

**LEZIONI DI STATISTICA E
CALCOLO DELLE PROBABILITA'**

UMBERTO MAGAGNOLI

**Materiale per il Corso di lezioni di
“STATISTICA”**

**Laurea magistrale in “Matematica”
Facoltà di Scienze Matematiche, Fisiche e Naturali
Università di Ferrara
Anno accademico 2010-11**

**PARTE TERZA
“Inferenza statistica: stima e verifiche d'ipotesi”**

1. Il ruolo della variabile casuale normale nell'inferenza statistica

Diverse sono le motivazioni dell'impiego preminente che la distribuzione normale svolge nell'ambito degli studi statistici, in particolare nel campo inferenziale. Possiamo limitarci a ricordare: a) teoria degli errori di misura; b) situazioni sperimentali in cui è valido il ricorso della distribuzione normale come approssimazione del modello ipotizzato; c) impiego dei teoremi “del limite centrale” legati alle proprietà asintotiche della distribuzione normale.

2. I teoremi del limite centrale

È disponibile una serie di teoremi detti del “limite centrale” o, forse meglio del “centrale limite” che dimostrano come la distribuzione normale standardizzata possa intendersi distribuzione limite asintotico di particolari successioni di v.c. da cui è possibile impiegare la distribuzione normale come approssimazione degli elementi di tali successioni.

Consideriamo per primo il “teorema di Lindeberg-Lévi” indicato come teorema del limite centrale (in forma debole).

Sia $\{X_n; n = 1, 2, \dots\}$ una successione di v.c. aventi: uguale legge di distribuzione $X_n \sim X$ per $n = 1, 2, \dots$; indipendenti, quindi *i.i.d.*; con l'esistenza dei momenti fino al 2° ordine, quindi con media $M(X_n) = M(X) = \mu$ e varianza $Var(X_n) = Var(X) = \sigma^2$.

Sia $\{Y_n; n = 1, 2, \dots\}$ la successione delle v.c. “somma”: $Y_n = \sum_{i=1}^n X_i$, con $M(Y_n) = n\mu$ e $Var(Y_n) = n\sigma^2$.

Si ottiene la successione di v.c. “somme standardizzate” $\{Z_n; n = 1, 2, \dots\}$ in cui

$$Z_n = \frac{Y_n - M(Y_n)}{[Var(Y_n)]^{1/2}} = \frac{(X_1 + X_2 + \dots + X_n) - n\mu}{\sqrt{n}\sigma}$$

con $M(Z_n) = 0$ e $Var(Z_n) = 1$.

Il *teorema del limite centrale* afferma che: alla successione di v.c. $\{Z_n\}$ corrisponde una successione di F.d.R. $\{F_n(z)\}$ che soddisfa la seguente relazione

$$\lim_{n \rightarrow \infty} F_n(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2/2} dt = \Phi(z) \text{ per } \forall z \in \mathfrak{R}$$

essendo $\Phi(z)$ la F.d.R. di una v.c. “normale standardizzata”, conseguentemente, la v.c. Z_n si distribuisce “asintoticamente” come una v.c. normale standardizzata $Z_n \rightarrow \sim N(0, 1)$, per $n \rightarrow \infty$.

In particolare, quando n è sufficientemente grande si può approssimare la F.d.R. $F_n(z)$ con la $\Phi(z)$. Analogamente si può approssimare la F.d.R. Y_n alla F.d.R. di una v.c. $N(\mu = M(Y_n), \sigma^2 = Var(Y_n))$.

Presentiamo l’enunciato del “teorema di Liapunov” indicato come teorema del limite centrale (in forma forte).

Sia $\{X_n; n = 1, 2, \dots\}$ una successione di v.c. aventi: diversa legge di distribuzione; indipendenti; con l’esistenza dei momenti fino al 2° ordine, quindi con media $M(X_n) = \mu_n$ e varianza $Var(X_n) = \sigma_n^2$; con l’esistenza del momento centrale assoluto del 3° ordine $M(|X_n - \mu_n|^3) = \beta_n$.

Sia $\{Y_n; n = 1, 2, \dots\}$ la successione delle v.c. “somma”: $Y_n = \sum_{i=1}^n X_i$, con $M(Y_n) = \sum_{i=1}^n \mu_i = A_n$, $Var(Y_n) = \sum_{i=1}^n \sigma_i^2 = S_n^2$, inoltre definiamo $B_n = \sqrt[3]{\sum_{i=1}^n \beta_i}$.

Se il limite $\lim_{n \rightarrow \infty} B_n/S_n = 0$, allora la v.c. “somma standardizzata” Z_n , data da

$$Z_n = \frac{Y_n - A_n}{S_n} \text{ con } M(Z_n) = 0 \text{ e } Var(Z_n) = 1$$

definisce una successione di v.c. $\{Y_n; n = 1, 2, \dots\}$ con F.d.R. $\{F_n(z); n = 1, 2, \dots\}$ che soddisfa la relazione

$$\lim_{n \rightarrow \infty} F_n(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2/2} dt = \Phi(z) \text{ per } \forall z \in \mathfrak{R}$$

Che costituisce l'enunciato *teorema del limite centrale*, nella forma forte, affermando che la v.c. Z_n si distribuisce "asintoticamente" come una corrisponde una v.c. "normale standardizzata" $Z_n \rightarrow \sim N(0, 1)$, per $n \rightarrow \infty$.

Se n è sufficientemente grande si può approssimare la *F.d.R.* $F_n(z)$ con la $\Phi(z)$, in assenza di similarità delle distribuzioni di $\{X_n\}$ e fatte salve le condizioni sui momenti. Analogamente si può approssimare la *F.d.R.* Y_n alla *F.d.R.* di una v.c. $N(A_n = M(Y_n), S_n^2 = Var(Y_n))$.

I teoremi del limite centrale, in modo particolare quello di Liapunov, permettono di giustificare l'impiego della distribuzione normale in ambiti diversissimi: fisica, chimica, biologia, astronomia, economia, psicologia, ecc., dove una grandezza di interesse può pensarsi dipendere dal contributo di molteplici fattori casuali aventi diverse leggi distributive e non direttamente osservabili.

Consideriamo alcune situazioni applicative dei teoremi del limite centrale:

1. La media "tipica" secondo Quetelet

Sia Y una caratteristica attinente il prodotto di un processo di lavorazione o di misurazione. Un processo di lavorazione può intendersi come una sequela di operazioni che si ripetono in condizioni oggettive costanti. Ci si attende un unico valore teorico per la caratteristica $Y \equiv \mu$. Dalla constatazione di osservazioni diverse di Y si può imputare la variabilità a cause accidentali che agiscono nelle diverse fasi della lavorazione.

