ogni } w \in \mathbb{R}$$

$$H_1: F > G \text{ per qualche } w \in \mathbb{R}$$

(e per almeno un w la disuguaglianza è stretta)

Questo significa in sostanza che, siccome la funzione di ripartizione F di X è minore della G di Y , in generale si ha buona probabilità che X assuma valori superiori ad Y , ovvero X tende ad assumere valori più grandi di Y .

3. Il terzo caso è:

$$H_0: F \geq G \text{ per ogni } w \in \mathbb{R}$$

$$H_1: F < G \text{ per qualche } w \in \mathbb{R}$$

(e per almeno un w la disuguaglianza è stretta)

In questo caso, significa che X tende ad assumere valori più piccoli di Y .

Osservazione

Speso per determinare l'omogeneità dei dati si esegue il confronto tra due semplici valori. Se ad esempio è noto che X ed Y hanno andamento normale con ugual varianza e medie incognite, allora confrontare le loro due funzioni di ripartizione equivale a confrontare tra loro i due valori:

$$\Phi\left(\frac{w - \mu_x}{\sigma^2}\right)$$

$$\Phi\left(\frac{w - \mu_y}{\sigma^2}\right)$$

Ovvero, equivale a confrontare tra loro le due medie. Notiamo perciò che i test sulla media sono in realtà casi particolari di test di omogeneità.

Il test di omogeneità sulle medie per dati accoppiati

Cerchiamo ora di capire meglio questo test. Abbiamo:

$$H_0: F = G \qquad H_1: F \neq G$$

Ovvero:

$$H_0: \mu_x = \mu_y \qquad H_1: \mu_x \neq \mu_y$$

Un'altra alternativa è quella di avere:

$$H_0: \mu_x \geq \mu_y \qquad H_1: \mu_x < \mu_y$$

L'idea base

Dato il problema:

$$H_0: \mu_x = \mu_y + \Delta \qquad H_1: \mu_x \neq \mu_y + \Delta$$

Possiamo riscriverlo nella forma:

$$H_0: \mu_x - \mu_y = \Delta \qquad H_1: \mu_x - \mu_y \neq \Delta$$

È dato inoltre il campione di dati accoppiati:

$$(X_1, Y_1), \dots, (X_n, Y_n) \text{ i. i. d. } \sim f(x, y, \boldsymbol{\theta})$$

Possiamo analizzare solo le differenze tra gli elementi di ogni singola coppia:

$$D_1, \dots, D_n$$

Dove:

$$D_i = X_i - Y_i$$

Si otterrà in questo modo, siccome ogni coppia è indipendente dalle altre, un campione:

$$D_1, D_2, \dots, D_n \text{ i. i. d.}$$

La media delle D_i sarà, naturalmente:

$$\mu_D = \mu_x - \mu_y$$

Il problema diventa allora:

$$H_0: \mu_D = \Delta \qquad H_1: \mu_D \neq \Delta$$

Questo problema è già stato analizzato:

1. Se i dati sono numerosi, possiamo usare il test del tipo:

$$\text{rifiuta } H_0 \text{ se: } \frac{|\bar{D} - \Delta|}{\sqrt{\frac{S_D^2}{n}}} \geq z_{1-\frac{\alpha}{2}}$$

Dove di solito si considerano campioni numerosi i campioni con $n \geq 30$ (ma non è una regola fissa).

2. Se i dati sono poco numerosi, ma possiamo ipotizzare che le D_j abbiano distribuzione gaussiana, allora adottiamo la regola:

$$\text{rifiuta } H_0 \text{ se: } \frac{|\bar{D} - \Delta|}{\sqrt{\frac{S_D^2}{n}}} \geq t_{n-1} \left(1 - \frac{\alpha}{2} \right)$$

Il modello tipico per i dati accoppiati è la distribuzione normale bidimensionale, e ogni sua trasformazione lineare è ancora una variabile aleatoria gaussiana, a perciò in molti casi la distribuzione delle D_j è effettivamente gaussiana.

Test sui dati accoppiati – test di indipendenza (dati gaussiani)

Campione gaussiano bidimensionale e coefficiente di correlazione

Un campione accoppiato gaussiano bidimensionale è un campione del tipo:

$$(X_1, Y_1), \dots, (X_n, Y_n) \text{ i. i. d. } \sim \mathcal{N}(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$$

Si ricorda il concetto di coefficiente di correlazione lineare ρ tra due variabili aleatorie X e Y :

$$\rho = \frac{\text{cov}(X, Y)}{\sqrt{\sigma_x^2 \sigma_y^2}}$$

Si ricorda inoltre che $|\rho| \leq 1$. Valgono le seguenti proprietà:

1. Se $|\rho| = 1$, allora con probabilità 1 le due variabili sono una la trasformazione lineare dell'altra:

$$Y = aX + 1$$

2. Se $\rho = 0$, allora le due variabili X e Y sono dette *scorrelate*.

Infine, si ha la seguente proprietà:

$$\text{Se } X \text{ e } Y \text{ sono indipendenti, allora: } E[X \cdot Y] = E[X] \cdot E[Y] \rightarrow \rho = 0$$

Non è però vero il viceversa.

Si ha allora:

$$f_{X,Y}(x, y) = \frac{1}{2\pi \cdot \sigma_x \cdot \sigma_y \cdot \sqrt{1 - \rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_x}{\sigma_x} \right)^2 + \left(\frac{y-\mu_y}{\sigma_y} \right)^2 - 2\rho \left(\frac{x-\mu_x}{\sigma_x} \cdot \frac{y-\mu_y}{\sigma_y} \right) \right]}$$

Nel caso in cui il coefficiente di correlazione lineare fosse nullo:

$$f_{X,Y}(x, y) = \frac{1}{2\pi \cdot \sigma_x \cdot \sigma_y} e^{-\frac{1}{2} \left[\left(\frac{x-\mu_x}{\sigma_x} \right)^2 + \left(\frac{y-\mu_y}{\sigma_y} \right)^2 \right]} = \frac{1}{\sqrt{2\pi} \cdot \sigma_x} e^{-\frac{1}{2} \left(\frac{x-\mu_x}{\sigma_x} \right)^2} \cdot \frac{1}{\sqrt{2\pi} \cdot \sigma_y} e^{-\frac{1}{2} \left(\frac{y-\mu_y}{\sigma_y} \right)^2}$$

Si identificano cioè i due fattori, che sono le distribuzioni di due gaussiane. Si nota quindi che, nel caso in cui si abbia un campione congiuntamente gaussiano (ma non nel caso generale):

$$X \text{ ed } Y \text{ sono indipendenti se e solo se } \rho = 0.$$

Di conseguenza, per eseguire un test di indipendenza tra due variabili aleatorie gaussiane è sufficiente eseguire un test d'indipendenza sul loro coefficiente di correlazione.

I vari casi di test d'indipendenza su gaussiane

Possiamo avere diversi tipi di ipotesi:

1. Caso 1:

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

2. Caso 2:

$$H_0: \rho = 0$$

$$H_1: \rho > 0$$

Si noti che testare $\rho > 0$ (*dipendenza positiva*) significa andare a verificare se, all'aumentare di X , aumenta anche Y . Esiste anche un caso simile (risolto con lo stesso test):

- 2.1 Si ha:

$$H_0: \rho \leq 0$$

$$H_1: \rho > 0$$

3. Caso 3:

$$H_0: \rho = 0$$

$$H_1: \rho < 0$$

Esiste anche un caso simile (risolto con lo stesso test):

- 3.1 Si ha:

$$H_0: \rho \geq 0$$

$$H_1: \rho < 0$$

Statistica test

Per prima cosa, vediamo come stimare ρ . Per analogia rispetto allo stimatore S^2 per la varianza, possiamo pensare di stimare la covarianza tra X ed Y mediante lo stimatore:

$$\widehat{cov}(X, Y) = \sum_{j=1}^n \frac{(X_j - \bar{X})(Y_j - \bar{Y})}{n - 1}$$

Siccome inoltre stimiamo:

$$\sigma_x^2 = S_x^2$$

$$\sigma_y^2 = S_y^2$$

Avremo:

$$\hat{\rho} = R = \frac{\widehat{cov}(X, Y)}{\sqrt{\sigma_x^2 \sigma_y^2}} = \frac{\sum_{j=1}^n \frac{(X_j - \bar{X})(Y_j - \bar{Y})}{n - 1}}{\sqrt{\sum_{j=1}^n \frac{(X_j - \bar{X})^2}{n - 1} \sum_{j=1}^n \frac{(Y_j - \bar{Y})^2}{n - 1}}} = \frac{\sum_{j=1}^n (X_j - \bar{X})(Y_j - \bar{Y})}{\sqrt{\sum_{j=1}^n (X_j - \bar{X})^2 \sum_{j=1}^n (Y_j - \bar{Y})^2}}$$

Continua a valere la proprietà:

$$|R| \leq 1$$

Possiamo inoltre introdurre la statistica test:

$$ST = \frac{R}{\sqrt{1 - R^2}} \sqrt{n - 2}$$

Che, nel caso in cui valga l'ipotesi H_0 , si dimostra avere una distribuzione t-Student:

$$ST = \frac{R}{\sqrt{1 - R^2}} \sqrt{n - 2} \sim t_{n-2}, \quad n \geq 3$$

Ci conviene perciò adottare proprio ST come statistica test. Si noti inoltre che ST è pari.

Come eseguire il test

Possiamo allora adottare le seguenti regole per eseguire i test:

1. Caso 1:

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

Allora la regola da seguire è:

Rifiuto l'ipotesi nulla H_0 se

$$|ST| \geq t_{n-2} \left(1 - \frac{\alpha}{2}\right)$$

2. Caso 2:

$$H_0: \rho = 0$$

$$H_1: \rho > 0$$

Allora la regola da seguire è:

Rifiuto l'ipotesi nulla H_0 se

$$ST \geq t_{n-2}(1 - \alpha)$$

3. Caso 3:

$$H_0: \rho = 0$$

$$H_1: \rho < 0$$

Allora la regola da seguire è:

Rifiuto l'ipotesi nulla H_0 se

$$ST \leq -t_{n-2}(1 - \alpha)$$

Test sui dati accoppiati: Test di Wilcoxon (omogeneità)

Sia dato un campione di dati accoppiati:

$$(X_1, X_2), \dots, (X_n, Y_n) \text{ i. i. d.}$$

E si voglia risolvere un problema del tipo:

$$H_0: F = G$$

$$H_1: F \leq G$$

Dove F è la funzione di ripartizione di X e G è quella di Y , e dove, come già accennato, $F \leq G$ significa che X tende ad assumere valori superiori rispetto ad Y .

Ipotizziamo che non ci sia alcuna ripetizione nei dati, né all'interno di una coppia, né all'esterno di una coppia. In questo caso, possiamo applicare il test di Wilcoxon.

Si noti che l'ipotesi imposta risulta essere verificata se si ipotizza che la distribuzione congiunta sia assolutamente continua:

$$P(X_i = Y_j) = P(X_i = X_j) = P(Y_i = Y_j) = 0 \quad \forall i = 1, \dots, n, \forall j = 1, \dots, n,$$

(escludendo naturalmente i casi in cui dalle due parti dell'uguale si ha la stessa variabile).

In ogni caso, questo non è l'unico caso in cui l'ipotesi risulti verificata. Si noti comunque che questa è l'unica ipotesi che occorre imporre.

Idee base

L'ipotesi nulla è $H_0: F = G$. Come abbiamo ipotizzato, si ha:

$$P(X = Y) = 0$$

Inoltre, siccome le due funzioni di ripartizioni sono uguali, varrà anche:

$$P(X > Y) = P(X < Y)$$

Unendo tali condizioni si ricava in modo ovvio:

$$p = P(X > Y) = P(X < Y) = \frac{1}{2}$$

Quindi possiamo riscrivere il problema nella forma:

$$H_0: P(X > Y) = \frac{1}{2}$$

$$H_1: P(X > Y) > \frac{1}{2}$$

Statistica test

La statistica test è:

$$ST = n^\circ \text{ di coppie con } X > Y$$

Si avrà così, nell'ipotesi che H_0 sia vera:

$$ST \sim \text{Bin}\left(n, \frac{1}{2}\right)$$

Regola

La regola di decisione è la seguente:

$$\text{Rifiuto } H_0 \text{ se} \\ ST > q_{\text{Bin}(n, \frac{1}{2})}(1 - \alpha)$$

Tuttavia, le tavole della binomiale non sono disponibili durante l'esame. Si deve quindi lavorare utilizzando il p-value.