Il risultato osservato è descritto come somma di n v.c. $\{E_i; i = 1, 2, \dots, n\}$

$$\underbrace{Y}_{\text{effetto}} = \underbrace{\mu}_{\text{media tipica}} + \underbrace{(E_1 + E_2 + \dots + E_n)}_{\text{cause accidentali}}$$

Per le v.c. accidentali si può ipotizzare la condizione di mutua indipendenza, media e varianza pari a: $M(E_i) = 0, Var(E_i) = \sigma_i^2$ per $i = 1, 2, \dots, n$. Conseguente che la media e la varianza di Y risultano: $M(Y) = \mu, Var(Y) = \sum_{i=1}^n \sigma_i^2 = \sigma^2$.

Per il teorema di Liapunov, al tendere di $n \rightarrow \infty$, la v.c. standardizzata $Z = (Y - \mu)/\sigma$ tende alla v.c. normale standardizzata $Z \rightarrow \sim N(0,1)$.

Essendo $Y = \mu + \sigma Z$ una combinazione lineare di Z , si distribuisce approssimativamente come una normale con media μ e varianza σ^2

$$Y \rightarrow \sim N(\mu, \sigma^2)$$

con *f.d.d.* di Y approssimabile a

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right\}$$

2. Distribuzione della media campionaria

Consideriamo la successione di n v.c. $\{X_i; i = 1, 2, \dots, n\}$, indipendenti con uguale distribuzione $X_i \sim X$, con media e varianza pari a: $M(X) = \mu$ e $Var(X) = \sigma^2$. Tali condizioni si verificano al livello delle rilevazioni nel caso di un campionamento casuale semplice di numerosità n .

La v.c. media “campionaria” è definita come: $\bar{X} = \sum_{i=1}^n X_i/n$ con media e varianza pari a: $M(\bar{X}) = \mu$ e $Var(\bar{X}) = \sigma^2/n$.

Per il teorema di Lindeberg-Lévi la v.c. standardizzata

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma}$$

converge in distribuzione, quando $n \rightarrow \infty$, alla v.c. normale standardizzata $N(0, 1)$.

Essendo $\bar{X} = \mu + \frac{\sigma}{\sqrt{n}}Z$ una combinazione lineare di Z , si distribuisce approssimativamente come una normale con media μ e varianza σ^2/n

$$\bar{X} \rightarrow \sim N(\mu, \sigma^2/n)$$

con *f.d.d.* di \bar{X} approssimabile a

$$f_Y(y) = \frac{1}{\sqrt{2\pi}(\sigma/\sqrt{n})} \exp\left\{-\frac{1}{2}\left(\frac{y-\mu}{\sigma/\sqrt{n}}\right)^2\right\}$$

per n sufficientemente grande.

3. Funzione di ripartizione di una v.c. binomiale

La v.c. binomiale $Y \sim Bi(n, p)$ per valori di n elevati presenta difficoltà di calcolo della *f.d.p.* e ancor più della *F.d.R.* in quanto possono essere coinvolti fattoriali di difficile valutazione esatta.

Ricordando che la Y può intendersi come somma di n v.c. bernoulliane (binomiali unitarie $X_i \sim Be(p) \equiv Bi(1, p)$ per $i = 1, 2, \dots, n$): $Y = \sum_{i=1}^n X_i$. Le v.c. X_i sono *i.i.d.* con *f.d.p.*

$$f_X(x) = \begin{cases} p^x(1-p)^{1-x} & \text{per } x = 0 \text{ e per } x = 1 \\ 0 & \text{per altrove} \end{cases}$$

La *F.d.R.* di Y risulta data da

$$F_Y(y) = \sum_{i \leq y} \binom{n}{i} p^i (1-p)^{n-i}$$

Introducendo la v.c. standardizzata $Z = \frac{Y-np}{\sqrt{np(1-p)}}$, in quanto $M(Y) = np$ e $Var(Y) = np(1-p)$

Possiamo scrivere l'identità

$$F_Y(y) = P_Y(Y \leq y) = P_Y\left(\frac{Y - np}{\sqrt{np(1-p)}} \leq \frac{y - np}{\sqrt{np(1-p)}}\right)$$

Per il teorema di Lindeberg-Lévi al divergere di n , si può approssimare

$$F_Y(y) \cong \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du = \Phi(z)$$

dove $z = \frac{y - np}{\sqrt{np(1-p)}}$.

Se si vuole la probabilità di valori di Y compresi tra gli interi h e k estremi inclusi, si dovrebbe calcolare, data la definizione della *F.d.R.* di Y , come

$$P_Y(a \leq Y \leq b) = F_Y(b) - F_Y(a - 1) = \sum_{i=a}^b \binom{n}{i} p^i (1-p)^{n-i}$$

Impiegando l'approssimazione normale si può ottenere quanto richiesto dalle tavole della *F.d.R.* normale standardizzata $\Phi(z)$, come

$$P_Y(a \leq Y \leq b) \cong \int_v^w \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du = \Phi(w) - \Phi(v)$$

dove gli estremi di integrazione, per migliorare l'approssimazione sono ottenuti come

$$v = \frac{(a - 0,5) - np}{\sqrt{np(1-p)}}, \quad w = \frac{(b + 0,5) - np}{\sqrt{np(1-p)}}$$

Trattandosi nel caso della Y di una *v.c.* “discreta” mentre *v.c.* la normale è “continua. Si osservi inoltre che l'approssimazione è tanto più migliore quanto più p è prossimo a 0,5 ossia tanto più la Y è una *v.c.* “simmetrica”, come lo è la *v.c.* normale.

Questa versione del teorema del limite centrale, relativa alla approssimazione della binomiale, è ricorda anche come “teorema di DeMoivre-Laplace”.