P-value

Sia s il valore della statistica test. Abbiamo:

$$\begin{aligned} pvalue &= P(ST > s) = P\left(\text{Bin}\left(n, \frac{1}{2}\right) > s\right) = 1 - P\left(\text{Bin}\left(n, \frac{1}{2}\right) \leq s\right) = 1 - \sum_{k=0}^s \binom{n}{k} p^k (1-p)^{n-k} = \\ &= 1 - \sum_{k=0}^s \binom{n}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k} = 1 - \left(\frac{1}{2}\right)^n \sum_{k=0}^s \binom{n}{k} \end{aligned}$$

Caso di grandi campioni (approssimazione con la normale)

Se n è grande, approssimativamente:

$$ST \sim \mathcal{N}(np, np(1-p))$$

Perciò, sotto H_0 :

$$ST \sim \mathcal{N}\left(\frac{n}{2}, \frac{n}{4}\right)$$

E quindi usiamo la regola:

$$\begin{aligned} &\text{Rifiuto } H_0 \text{ se} \\ &ST > q_{\text{Bin}(n, \frac{1}{2})}(1-\alpha) \cong \sqrt{\frac{n}{4}} \cdot z_{1-\alpha} + \frac{n}{2} \end{aligned}$$

Ipotesi diverse

Se consideriamo il caso in cui le ipotesi siano:

$$H_0: F = G$$

$$H_1: F \geq G$$

Allora la regola diventa:

$$\begin{aligned} &\text{Rifiuto } H_0 \text{ se} \\ &ST < q_{\text{Bin}(n, \frac{1}{2})}(\alpha) \end{aligned}$$

E il p-value è in questo caso:

$$pvalue = P\left(\text{Bin}\left(n, \frac{1}{2}\right) < s\right) = \left(\frac{1}{2}\right)^n \sum_{k=0}^{s-1} \binom{n}{k}$$

Test di Wilcoxon-Mann-Whitney (omogeneità dati non accoppiati)

Si considerino ora due campioni indipendenti:

$$X_1, X_2, \dots, X_m \text{ i. i. d. } \sim F$$

$$Y_1, Y_2, \dots, Y_n \text{ i. i. d. } \sim G$$

E si voglia ancora una volta eseguire un test del tipo:

$$H_0: F = G$$

$$H_1: F \leq G$$

Nel caso in cui si ipotizza che non ci siano ripetizioni sui dati:

$$P(X_i = Y_j) = P(X_i = X_j) = P(Y_i = Y_j) = 0 \quad \forall i, j$$

(escludendo naturalmente i casi in cui dalle due parti dell'uguale si ha la stessa variabile).

In particolare, questa ipotesi è certamente verificata se le distribuzioni F e G sono continue.

Statistica test

Introduciamo la statistica:

$$U = n^\circ \text{ di coppie in cui } X \text{ è maggiore di } Y \text{ sulle } m \cdot n \text{ coppie } (X_i, Y_j), \text{ per } i = 1, \dots, n \text{ e } j = 1, \dots, n$$

Nel caso in cui l'ipotesi nulla sia vera, ci si aspetta di avere (lo si può verificare, non lo facciamo):

$$E_0[U] = \frac{m \cdot n}{2} \qquad \text{VAR}_0[U] = \frac{m \cdot n \cdot (m + n + 1)}{12}$$

Sulla base di U possiamo allora costruire la nostra statistica test, ovvero:

$$T_x = R_1 + R_2 + \dots + R_m$$

Dove R_i viene ottenuto considerando la "graduatoria" finale che si ottiene ordinando tutti i valori di X e di Y secondo l'ordine crescente, e rappresenta in particolare la posizione dell' i -esima X che si incontra scorrendo la graduatoria dal più piccolo valore al più grande.

Avremo allora:

$$\min(T_X) = 1 + 2 + \dots + m = \frac{m(m+1)}{2}$$

$$\max(T_X) = (n+1) + (n+2) + \dots + (n+m) = nm + 1 + 2 + \dots + m = nm + \frac{m(m+1)}{2} = m\left(n + \frac{m+1}{2}\right)$$

Si può inoltre osservare che:

$$U = (R_1 - 1) + (R_2 - 2) + \dots + (R_m - m) = R_1 + R_2 + \dots + R_m - (1 + 2 + \dots + m) = T_X - \frac{m(m+1)}{2} \rightarrow T_X = U + \frac{m(m+1)}{2}$$

Perciò:

$$E_0[T_X] = E_0[U] - \frac{m(m+1)}{2} = \frac{mn}{2} - \frac{m(m+1)}{2} = \frac{m(m+n+1)}{2}$$

$$VAR_0[T_X] = VAR_0[U] = \frac{mn(m+n+1)}{12}$$

La regola di rifiuto

Possiamo allora usare la seguente regola:

$$\text{Rifiuto } H_0 \text{ se: } T_X > W_{m,n}(1 - \alpha)$$

Dove $W_{m,n}$ è un valore che si trova sulle tavole di Mann-Whitney, che però riportano solamente valori piccoli.

Nel caso in cui i valori m e n siano grandi (in genere $m > 7$ e $n > 7$), sotto l'ipotesi nulla si ha:

$$T_X \sim \mathcal{N}\left(\frac{m(m+n+1)}{2}, \frac{mn(m+n+1)}{12}\right)$$

E perciò si usano i quantili della normale:

$$\text{Rifiuto } H_0 \text{ se: } T_X > \frac{m(m+n+1)}{2} + z_{1-\alpha} \sqrt{\frac{mn(m+n+1)}{12}}$$

Osservazioni

1. Un basato su T_Y porterebbe alle stesse conclusioni alle quali porta il test basato su T_X . Infatti:

$$T_Y + T_X = 1 + 2 + \dots + m + n = \frac{(m+n)(m+n+1)}{2}$$

Quindi possiamo dire che T_X è una trasformazione lineare di T_Y . Questo vale perché abbiamo ipotizzato che non ci siano ripetizioni nei campioni dati.

2. Se si hanno poche ripetizioni, si usa lo stesso meccanismo, ma si assegnando agli elementi che occupano la stessa posizione un valore intermedio. Ad esempio, se si hanno 3 elementi in 6° posizione, si associa a ciascuno di essi il 7° posto, e al successivo si assegna il 9°.
3. La tavola di Mann-Whitney contiene direttamente i valori in funzione di α non superiori al 10%, perciò se si cercano i valori $1 - \alpha$ superiori al 10%, occorre utilizzare la relazione (riportata anche sulle tavole):

$$w_p = m(m+n+1) - w_{1-p}$$

Test di omogeneità su campioni gaussiani indipendenti

Siano dati due campioni gaussiani indipendenti:

$$X_1, \dots, X_m \text{ i. i. d. } \sim \mathcal{N}(\mu_X, \sigma_X^2)$$

$$Y_1, \dots, Y_m \text{ i. i. d. } \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$$

Si voglia eseguire un test che permetta di stabilire se le due distribuzioni sono uguali, ovvero:

$$H_0: \begin{cases} \mu_X = \mu_Y \\ \sigma_X^2 = \sigma_Y^2 \end{cases}$$

$$H_1: \mu_X \neq \mu_Y \text{ oppure } \sigma_X^2 \neq \sigma_Y^2$$

Il test da eseguire è sequenziale, ovvero:

1. Eseguiamo il test sulla varianza; se non rifiutiamo l'ipotesi secondo la quale le due varianze sono uguali
2. Eseguiamo il test sulla media; se non rifiutiamo l'ipotesi secondo la quale le due medie sono uguali, allora non rifiutiamo nemmeno l'ipotesi che le due distribuzioni siano uguali.

Se anche solo una delle due ipotesi nulle viene rifiutata, concludiamo che i due modelli sono diversi.

Quindi dobbiamo eseguire:

1. Il test di confronto sulle varianze:

$$H_0: \sigma_X^2 = \sigma_Y^2$$

$$H_1: \sigma_X^2 \neq \sigma_Y^2$$

Chiamiamo α_1 il livello di significatività col quale eseguiamo tale test.

2. Il test di confronto sulle medie:

$$H_0: \mu_X = \mu_Y$$

$$H_1: \mu_X \neq \mu_Y$$

Ma solo a patto che, con livello α_1 non sia stata rifiutata l'ipotesi $H_0: \sigma_X^2 = \sigma_Y^2$. Il secondo test viene eseguito con un certo livello di significatività che indichiamo con α_2 .

Il livello di significatività complessivo del test è:

$$1 - (1 - \alpha_1)(1 - \alpha_2) = 1 - (1 - \alpha_2 - \alpha_1 + \alpha_1\alpha_2) = \alpha_1 + \alpha_2 - \alpha_1\alpha_2 \cong \alpha_1 + \alpha_2$$

Vediamo ora nel dettaglio i due passi.

Passo 1 – test sulla varianza (anche nel caso generale legato dal problema in analisi) – F-test

Si considerano, come abbiamo già visto:

$$H_0: \sigma_X^2 = \sigma_Y^2$$

$$H_1: \sigma_X^2 \neq \sigma_Y^2$$

Con dati:

$$X_1, \dots, X_m \text{ i. i. d. } \sim \mathcal{N}(\mu_X, \sigma_X^2)$$

$$Y_1, \dots, Y_m \text{ i. i. d. } \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$$

E con livello di significatività α_1 . Le medie sono incognite. Tuttavia, potremmo considerare questo problema isolatamente rispetto al contesto in cui stiamo operando, perciò considereremo in seguito anche il caso di medie note, e analizzeremo anche alcuni particolari problemi simili, come:

$$H_0: \sigma_X^2 \leq \sigma_Y^2$$

$$H_1: \sigma_X^2 > \sigma_Y^2$$

$$H_0: \sigma_X^2 \geq \sigma_Y^2$$

$$H_1: \sigma_X^2 < \sigma_Y^2$$

che prevedono l'uso della stessa statistica test. Restiamo però al caso di medie incognite:

Statistica test

Concentriamoci ancora sul problema:

$$H_0: \sigma_X^2 = \sigma_Y^2$$

$$H_1: \sigma_X^2 \neq \sigma_Y^2$$

In questo caso, possiamo riscrivere in maniera ovvia il problema come:

$$H_0: \frac{\sigma_X^2}{\sigma_Y^2} = 1$$

$$H_1: \frac{\sigma_X^2}{\sigma_Y^2} \neq 1$$

Perciò possiamo cercare di stimare il rapporto tra le due varianze. Sappiamo che la stima per σ_X^2 è S_X^2 e che la stima per σ_Y^2 è S_Y^2 , perciò è ovvio usare lo stimatore:

$$ST = \frac{S_X^2}{S_Y^2}$$

Che ricopre anche il ruolo di statistica test.

Nota: La F-Fisher

Introduciamo ora una distribuzione che ci sarà utile per eseguire questo test: la F di Fisher. Siano date due variabili aleatorie W_1 e W_2 indipendenti:

$$W_1 \sim \chi_a^2 \qquad W_2 \sim \chi_b^2$$

E sia Z la variabile aleatoria così definita:

$$Z = \frac{\frac{W_1}{a}}{\frac{W_2}{b}} = \frac{bW_1}{aW_2}$$

Allora $Z \geq 0$ e Z è una variabile aleatoria continua; la sua densità è detta F di Fisher con a gradi di libertà al numeratore e b gradi di libertà al denominatore, e la si indica:

$$Z \sim F_{a,b}$$

Il quantile di ordine p di tale distribuzione di probabilità si indica con:

$$F_{a,b}(p)$$

Se si conosce il valore $F_{a,b}(1-p)$ e si vuole determinare $F_{a,b}(p)$, si può seguire il procedimento così descritto:

$$p = P\left(Z \leq F_{a,b}(p)\right) = P\left(\frac{\frac{W_1}{a}}{\frac{W_2}{b}} \leq F_{a,b}(p)\right) = P\left(\frac{\frac{W_2}{b}}{\frac{W_1}{a}} \geq \frac{1}{F_{a,b}(p)}\right)$$

Da cui si ricava:

$$1 - p = P\left(F_{b,a} \leq \frac{1}{F_{a,b}(p)}\right)$$

E quindi:

$$F_{b,a}(1-p) = \frac{1}{F_{a,b}(p)}$$

O, equivalentemente:

$$F_{a,b}(p) = \frac{1}{F_{b,a}(1-p)}$$

Distribuzione della statistica test

Sappiamo che:

$$\frac{S_X^2(m-1)}{\sigma_X^2} \sim \chi_{m-1}^2 \qquad \frac{S_Y^2(n-1)}{\sigma_Y^2} \sim \chi_{n-1}^2$$

Quindi, dalla definizione fornita di *F di Fisher* discende che:

$$\frac{\frac{S_X^2(m-1)}{\sigma_X^2} \cdot \frac{1}{m-1}}{\frac{S_Y^2(n-1)}{\sigma_Y^2} \cdot \frac{1}{n-1}} = \frac{\sigma_Y^2}{\sigma_X^2} \frac{S_X^2}{S_Y^2} \sim F_{m-1,n-1}$$

Nel caso in cui H_0 sia vera, abbiamo $\sigma_Y^2 = \sigma_X^2$, perciò ricaviamo che in questo caso:

$$ST = \frac{S_X^2}{S_Y^2} \sim F_{m-1,n-1}$$

La regione critica

Si rifiuta l'ipotesi nulla H_0 se:

$$ST \leq F_{m-1,n-1}\left(\frac{\alpha}{2}\right) \qquad \text{oppure} \qquad ST \geq F_{m-1,n-1}\left(1 - \frac{\alpha}{2}\right)$$

Il caso di test unilatero

Consideriamo ora il caso in cui il test sia unilatero:

$$H_0: \sigma_X^2 \leq \sigma_Y^2 \qquad H_1: \sigma_X^2 > \sigma_Y^2$$

In questo caso la regione critica viene così modificata:

$$\text{Rifiuto } H_0 \text{ se } ST \geq F_{m-1, n-1}(1 - \alpha)$$

Calcolo della potenza

In un F-test è sempre possibile calcolare in maniera analitica la potenza del test. Ad esempio:

$$P_{\sigma_X^2=2\sigma_Y^2} \left(\frac{S_X^2}{S_Y^2} \geq F_{m-1, n-1}(1 - \alpha) \right) = P_{\sigma_X^2=2\sigma_Y^2} \left(\frac{S_X^2}{2S_Y^2} \geq \frac{1}{2} F_{m-1, n-1}(1 - \alpha) \right)$$

Il caso a medie note

Nel caso in cui le medie siano note, si procede come finora visto, utilizzando però la seguente intuitiva "tabella delle sostituzioni":

Al posto di...	Si usa...
S_X^2, S_Y^2	S_{0X}^2, S_{0Y}^2
$F_{m-1, n-1}$	$F_{m, n}$

Osservazione

Questi test possono essere eseguiti solo se i campioni sono *effettivamente* gaussiani.