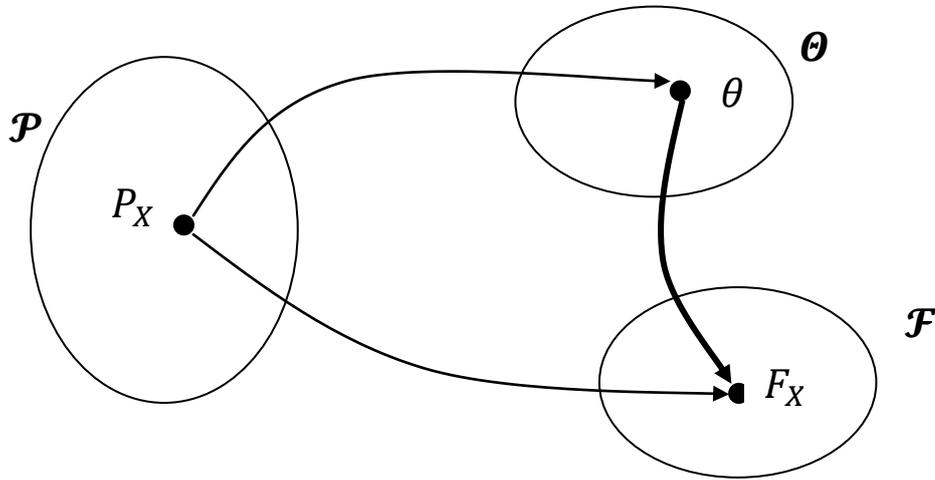
3. Inferenza e induzione statistica

Qualora sia completamente disponibile un modello interpretativo di un fenomeno, nelle sue due componenti strutturale e accidentale, si è in grado di trarre conclusioni sulla realtà, prendere decisioni operative, mettere in discussione la validità del modello stesso mediante la raccolta di ulteriori informazioni.

Per quanto riguarda la componente accidentale interpretata mediante una *v.c.*, unidimensionale o multidimensionale, il disporre della legge di distribuzione della *v.c.* permette di studiare il fenomeno in modo completo avvalendosi della teoria delle probabilità.

Spesso non si dispone della distribuzione nella sua forma completa e si deve quindi ricorrere a scegliere la *F.d.R.* o i valori dei parametri che specificano tale funzione mediante l’osservazione o la rilevazione sperimentale campionaria che, mediante la “teoria inferenziale” o “induzione statistica”, permetta di disporre delle informazioni sulla legge di distribuzione (inferenza non parametrica) o sui parametri che la identificano (inferenza parametrica). Ci si concentra nel seguito allo studio della inferenza parametrica in cui definita una *v.c.* X , e il corrispondente spazio probabilistico $\{\Omega \equiv \mathfrak{R}, \mathcal{B} \equiv \mathcal{B}_1, P_X \in \mathcal{P}\}$ dove \mathcal{P} è una famiglia di distribuzioni (p.es. normali, binomiali, di Poisson, ecc.). Per distinguere ogni singola distribuzione P_X la si caratterizza mediante il “parametro” θ , unidimensionale o multidimensionale, contenuto in Θ , insieme dello spazio parametrico. Potendosi evidenziare l’applicazione $\theta: \mathcal{P} \rightarrow \Theta$ di cui è importante per l’identificazione che tale applicazione sia “biettiva”.

Essendo che a ogni distribuzione di probabilità è associata una *F.d.R.* è possibile fare corrispondere ad ogni elemento di \mathcal{P} una *F.d.R.* $F_X: \mathcal{P} \rightarrow \mathcal{F}$ dove \mathcal{F} è l'insieme di *F.d.R.*

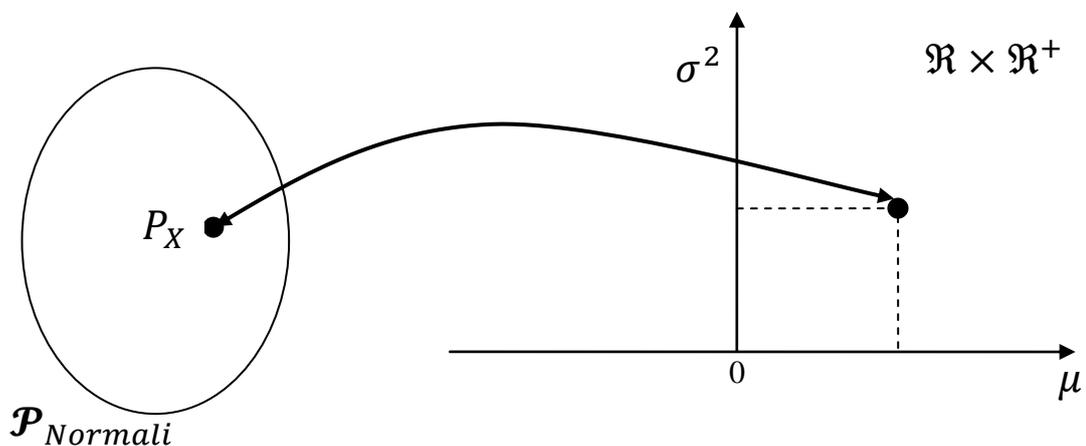


Esempio 1.

Sia $X \sim N(\mu, \sigma^2)$ una v.c. normale, la famiglia di distribuzioni normali è caratterizzata dalla *f.d.d.*

$$f_X(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right\} \quad -\infty < x < \infty$$

Il parametro θ , in questo caso bidimensionale, è $\theta = (\mu, \sigma^2)$ avendosi una applicazione biettiva $\theta: \mathcal{P}_{Normale} \rightarrow (\mathcal{R} \times \mathcal{R}^+)$

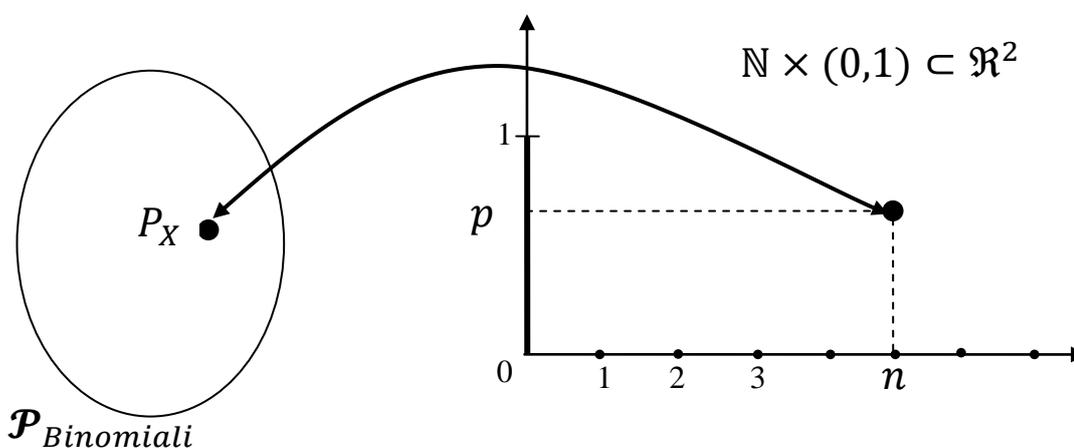


Esempio 2.