Passo 2 – test sulla media (anche nel caso generale non legato a questo problema)

Prendiamo adesso in analisi il test sulle medie. Abbiamo ancora 2 campioni gaussiani indipendenti:

$$X_1, \dots, X_m \text{ i. i. d. } \sim \mathcal{N}(\mu_X, \sigma_X^2) \qquad Y_1, \dots, Y_n \text{ i. i. d. } \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$$

Anche in questo caso, le varianze possono essere note o incognite. Nel problema che stiamo analizzando (quello del test di omogeneità su due distribuzioni gaussiane) sono incognite, ma lo vediamo qui nel caso generale.

Possiamo scrivere il test da eseguire, ovvero:

$$H_0: \mu_X = \mu_Y \qquad H_1: \mu_X \neq \mu_Y$$

Come:

$$H_0: \mu_X - \mu_Y = \Delta \qquad H_1: \mu_X - \mu_Y \neq \Delta$$

Dove Δ è un numero prefissato. Iniziamo analizzando il caso a varianze note.

Idee base

1. Stimiamo la differenza tra le medie, $\mu_X - \mu_Y$, con:

$$\bar{X} - \bar{Y}$$

2. Abbiamo:

$$(\bar{X} - \bar{Y}) \sim \mathcal{N}(\mu_X - \mu_Y, \sigma_{\bar{X}-\bar{Y}}^2)$$

Dove:

$$\sigma_{\bar{X}-\bar{Y}}^2 = \text{VAR}(\bar{X} - \bar{Y}) = \text{VAR}(\bar{X}) + \text{VAR}(\bar{Y}) - 2\text{cov}(\bar{X}, \bar{Y}) = \text{VAR}(\bar{X}) + \text{VAR}(\bar{Y}) = \frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}$$

Perciò:

$$\frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}} \sim \mathcal{N}(0,1)$$

Statistica test

Sotto l'ipotesi nulla, abbiamo $\mu_X - \mu_Y = \Delta$, perciò possiamo definire la statistica test:

$$ST = \frac{(\bar{X} - \bar{Y}) - \Delta}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}}$$

Il cui andamento, come abbiamo appena dimostrato è quello di una gaussiana standard.

Regione critica

Di conseguenza, utilizziamo la seguente regola di decisione:

$$\text{Rifiuto } H_0 \text{ se} \\ |ST| \geq z \left(1 - \frac{\alpha}{2}\right)$$

Osservazione

Questo test può essere utilizzato anche nel caso in cui i dati raccolti non siano normali, ma si hanno grandi campioni, sempre a patto che siano note le varianze.

Il caso a varianze incognite

1. Nel caso in cui le varianze siano incognite, la statistica test può essere scelta come:

$$ST_1 = \frac{(\bar{X} - \bar{Y}) - \Delta}{\sqrt{\frac{S_X^2}{m} + \frac{S_Y^2}{n}}}$$

Solo a patto che si abbiano grandi campioni. In questo caso:

$$\text{Rifiuto } H_0 \text{ se } |ST_1| \geq z_{1-\frac{\alpha}{2}}$$

2. Nel caso in cui m e/o n siano piccoli, questo test non va bene, neppure nel caso in cui la loro distribuzione sia effettivamente una gaussiana. Il problema di individuare un test per questo caso è tutt'oggi aperto, anche se sono state proposte varie soluzioni valide in casi diversi (tra queste però non è ancora emersa alcuna).

3. Nel caso particolare in cui si sia a conoscenza del fatto che le due varianze, pur essendo incognite, sono uguali tra loro, possiamo costruire un ottimale nella classe dei test non distorti. Per questo motivo è necessario eseguire prima il test sulla varianza e solo in seguito il test sulle medie.

In particolare, in questo caso, detta:

$$\sigma^2 = \sigma_X^2 = \sigma_Y^2$$

Abbiamo:

$$\frac{(\bar{X} - \bar{Y}) - \Delta}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}} = \frac{(\bar{X} - \bar{Y}) - \Delta}{\sqrt{\sigma^2 \left(\frac{1}{m} + \frac{1}{n}\right)}} \sim \mathcal{N}(0,1)$$

Naturalmente rimane il problema che σ^2 è incognita. Possiamo allora stimarla partendo dalla quantità:

$$\text{somme} = \sum_{i=1}^m (X_i - \bar{X})^2 + \sum_{i=1}^n (Y_i - \bar{Y})^2$$

Affinché $\hat{\sigma}^2$ non sia distorto, dobbiamo calcolare:

$$E[\text{somme}] = E[(m-1)S_X^2 + (n-1)S_Y^2] = (m-1)\sigma^2 + (n-1)\sigma^2 = (m+n-2)\sigma^2$$

Quindi possiamo costruire lo stimatore seguente:

$$S_{pooled}^2 = \frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}$$

Che, per come è stato costruito, è uno stimatore non distorto di σ^2 . Possiamo ora sfruttare l'ipotesi di normalità del campione; infatti abbiamo:

$$\frac{S_{pooled}^2(m+n-2)}{\sigma^2} = \frac{(m-1)S_X^2 + (n-1)S_Y^2}{\sigma^2} = \frac{(m-1)S_X^2}{\sigma^2} + \frac{(n-1)S_Y^2}{\sigma^2}$$

Overo abbiamo ottenuto la somma tra una χ_{m-1}^2 e una χ_{n-1}^2 , indipendenti tra loro (a seguito del fatto che sono le varianze campionarie di campioni indipendenti). Perciò:

$$\frac{S_{pooled}^2(m+n-2)}{\sigma^2} \sim \chi_{m+n-2}^2$$

Usiamo quindi come statistica test:

$$ST = \frac{(\bar{X} - \bar{Y}) - \Delta}{\sqrt{S_{pooled}^2 \left(\frac{1}{m} + \frac{1}{n}\right)}}$$

E usiamo come quantità pivotale:

$$\frac{S_{pooled}^2(m+n-2)}{\sigma^2} \sim \chi_{m+n-2}^2$$

Sotto l'ipotesi nulla H_0 , notiamo che numeratore e denominatore di ST sono indipendenti, perché la media e la varianza campionaria di un campione gaussiano godono di tale proprietà. Perciò:

$$ST \sim t_{m+n-2}$$

In conclusione, con livello α adottiamo la seguente regione critica:

$$\mathcal{G} = \left\{ (X_1, X_2, \dots, X_m, Y_1, Y_2, \dots, Y_n) \mid \left| \frac{(\bar{X} - \bar{Y}) - \Delta}{\sqrt{S_{pooled}^2 \left(\frac{1}{m} + \frac{1}{n}\right)}} \right| \geq t_{n+m-2} \left(1 - \frac{\alpha}{2}\right) \right\}$$

Osservazioni

1. Se la varianza è incognita, non possiamo calcolare la potenza del test sulle medie, salvo che nel caso in cui si abbiano tanti dati, perché si dovrebbe ipotizzare $\mu_X - \mu_Y \neq \Delta$, e quindi in tal caso si otterrebbe una distribuzione t-student non centrata, e operare con tale tipo di distribuzione richiede strumenti che in questo corso non vengono introdotti.
2. Se si hanno dati la cui distribuzione non è gaussiana con certezza, è più opportuno utilizzare il test di Wilcoxon-Mann-Whitney, mentre nel caso in cui i dati sono certamente normali è meglio utilizzare il test appena introdotti.

Test chi-quadro di Pearson – per il buon adattamento (goodness of fit)

I test di buon adattamento

I test di buon adattamento sono test che vengono utilizzati per validare un modello probabilistico, ovvero per verificare se tale modello si adatta ad un set di dati di cui si dispone. Le ipotesi sono allora:

$$H_0: X \sim F_0(\vartheta_1, \dots, \vartheta_m)$$

$$H_1: X \not\sim F_0(\vartheta_1, \dots, \vartheta_m)$$

Per eseguire il test, disponiamo di un campione di dati:

$$X_1, \dots, X_n \text{ i. i. d.}$$

Se supponiamo che sia vera l'ipotesi nulla, allora il modello può essere completamente specificato oppure no, in quanto è possibile che l'ipotesi nulla specifichi solamente una certa famiglia di modelli, e non un particolare modello specifico. Ad esempio, l'ipotesi nulla potrebbe essere del tipo "X ha andamento normale", senza però specificarne media e varianza. Si hanno allora due diversi casi:

1. H_0 semplice: F_0 è completamente specificata, cioè indica un valore per ciascuno dei parametri $\vartheta_1, \dots, \vartheta_m$.
2. H_0 composta: almeno un parametro della distribuzione di probabilità F_0 non è fissato da H_0 .

Si noti che i dati possono essere:

- a) Discreti
- b) Continui
- c) Categorici (ad esempio, fasce di età)

Proprio a seguito della possibilità di avere dati categorici, si parla di modello probabilistico e non di funzione di ripartizione, in quanto nel caso di dati categorici non esiste un ordinamento tra i dati stessi, e perciò non esiste il concetto di funzione di ripartizione.

Il test chi quadro di Pearson

Il test chi quadro di Pearson è un test asintotico (valido cioè solo per grandi campioni), anche se ne esistono varianti anche per piccoli campioni (che però non verranno qui analizzate).

Supponiamo ad esempio che l'ipotesi nulla sia del tipo:

$H_0: X$ ha densità:

X	a_1	...	a_k
Densità	f_{0_1}	...	f_{0_k}

Abbiamo allora:

$$f_{0_j} = P(X = a_j)$$

Consideriamo in particolare:

$$f_{0_j} = \frac{1}{k} \text{ per ogni } j = 1, \dots, k$$

Consideriamo invece come ipotesi alternativa:

$$H_1: f_{0_j} \neq f_j \text{ per almeno un } j$$

Statistica test

La statistica test può essere costruita contando quante volte ogni possibile modalità si ripete nei dati raccolti:

X	a_1	...	a_k
Densità $H_0 = f_{0_j}$	f_{0_1}	...	f_{0_k}
Freq. attese	nf_{0_1}	...	nf_{0_k}
Freq. reali N_j	N_1	...	N_k

Dove N_j è il numero di osservazioni su n nelle quali si ha $X = a_j$. Sappiamo che:

$$N_1 \sim \text{Bin}(n, f_{0_1})$$

In particolare, sotto l'ipotesi nulla, avremo:

$$N_1 \sim \text{Bin}(n, f_{0_1})$$

Da cui ricaviamo facilmente:

$$E[N_1] = nf_{0_1}$$

Possiamo allora aggiungere alla precedente tabella una riga contenente la misura della distanza dal valore atteso per la frequenza di ogni diversa modalità:

X	a_1	...	a_k
Densità $H_0 = f_{0_j}$	f_{0_1}	...	f_{0_k}
Freq. attese	nf_{0_1}	...	nf_{0_k}
Freq. reali N_j	N_1	...	N_k
$\frac{(N_j - nf_{0_j})^2}{nf_{0_j}}$	$\frac{(N_1 - nf_{0_1})^2}{nf_{0_1}}$...	$\frac{(N_k - nf_{0_k})^2}{nf_{0_k}}$

Definiamo a questo punto la statistica:

$$Q = \sum_{j=1}^k \left[\frac{(N_j - nf_{0_j})^2}{nf_{0_j}} \right]$$

Che rappresenta proprio la statistica test del test chi-quadro di buon adattamento di Pearson.

Si noti che, anche se Q è definito dalla precedente espressione, può anche essere calcolato come:

$$Q = \sum_{j=1}^k \left[\frac{N_j^2 + n^2 f_{0_j}^2 - 2nf_{0_j}N_j}{nf_{0_j}} \right] = \sum_{j=1}^k \left[\frac{N_j^2}{nf_{0_j}} + nf_{0_j} - 2N_j \right] = \sum_{j=1}^k \frac{N_j^2}{nf_{0_j}} + n - 2n = \sum_{j=1}^k \frac{N_j^2}{nf_{0_j}} - n$$

La regione critica

Il test prevede che:

$$\text{Rifiuto } H_0 \text{ se } Q \geq K^*, \text{ dove } K^*: P_{H_0}(\text{rifiuto } H_0) = \alpha$$

Per meglio definire la regione critica, è quindi necessario valutare K^* , che si ricava facilmente essere:

$$K = q_{Q_{H_0}}(1 - \alpha)$$

Siccome si ha:

$$\lim_{n \rightarrow +\infty} P(Q \leq x) = F_{\chi_{k-1}^2}(x) \quad \forall x \in \mathbb{R}$$

Possiamo affermare che, all'aumentare delle osservazioni, il modello è simile al modello chi-quadro; per n sufficiente grande, avremo allora:

$$K = q_{Q_{H_0}}(1 - \alpha) \cong \chi_{k-1}^2(1 - \alpha)$$

Condizione pratica per l'uso del test

Nella pratica si può usare questo test se valgono le seguenti ipotesi:

- $n \geq 30$
- $nf_{0_j} \geq 5$ per ogni j .