Sia $X \sim Bi(n, p)$ una v.c. binomiale, la famiglia di distribuzioni binomiali è caratterizzata dalla f.d.p.

$$f_X(x; n, p) = p_x = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0, 1, \dots, n$$

Il parametro θ , in questo caso bidimensionale, è $\theta = (n, p)$ avendosi una applicazione biettiva $\theta: \mathcal{P}_{Binomiale} \rightarrow (\mathbb{N} \times (0,1))$



Si osservi, nei due precedenti esempi, se si considera come parametro solamente μ o solamente p , rispettivamente, l'applicazione non è "biettiva".

4. Campionamento casuale da una variabile casuale

Consideriamo un esperimento casuale E che genera una v.c. X (per semplicità considerata unidimensionale). L'esperimento si può pensare replicato n volte dando origine a una successione di esperimenti casuali $\{E_1, E_2, \dots, E_n\}$ in cui ciascuna componente genera una corrispondente v.c. appartenente alla successione $\mathbf{X}_n = \{X_1, X_2, \dots, X_n\}$.

Se tutte le componenti sono "somiglianti" cioè presentano la stessa distribuzione di X

$$\{X_1 \sim X, X_2 \sim X, \dots, X_n \sim X\}$$

L'esperimento considerato costituisce un "campionamento casuale" in quanto gli n valori ottenuti si possono considerare derivati dalla stessa distribuzione di X , definita mediante la *F.d.R.* $F_X(x)$

$$F_X(x) = P_X(X \leq x) \quad \forall x \in \mathfrak{R}$$

Per quanto detto ogni manifestazione x_i dell'esperimento è una determinazione, in senso probabilistico, della *v.c.* X e ne costituisce un suo informatore.

5. Il campionamento casuale semplice da una *v.c.* unidimensionale X

Si ha un campionamento "casuale semplice", che si indicherà con *c.s.*, quando le *v.c.* componenti di \mathbf{X}_n sono stocasticamente indipendenti, cioè la *F.d.R.* di \mathbf{X}_n risulta fattorizzabile in termini della *F.d.R.* della *v.c.* X .

$$\begin{aligned} F_{\mathbf{X}}(\mathbf{x}_n) &= F_X(x_1, x_2, \dots, x_n) = F_1(x_1)F_2(x_2) \cdots F_n(x_n) = \\ &= F(x_1)F(x_2) \cdots F(x_n) = \prod_{i=1}^n F(x_i) \end{aligned} \quad (\text{a})$$

La condizione di indipendenza tra le successive osservazioni è tale da non aggiungere informazioni esogene al modo di manifestarsi, in senso probabilistico, della *v.c.* X di partenza.

E' possibile fare alcune osservazioni.

- Ad esempio, per $n = 2$, se il campionamento "non è" *c.s.* tra le due *v.c.* componenti X_1 e X_2 può esistere una correlazione lineare,

misurata da $\rho = C\{X_1, X_2\} / [V\{X_1\}V\{X_2\}]^{1/2} \neq 0$, quindi il valore di x_1 osservato influisce sul modo di verificarsi di x_2 , questo ovviamente non si verifica se X_1 e X_2 sono indipendenti, in tale condizione $\rho = 0$.

- La condizione di indipendenza tra le componenti X_i della v.c. campionaria ($i = 1, 2, \dots, n$) è strettamente legata alle modalità di conduzione dell'esperimento casuale che è all'origine della v.c. \mathbf{X}_n .
- Il fatto che alle componenti di \mathbf{X}_n , X_i per $i = 1, 2, \dots, n$, sia associata la stessa *F.d.R.* $F_i(x) \equiv F(x)$ di X si può pensare, nel caso di estrazione da una popolazione finita, come una replicazione dell'esperimento nelle medesime condizioni ossia con re-immissione. In questo senso il campione può essere inteso come manifestazione replicata della stessa grandezza aleatoria.
- Se la v.c. X è di tipo continuo (più precisamente "assolutamente continuo"), la *F.d.R.* di X può scriversi

$$F(x) = \int_{-\infty}^x f(x) dx$$

con $f(x) \geq 0$ per $\forall x \in \mathfrak{R}$, la "funzione di densità di probabilità" (*f.d.d.*). Allora la *F.d.R.* della v.c. campionaria \mathbf{X}_n è

$$\begin{aligned} F_{\mathbf{X}}(x_1, x_2, \dots, x_n) &= F(x_1)F(x_2) \cdots F(x_n) = \\ &= \int_{-\infty}^{x_1} f(u_1) du_1 \cdot \int_{-\infty}^{x_2} f(u_2) du_2 \cdots \int_{-\infty}^{x_n} f(u_n) du_n = \\ &= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \int_{-\infty}^{x_n} [f(u_1)f(u_2) \cdots f(u_n)] du_1 du_2 \cdots du_n = \\ &= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \int_{-\infty}^{x_n} f_{\mathbf{X}}(u_1, u_2, \dots, u_n) du_1 du_2 \cdots du_n \end{aligned}$$

dove $f_{\mathbf{X}}(x_1, x_2, \dots, x_n)$ è la *f.d.d.* della v.c. multidimensionale \mathbf{X}_n che si esprime in forma fattoriale delle *f.d.d.* delle v.c. componenti

$$f_X(x_1, x_2, \dots, x_n) = f(x_1)f(x_2) \cdots f(x_n) = \prod_{i=1}^n f(x_i) \quad (b)$$

- Se la v.c. X è di tipo discreto, con funzione di probabilità (*f.d.p.*) $p_k = P(X = x_k) > 0$ per $x_k \in \mathfrak{X} \subset \mathfrak{R}$ e $k = 1, 2, \dots$, la *F.d.R.* di X è data da:

$$F(x) = \sum_{x_k \leq x} p_k$$

Nel caso di campionamento *c.s.* la *f.d.p.* della v.c. \mathbf{X}_n si ottiene in forma fattorizzata in termini della *f.d.p.* della v.c. X .

$$p_X(x_1, x_2, \dots, x_n) = p(x_1)p(x_2) \cdots p(x_n) = \prod_{i=1}^n p(x_i) \quad (c)$$

essendo $p_X(x_1, x_2, \dots, x_n)$ la *f.d.p.* della v.c. \mathbf{X}_n dove $x_i \in \mathfrak{X}$ per $\forall i$ e $(x_1, x_2, \dots, x_n) \in \mathfrak{X}^n$.

- Le due relazioni (b) e (c) di fattorizzazione, riguardanti rispettivamente le *f.d.d.* e *f.d.p.*, possono impiegarsi in modo alternativo al posto della relazione (a), corrispondente alla *F.d.R.* della v.c. \mathbf{X}_n , come condizioni valide nel caso di campionamento *c.s.*.

6. Campione casuale semplice, stima e stimatore di un parametro

Qualora la v.c. $\mathbf{X}_n = \{X_1, X_2, \dots, X_n\}$ considerata nel paragrafo precedente, oltre avere le componenti somiglianti sono anche “mutualmente indipendenti”, il campionamento generato dalla replicazione dell’esperimento è detto “campionamento casuale semplice”. Ricordiamo che la condizione di indipendenza è legata alla struttura propria del campionamento. Nel seguito si considererà di provarsi sempre in tali condizioni.

Si indicherà con “campione” una determinazione $\mathbf{x}_n = \{x_1, x_2, \dots, x_n\}$ della v.c. multipla $\mathbf{X}_n = \{X_1, X_2, \dots, X_n\}$. L’inferenza statistica consiste nel basarsi sui risultati ottenuti dal campione per sintetizzare tali valori, ad

esempio con gli strumenti propri della statistica descrittiva, al fine di disporre elementi di valutazione dei parametri del modello probabilistico ipotizzato.

L'esperimento casuale può condurre a manifestarsi di successivi campioni come è rappresentato nella seguente figura

	E_1	E_2	E_n
	↓	↓	↓
x_1	x_{11}	x_{21}	x_{n1}
x_2	x_{12}	x_{22}	x_{n2}
⋮	⋮	⋮			⋮
⋮	⋮	⋮			⋮
x_m	x_{1m}	x_{2m}	x_{nm}
↓	↓	↓			↓
X	X_1	X_2	X_n

1° indice: posizione del valore occupato nel campione;
2° indice: n° d'ordine del campione rilevato.

Al tendere di $m \rightarrow \infty$ ogni colonna dà origine al manifestarsi di una delle v.c. componenti X_1, X_2, \dots, X_n , tutte “somiglianti” a X e “mutualmente indipendenti”. Quando detto è valido sia che X sia una v.c. “discreta” o “continua”, nel rispetto che per ogni componente sia verificata la “legge empirica del caso”.

La *F.d.R.* di X_n risulta, in queste condizioni, fattorizzabile in termini della *F.d.R.* della v.c. X .

$$\begin{aligned}
 F_X(\mathbf{x}_n) &= F_X(x_1, x_2, \dots, x_n) = F_1(x_1)F_2(x_2) \cdots F_n(x_n) = \\
 &= F(x_1)F(x_2) \cdots F(x_n) = \prod_{i=1}^n F(x_i) \quad (a)
 \end{aligned}$$

La condizione di indipendenza tra le successive osservazioni è tale da non aggiungere informazioni esogene al modo di manifestarsi, in senso probabilistico, della v.c. X di partenza.

Sorge l'esigenza di sintetizzare i dati a disposizione per meglio comprendere il fenomeno allo studio, particolarmente se n è elevato, si introducono delle funzioni appartenenti a spazi aventi una dimensione ridotta rispetto a quella originaria del campione. Ad esempio se i valori di X sono contenuti in $\mathfrak{R} \equiv (-\infty, \infty)$, dallo spazio \mathfrak{R}^n si passa allo spazio \mathfrak{R}^r con $r < n$.

Si consideri, ad esempio, la seguente coppia di funzioni di \mathbf{x}_n :

$$\bar{x} = g_1(\mathbf{x}_n) = \sum_{i=1}^n x_i/n, \quad s^* = g_2(\mathbf{x}_n) = \left[\sum_{i=1}^n (x_i - \bar{x})^2/n \right]^{1/2}$$

note, rispettivamente, come “media aritmetica” e “scarto quadratico medio” (*s.q.m*) dei dati campionari. Si sono così ridotti tali dati di numerosità n a due sole dimensioni (rispettivamente un indicatore di locazione e uno di dispersione).

Altre possibilità di sintesi, quali vengono presentate in Statistica descrittiva, sono

- *Il riordinamento dei dati del campione in ordine non decrescente:*

$$\mathbf{x}_{(n)} = g_3(\mathbf{x}_n) = (x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)})$$

dove, in particolare $x_{(1)} = \min \{x_i: i = 1, 2, \dots, n\}$ è il valore minimo delle osservazioni e $x_{(n)} = \max \{x_i: i = 1, 2, \dots, n\}$ è il valore massimo;

- *la Mediana campionaria*

$$\tilde{x} = g_4(\mathbf{x}_n) = \begin{cases} x_{(\frac{n+1}{2})} & \text{se } n \text{ e' dispari} \\ [x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}]/2 & \text{se } n \text{ e' pari} \end{cases}$$

- il *Campo di variazione*

$$r = g_5(\mathbf{x}_n) = x_{(n)} - x_{(1)}$$

- il *Momento di ordine 2 dall'origine*

$$m_2 = g_6(\mathbf{x}_n) = \sum_{i=1}^n x_i^2 / n$$

- la *Differenza media assoluta*

$$\Delta = g_7(\mathbf{x}_n) = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j=1}^n |x_i - x_j| \quad \text{con } x_i \geq 0 \quad \text{per } \forall x_i$$

Ognuna delle funzioni dei dati campionari \mathbf{x}_n può intendersi come una determinazione di una v.c. casuale funzione della v.c. multidimensionale $\mathbf{X}_n = (X_1, X_2, \dots, X_i, \dots, X_n)$ che è detta “informatore statistico” $t = t(\mathbf{X}_n)$ o semplicemente “statistica”.

Si osservi che spesso nelle analisi inferenziali si richiede di considerare al posto dei singoli componenti del vettore dei parametri Θ una qualche funzione degli stessi, che è detta “funzione parametrica” $g: \Theta \rightarrow \Gamma \subset \mathfrak{R}^k$.