Alcuni libri di testo usano una diversa condizione pratica, imponendo $nf_{0_j} \geq 1$ per ogni j e $nf_{0_j} > 5$ per almeno l'80% delle modalità.

Dati non categorici – ipotesi nulla semplice

Fino ad ora abbiamo considerato solamente dati categorici. Tuttavia, abbiamo già affermato che si può in realtà trattare anche di dati continui o discreti, per i quali è definito il concetto di funzione di ripartizione. In tal caso, avremo un'ipotesi nulla del tipo:

$$H_0: X \sim F_0(\vartheta_{0_1}, \dots, \vartheta_{0_m})$$

Dove X è una variabile aleatoria continua o una variabile aleatoria discreta numerabili. Considereremo inizialmente solo ipotesi nulle semplici.

In questo caso, si parte dai dati grezzi, i quali vengono discretizzati tramite la costruzione di un certo numero k di intervalli del tipo:

A_1	A_2	...	A_k
$(a_0, a_1]$	$(a_1, a_2]$...	$(a_{k-1}, a_k]$

Si noti che l'estremo superiore a_k può anche essere infinito.

Dopo tale operazione, si contano quante osservazioni cadono in ciascuna delle classi così individuate e, analogamente a quanto visto nel caso precedente, chiamiamo N_j tale quantità.

Si calcolano poi le probabilità che una distribuzione del tipo specificato dall'ipotesi nulla assuma un valore, di volta in volta, in ciascuno degli intervalli A_j :

$$f_{0j} = P_{H_0}(X \in A_j) = F_0(a_j) - F_0(a_{j-1})$$

Si ottiene così una tabella del tipo:

A_j	A_1	A_2	...	A_k
N_j	N_1	N_1	...	N_k
f_{0j}	f_{01}	f_{02}	...	f_{0k}
nf_{0j}	nf_{01}	nf_{02}	...	nf_{0k}

A questo punto possiamo procedere esattamente come descritto nel caso con modalità.

Osservazioni

1. Alcune classi possono in realtà essere dei valori singoli, altre possono essere intervalli limitati o illimitati.
2. La scelta nel numero di classi è un punto critico: più classi si costruiscono, più il test è affidabile; d'altro canto però se k è troppo grande, si ha il rischio di avere classi con un numero di elementi troppo bassi, al punto da far cadere la validità delle approssimazioni asintotico.
3. Il posizionamento dei tagli è un altro punto difficile da risolvere. Se l'intervallo è limitato, allora è possibile posizionare i tagli tra i vari intervalli in modo da ripartire uniformemente lo spazio dei valori che possono essere assunti; altrimenti, si sceglie in genere una ripartizione uniforme in termini di probabilità (calcolate ovviamente sotto l'ipotesi di validità di H_0).

Dati non categorici – ipotesi nulla composta

Nel caso in cui l'ipotesi nulla sia composta, ovvero del tipo:

$$H_0: X \sim F_0(\vartheta_{01}, \dots, \vartheta_{0m})$$

Con almeno uno dei parametri incogniti, è chiaro che le quantità:

$$f_{0j} = F_0(a_j) - F_0(a_{j-1})$$

Dipendono da parametri incogniti; si stimeranno perciò tali parametri con opportuni stimatori $\hat{\vartheta}_i$, e quindi si potrà calcolare:

$$\hat{f}_{0j} = F_0(a_j; \hat{\vartheta}_1, \dots, \hat{\vartheta}_m) - F_0(a_{j-1}; \hat{\vartheta}_1, \dots, \hat{\vartheta}_m)$$

Dopodiché il modo di procedere è lo stesso descritto nei precedenti casi, salvo il fatto che al termine la statistica test da considerare è:

$$ST = \sum_{j=1}^k \left[\frac{(N_j - n\hat{f}_{0j})^2}{n\hat{f}_{0j}} \right]$$

E che l'ipotesi nulla viene rifiutata nel caso:

$$ST \geq \chi_{k-1-r}^2(1 - \alpha)$$

Dove r è il numero di parametri stimati sotto H_0 .

Test chi-quadro di indipendenza

Abbiamo già introdotto in passato il T-test di indipendenza, valido per campioni gaussiani. Cerchiamo ora di introdurre un test più generico che ci consenta di capire se si può escludere l'indipendenza tra 2 caratteri X e Y , che possono essere discreti, categorici o continui.

Avremo allora:

$$H_0: X \text{ e } Y \text{ sono indipendenti}$$

$$H_1: X \text{ e } Y \text{ non sono indipendenti}$$

Dati categorici

Supponiamo che X e Y possano assumere rispettivamente le modalità a_i e b_j in tabella; nella tabella sono indicate anche le frequenze. In particolare, si ha a monte un campione accoppiato bidimensionale:

$$(X_1, Y_1), \dots, (X_n, Y_n) \text{ i. i. d.}$$

E quindi possiamo contare quante sono le coppie N_{11} nelle quali la X ha modalità a_1 e la Y ha modalità b_1 , e così via:

X \ Y	b_1	b_2	...	b_{r_2}	$N_{i \cdot}$
a_1	N_{11}	N_{12}	...	N_{1r_2}	$N_{1 \cdot} = \sum_{j=1}^{r_2} N_{1j}$
a_2	N_{21}	N_{22}	...	N_{2r_2}	$N_{2 \cdot} = \sum_{j=1}^{r_2} N_{2j}$
...
a_{r_1}	$N_{r_1 1}$	$N_{r_1 2}$...	$N_{r_1 r_2}$	$N_{r_1 \cdot} = \sum_{j=1}^{r_2} N_{r_1 j}$
$N_{\cdot j}$	$N_{\cdot 1} = \sum_{i=1}^{r_1} N_{i1}$	$N_{\cdot 2} = \sum_{i=1}^{r_1} N_{i2}$...	$N_{\cdot r_2} = \sum_{i=1}^{r_1} N_{i r_2}$	n

Nella tabella sono riportate inoltre le somme per righe e per colonne. Ovviamente:

$$\sum_{i=1}^{r_1} \sum_{j=1}^{r_2} N_{ij} = n$$

Possiamo inoltre tradurre l'ipotesi iniziale scrivendola come:

$$H_0: P(X = a_i, Y = b_j) = p_i q_j = P(X = a_i) \cdot P(Y = b_j), \quad \forall i, j$$