Qualora un parametro θ della legge di distribuzione di X sia ignoto si ricorre a una “statistica” $\hat{\theta}$ che chiamiamo “stima”, ottenuta come funzione dei dati campionari

$$\hat{\theta} = \hat{\theta}(\mathbf{x}_n) = \hat{\theta}(x_1, x_2, \dots, x_i, \dots, x_n)$$

Con la quale si cerca di approssimare il parametro ignoto. Se, ad esempio μ (media di X) è il parametro desiderato una sua stima ci è data dalla statistica “media aritmetica del campione”

$$\bar{x} = \hat{\mu}(\mathbf{x}_n) = \frac{1}{n} \sum_{i=1}^n x_i$$

Il valore stima $\hat{\theta} = \hat{\theta}(\mathbf{x}_n)$ è una determinazione della v.c. “stimatore” che risulta funzione della v.c. multidimensionale campionaria \mathbf{X}_n che possiamo definire come

$$\hat{\theta} = \hat{\theta}(\mathbf{x}_n) = \hat{\theta}(X_1, X_2, \dots, X_i, \dots, X_n)$$

Lo stimatore di $\mu = M(X)$, dato dalla “media campionaria”, è la v.c. \bar{X}

$$\bar{X} = \hat{\mu}(\mathbf{X}_n) = \frac{1}{n} \sum_{i=1}^n X_i$$

In particolare se $X \sim N(\mu, \sigma^2)$, trattandosi di un campionamento casuale semplice (c.s.), lo stimatore \bar{X} è una v.c. con legge di distribuzione $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$.

Di fronte ai diversi parametri che definiscono la legge di distribuzione della v.c. X si possono proporre numerose e differenti funzioni dei dati osservati come stime dei corrispondenti parametri ignoti, occorre quindi disporre di criteri di qualificazione degli stimatori e di metodi generali di costruzione delle funzioni di sintesi. Tali aspetti sono argomento dei prossimi due paragrafi.

7. Criteri di qualificazione degli stimatori

Tra i criteri che permettono di qualificare uno stimatore di un parametro θ relativo a una v.c. X e che indirizzano nella scelta operativa, ricordiamo:

- A. Criterio della “consistenza”;
- B. Criterio della “non distorsione” o “correttezza”;
- C. Criterio della “efficienza”;
- D. Criterio della “minima varianza”.

Di ciascuno di tali criteri diamo una definizione e alcune proprietà.

A. Criterio della “consistenza”

Si tratta di una proprietà asintotica: uno stimatore $\hat{\theta}_n = \hat{\theta}(X_n) = \hat{\theta}(X_1, X_2, \dots, X_n)$ è detto “consistente” se al divergere di n converge in probabilità al parametro θ .

In altre parole se, prefissato un valore $\varepsilon > 0$, si ha

$$\lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| < \varepsilon) = 1$$

ossia al crescere di n la probabilità che $\hat{\theta}_n$ cada in un intorno di θ , $(\theta \pm \varepsilon)$, tende a 1 (evento certo), cioè la v.c. $\hat{\theta}_n$ degenera assumendo un valore costante pari al parametro θ .

Esempio 3.

Sia $X \sim N(\mu, \sigma^2)$ con μ parametro ignoto mentre σ^2 sia noto e maggiore di zero.

Consideriamo come stimatore di μ la v.c. “media campionaria” $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$, che si distribuisce come $\bar{X}_n \sim N(\mu, \sigma^2/n)$, essendo $M(\bar{X}_n) = \mu$ e $Var(\bar{X}_n) = \sigma^2/n$.

Al divergere di n abbiamo: $\lim_{n \rightarrow \infty} Var(\bar{X}_n) = \lim_{n \rightarrow \infty} \sigma^2/n = 0$ ossia per $n \rightarrow \infty$ la $\bar{X}_n \sim N(\mu, \sigma^2/n)$ è degenera in quanto ha dispersione nulla e si concentra attorno al valor medio μ che è il parametro da stimare. \bar{X}_n è uno stimatore di μ consistente, come può verificarsi applicando il teorema di Chebyshev $P_{\bar{X}}(|\bar{X}_n - \mu| < h\sigma_{\bar{X}}) \geq 1 - 1/h^2$, in cui ponendo $\varepsilon = h\sigma_{\bar{X}} = h(\sigma/\sqrt{n})$ abbiamo $h = \varepsilon\sqrt{n}/\sigma$ e quindi

$$P_{\bar{X}}(|\bar{X}_n - \mu| < \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2 n}$$

da cui

$$\lim_{n \rightarrow \infty} P_{\bar{X}}(|\bar{X}_n - \mu| < \varepsilon) \geq \lim_{n \rightarrow \infty} \left(1 - \frac{\sigma^2}{\varepsilon^2 n}\right) = 1$$

Essendo il 1° membro una probabilità e quindi (≤ 1) deve verificarsi

$$\lim_{n \rightarrow \infty} P_{\bar{X}}(|\bar{X}_n - \mu| < \varepsilon) = 1$$

Rispettando la condizione di convergenza in probabilità di \bar{X}_n , richiesta dal criterio di “consistenza”.

La proprietà della consistenza di uno stimatore è quella che generalmente è richiesta come condizione irrinunciabile che permette di distinguere una qualsiasi “statistica” da una “stima” di un parametro. La verifica presenta difficoltà analitiche in situazioni complesse dato lo studio della distribuzione dello stimatore al divergere della numerosità campionaria.

B. Criterio della non distorsione

Una proprietà richiesta a uno stimatore $\hat{\theta}_n = \hat{\theta}(X_n) = \hat{\theta}(X_1, X_2, \dots, X_n)$ di un parametro θ , valida per n “finito” è quella della “non distorsione” o “correttezza” che riguarda il valore atteso di $\hat{\theta}_n$.

Uno stimatore $\hat{\theta}_n$ di un parametro θ è detto “non distorto” se la sua media è pari al valore del parametro

$$M(\hat{\theta}_n) = \theta$$

Il criterio trova una sua giustificazione nelle situazioni concrete in cui $\hat{\theta}_n$ tende a distribuirsi con legge normale o gaussiana, per la validità del teorema del “limite centrale”, in quanto si assicura la massima concentrazione di valori di $\hat{\theta}_n$ in un intorno della relativa media.

Esempio 4.

Sia la media μ il parametro da stimare di una v.c. X (come avviene per esempio nel caso di legge normale o di Poisson), consideriamo come

stimatore $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ la “media campionaria”. Sappiamo che $M(\bar{X}_n) = \mu$ per qualunque n , quindi è verificata la proprietà della non distorsione.

In particolare per la legge di distribuzione normale lo stimatore \bar{X}_n di μ è “non distorto” e “consistente”.

Possiamo generalizzare tale risultato a uno stimatore $\hat{\theta}_n$ di un parametro θ che soddisfi le condizioni

$$M(\hat{\theta}_n) = \theta \quad \forall n; \quad \lim_{n \rightarrow \infty} \text{Var}(\hat{\theta}_n) = 0$$

Un tale stimatore, oltre ad essere corretto è anche “consistente” qualunque sia la legge di distribuzione dello stesso. Per la dimostrazione si ricorre alla diseguaglianza di Chebyshev, richiedendosi oltre all’esistenza della media anche a quella della varianza di $\hat{\theta}_n$.

Se di un parametro θ esistono due stimatori corretti in effetti ne esistono infiniti, infatti, dato $M({}_1\hat{\theta}_n) = \theta$ e $M({}_2\hat{\theta}_n) = \theta$, anche ogni combinazione lineare $a {}_1\hat{\theta}_n + b {}_2\hat{\theta}_n$ è uno stimatore corretto $M(a {}_1\hat{\theta}_n + b {}_2\hat{\theta}_n) = \theta$, sorge quindi il problema di come scegliere tra più stimatori corretti. Questo compito è demandato al criterio dell’efficienza.

C. Criterio dell’efficienza

Consideriamo un criterio di confronto tra due o più stimatori di un parametro θ di una v.c. X . Dato un campionamento di numerosità n e siano ${}_1\hat{\theta}_n, {}_2\hat{\theta}_n$ due stimatori “corretti” di θ , è opportuno preferire quello che presenta una minore dispersione. Introduciamo il rapporto di “efficienza” in termini di varianza degli stimatori

$$Re({}_1\hat{\theta}_n, {}_2\hat{\theta}_n) = Re(1,2) = \frac{1/\text{Var}({}_1\hat{\theta}_n)}{1/\text{Var}({}_2\hat{\theta}_n)} = \frac{\text{Var}({}_2\hat{\theta}_n)}{\text{Var}({}_1\hat{\theta}_n)}$$

- Se $Re(1,2) > 1$ ${}_1\hat{\theta}_n$ è più efficiente di ${}_2\hat{\theta}_n$;
- Se $Re(1,2) = 1$ ${}_1\hat{\theta}_n$ è efficiente come ${}_2\hat{\theta}_n$;

- Se $Re(1,2) < 1$ ${}_1\hat{\theta}_n$ è meno efficiente di ${}_2\hat{\theta}_n$.

Esempio 5.

Sia $X \sim N(\mu, \sigma^2)$ due stimatori corretti di μ sono dati da: a) media campionaria \bar{X} ; b) mediana campionaria \tilde{X} . Si dimostra che $Var(\bar{X}) = \sigma^2/n$ e $Var(\tilde{X}) = (1,2533)^2 \sigma^2/n$, il rapporto di efficienza risulta

$$Re(\bar{X}, \tilde{X}) = \frac{1/Var(\bar{X})}{1/Var(\tilde{X})} = \frac{Var(\tilde{X})}{Var(\bar{X})} = (1,2533)^2$$

quindi \bar{X} è più efficiente di \tilde{X} .

D. Criterio della “minima varianza”

Qualora si disponga di una classe di stimatori non distorti, la ricerca dello stimatore “ottimale”, o più efficiente, si riconduce a determinare quello che ha minima varianza. Il teorema seguente, di cui daremo solo l’enunciato, dovuto a Rao-Cramér, definisce il valore minimo che la varianza di uno stimatore non distorto può assumere, valore che costituisce il limiti inferiore con la conseguenza che se esiste uno stimatore che ha tale varianza esso è il più efficiente nella classe degli stimatori non distorti in quanto il rapporto di efficienza di qualunque altro stimatore non può risultare maggiore di uno, ma al limite uguale a uno.

Limite inferiore della varianza di uno stimatore non distorto secondo Rao-Cramér

Sia $\hat{\theta}_n$ uno stimatore non distorto di un parametro θ di una v.c. X , avente f.d.d. o f.d.p. $f_X(x; \theta)$, allora

$$Var(\hat{\theta}_n) \geq \frac{1}{n M_X \left\{ \left[\frac{\partial}{\partial \theta} \ln(f_X(X; \theta)) \right]^2 \right\}}$$

Esempio 6.

Sia $X \sim N(\mu, \sigma^2)$ una v.c. con σ^2 noto e quindi con μ il parametro ignoto. Si consideri come stimatore di μ la media campionaria \bar{X} ottenuta dal campione c.s. $\mathbf{X}_n = (X_1, X_2, \dots, X_n)$, sappiamo che $M(\bar{X}) = \mu$ e $Var(\bar{X}) = \sigma^2/n$ ebbene in questo caso si dimostra che $M \left\{ \left[\frac{\partial}{\partial \theta} \ln(f_X(X; \theta)) \right]^2 \right\} = 1/\sigma^2$ e quindi essendo $Var(\hat{\theta}) \geq \sigma^2/n$ la media campionaria è uno stimatore efficiente di μ .

8. Metodi di ottenimento degli stimatori

Abbiamo visto come sia possibile verificare se un determinato stimatore possieda particolari proprietà quali quelle della: consistenza, non distorsione, efficienza, ecc.. Si richiede anche di disporre metodi che siano in grado di generare stimatori che possiedano alcune delle proprietà richieste mediante l'impiego di una procedura operativa semplice e di generale applicabilità. Uno di questi metodi è quello della “massima verosimiglianza” (*maximum-likelihood*) un altro è quello dei “momenti”.

9. Il metodo di stima della massima verosimiglianza

Sia X una v.c. con distribuzione di probabilità P_θ , e, per non appesantire la simbologia, si consideri che la X sia unidimensionale e il parametro θ uno scalare; si voglia, quindi, avere informazioni su θ partendo dai dati forniti da un campione $\mathbf{x}_n = (x_1, x_2, \dots, x_i, \dots, x_n)$ estratto in modo casuale dalla v.c. X .

La funzione $V(\theta; \mathbf{x}_n)$, intesa come funzione di θ per assegnato campione \mathbf{x}_n , è detta “funzione di verosimiglianza”, ed è data da

$$V(\theta; \mathbf{x}_n) = f_X(\mathbf{x}_n; \theta) = f_X(x_1, \dots, x_i, \dots, x_n; \theta)$$

dove $f_X(\cdot; \cdot)$ è la f.d.d. o f.d.p. della v.c. campionaria \mathbf{X}_n associata al parametro θ . La funzione $V(\theta; \mathbf{x}_n)$ fornisce una “misura dell'evidenza”

che il campione osservato possa provenire da una distribuzione di probabilità con legge nota e specificatamente individuata dal parametro θ .