Quindi, calcoliamo il numero atteso di coppie (X, Y) su n nelle quali si ha $X = a_i$ e $Y = b_j$ sotto l'ipotesi che H_0 sia vera; tale quantità sarà:

$$n \cdot p_i \cdot q_j$$

Se p_i è incognito, possiamo stimarlo con:

$$\hat{p}_i = \frac{\text{n}^\circ \text{ di } X = a_i}{n} = \frac{N_{i \cdot}}{n}$$

E, allo stesso modo:

$$\hat{q}_j = \frac{\text{n}^\circ \text{ di } Y = b_j}{n} = \frac{N_{\cdot j}}{n}$$

Perciò, il numero di coppie (X, Y) su n che hanno $X = a_i$ e $Y = b_j$ è pari a:

$$n \cdot \hat{p}_i \cdot \hat{q}_j = n \cdot \frac{N_{i \cdot}}{n} \cdot \frac{N_{\cdot j}}{n} = \frac{N_{i \cdot} \cdot N_{\cdot j}}{n}$$

Per valutare la *distanza* tra il modello specificato dalla H_0 per (X, Y) ed i dati, si calcola:

$$ST = \sum_{i=1}^{r_1} \sum_{j=1}^{r_2} \frac{\left(N_{ij} - \frac{N_{i \cdot} \cdot N_{\cdot j}}{n}\right)^2}{\frac{N_{i \cdot} \cdot N_{\cdot j}}{n}}$$

che ricopre il ruolo di statistica test.

Regione critica

Se n è sufficientemente grande, rifiuto H_0 a livello α se

$$ST \geq q(1 - \alpha) = \chi_{r_1 r_2 - 1 - \#parametri\ stimati}^2(1 - \alpha)$$

In particolare, per valutare il numero di parametri stimati, dobbiamo ricordare che in sostanza abbiamo stimato tutte le marginali p_i e q_j , per un totale rispettivamente di r_1 e r_2 parametri; tuttavia, l'ultima marginale può sempre essere ottenuta come differenza tra uno e la somma delle precedenti, quindi:

$$\#parametri\ stimati = r_1 - 1 + r_2 - 1$$

E in conclusione il numero di gradi di libertà sarà:

$$r_1 r_2 - 1 - r_1 + 1 - r_2 + 1 = r_1 r_2 - r_1 + 1 - r_2 = r_1(r_2 - 1) - (r_2 - 1) = (r_1 - 1)(r_2 - 1)$$

Condizione pratica

Affinché il test possa essere eseguito è necessario in pratica che si abbia $n \geq 30$ e che si abbiano almeno 5 osservazioni in ogni classe.

Caso continuo

Nel caso in cui le variabili aleatorie di partenza siano continue, anche in questo caso dobbiamo semplicemente eseguire dei raggruppamenti in intervalli opportunamente scelti, così come nel caso del test χ^2 di buon adattamento.

Osservazioni

Il test appena introdotto è un test non parametrico, perché confronta la statistica test con lo stesso quantile, che dipende solamente dal livello di significatività del test. Questi tipi di test sono anche detti *distribution-free*.

Test di Kolmogorov-Smirnov (buon adattamento)

Il test

Il test di Kolmogorov-Smirnov è un test di buon adattamento:

$$H_0: X \sim F_0 \qquad H_1: X \not\sim F_0$$

Che si esegue partendo da un campione del tipo:

$$X_1, \dots, X_n \text{ i. i. d. } \sim F$$

Si noti che la distribuzione indicata nell'ipotesi nulla deve essere una distribuzione di probabilità continua completamente specificata.

L'idea di base: funzione di ripartizione empirica

Il test si basa sull'uso della *funzione di ripartizione empirica*, ovvero:

$$\hat{F}_n(x) = \frac{n^\circ \text{ di } X_j \leq x}{n} \quad \forall x \in \mathbb{R}$$

La funzione di ripartizione empirica gode delle seguenti proprietà:

1. Siccome la variabile aleatoria:

$$n^\circ \text{ di } X_j \leq x^*$$

È una variabile binomiale con distribuzione binomiale:

$$Bin(n, F(x^*))$$

Abbiamo:

$$E[\hat{F}_n(x^*)] = E\left[\frac{Bin(n, F(x^*))}{n}\right] = \frac{nF(x^*)}{n} = F(x^*)$$

Possiamo quindi affermare che si tratta di uno stimatore puntualmente non distorto.

2. La varianza della funzione di ripartizione campionaria è (per lo stesso motivo):

$$VAR[\hat{F}_n(x^*)] = \frac{nF(x^*)(1 - F(x^*))}{n^2} = \frac{F(x^*)(1 - F(x^*))}{n}$$

Quindi si nota che la varianza tende a 0 per $n \rightarrow +\infty$: si ha di conseguenza anche la proprietà di consistenza in media quadratica.

3. Per il teorema centrale del limite, la legge asintotica di $\hat{F}_n(x^*)$ è;

$$\mathcal{N}\left(F(x^*), \frac{F(x^*)(1 - F(x^*))}{n}\right)$$

4. Definito con D_n il numero:

$$\sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| := D_n$$

Vale il teorema di Glivenko-Cantelli:

$$\lim_{n \rightarrow +\infty} D_n = 0 \text{ con probabilità } 1$$

5. Se si rappresenta il grafico della funzione di ripartizione, si ottiene necessariamente una funzione costante a tratti e monotona crescente, continua da destra, con asintoto a 0 per $x \rightarrow -\infty$ e ad 1 per $x \rightarrow +\infty$. In sostanza quindi il grafico rispetta il tipico andamento di una funzione di ripartizione di variabile aleatoria discreta.

I salti si hanno sempre e solo nei punti che corrispondono ai valori dei dati del campione considerato. Il salto ha ampiezza pari alla frequenza relativa. L'unica informazione che viene persa rappresentando il grafico della funzione di ripartizione rispetto a fornire tutti i dati del campione è l'ordinamento dei dati, che però nel nostro caso non ha alcuna importanza.

Statistica test

La statistica test è:

$$D_n = \sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)|$$

Risultati

Se F_0 è la vera funzione di ripartizione che ha generato i dati, allora D_n non dipende da F_0 ed è quella tabulata nelle tavole dei quantili di Kolmogorov-Smirnov.

Possiamo allora utilizzare la regione critica descrivibile mediante la frase:

$$\text{Rifiuta } H_0 \text{ se } D_n > q_{KS}(1 - \alpha)$$

Osservazione

Per il calcolo pratico di D_n è sufficiente valutare ciò che accade a sinistra e a destra dei punti nei quali si hanno i salti, ovvero calcolare per ogni dato le differenze:

$$|\hat{F}_n(x_i) - F(x_i)|$$

E

$$|\hat{F}_n(x_{i-1}) - F(x_i)|$$

Si individua poi il massimo tra tutti questi valori.