E' opportuno fare una serie di osservazioni.

- Il termine “verosimiglianza”, in inglese “*likelihood*”, è stato introdotto da R.A. Fisher (1921) in relazione a problemi di stima del coefficiente di correlazione lineare ρ , e spesso viene indicato con $L(\theta; \mathbf{x}_n)$.
- La funzione di verosimiglianza assume valori reali, non negativi e, in particolare, se la v.c. è discreta risulta minore di uno.
- Data la precedente definizione la $V(\theta; \mathbf{x}_n)$ è una funzione di $\theta \in \Theta$, con legge appartenente alla classe \mathcal{P} per un assegnato \mathbf{x}_n .
- Spesso, al posto della funzione di verosimiglianza $V(\theta; \mathbf{x}_n)$, si impiega il suo logaritmo in base “e” $v(\theta; \mathbf{x}_n) = \ln V(\theta; \mathbf{x}_n)$, detta “funzione di log-verosimiglianza”, ed essendo la funzione logaritmo monotona crescente rimangono invariate molte proprietà, quali quelle di ordinamento, massimi e minimi.
- Se il campione è c.s. è possibile fattorizzare la F.d.R. di \mathbf{X}_n e analogamente le relative f.d.d. e f.d.p. si ha

$$V(\theta; \mathbf{x}_n) = f_X(x_1; \theta) f_X(x_2; \theta) \cdots f_X(x_n; \theta) = \prod_{i=1}^n f_X(x_i; \theta)$$

$$v(\theta; \mathbf{x}_n) = \ln V(\theta; \mathbf{x}_n) = \sum_{i=1}^n \ln f_X(x_i; \theta)$$

Molte leggi di distribuzione di frequente impiego presentano una struttura di tipo esponenziale si comprende, quindi, come la funzione di log-verosimiglianza sia preferibile data la sua forma additiva.

- La funzione $V(\theta; \mathbf{x}_n) \equiv f_X(\mathbf{x}_n; \theta)$ può interpretarsi anche come un informatore statistico, applicazione $\mathfrak{R}^n \rightarrow \mathfrak{R}$ di cui può interessare la distribuzione $V(\theta; \mathbf{X}_n) \equiv f_X(\mathbf{X}_n; \theta)$.

Esempio 7.

Si consideri che da una v.c. di Poisson $X \sim Poi(\lambda)$ con parametro ignoto λ , si siano osservati due valori $(x_1 = 1, x_2 = 0)$ come dati da un campionamento c.s..

$$\begin{cases} f_X(x_1 = 1, \lambda) = e^{-\lambda} \frac{\lambda^1}{1!} = e^{-\lambda} \lambda \\ f_X(x_2 = 0, \lambda) = e^{-\lambda} \frac{\lambda^0}{0!} = e^{-\lambda} \end{cases}$$

da cui si ottiene

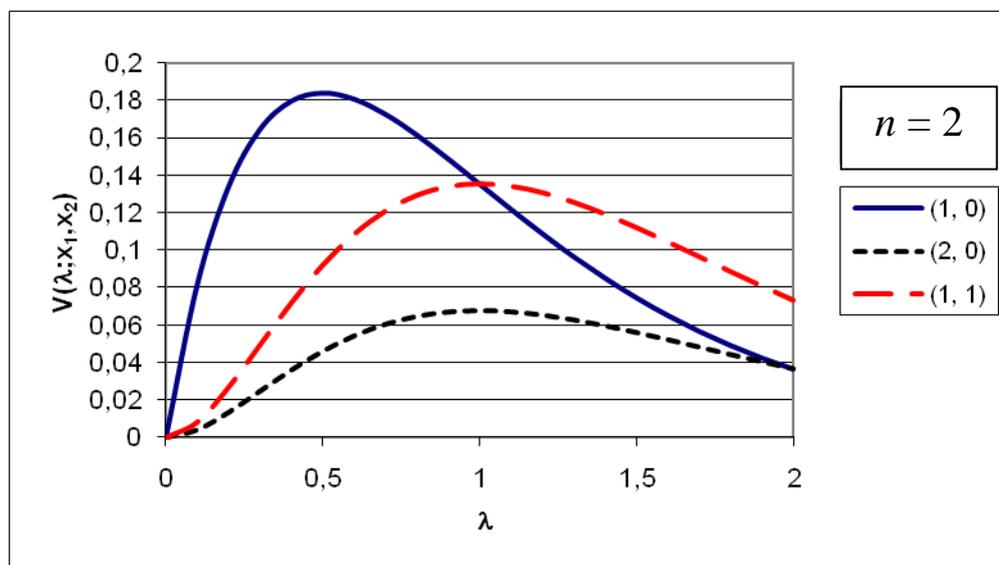
$$V(\lambda; (1, 0)) = e^{-2\lambda} \lambda$$

$$v(\lambda; (1, 0)) = \ln V(\lambda; (1, 0)) = \ln(e^{-2\lambda} \lambda) = -2\lambda + \ln \lambda$$

Il grafico seguente evidenzia l'andamento di $V(\cdot)$ al variare di λ e come possa ritenersi che, dato il campione osservato, sia maggiormente verosimile che λ presenti valori attorno a 0,5. Nella stessa figura si è riportato l'andamento di $V(\cdot)$ anche per

$$(x_1 = 2, x_2 = 0) \Rightarrow V(\lambda; \cdot) = \frac{e^{-2\lambda} \lambda^2}{2}$$

$$(x_1 = 1, x_2 = 1) \Rightarrow V(\lambda; \cdot) = e^{-2\lambda} \lambda^2$$



Possiamo definire la stima di massima verosimiglianza di θ come ogni $\hat{\theta}(\mathbf{x}_n) = \hat{\theta}(x_1, x_2, \dots, x_n)$ tale che

$$V(\hat{\theta}; \mathbf{x}_n) = \sup\{V(\theta; \mathbf{x}_n; \theta \in \Theta)\}$$

La relazione corrispondente in termini di v.c. $\hat{\theta}(\mathbf{X}_n) = \hat{\theta}(X_1, X_2, \dots, X_n)$ si interpreta come stimatore di massima verosimiglianza di θ e quindi funzione della v.c. \mathbf{X}_n . Se la funzione $V(\theta; \mathbf{x}_n)$ è derivabile rispetto a $\theta \in \Theta$, condizione necessaria per ottenere $\hat{\theta}$ è di uguagliare a zero la derivata prima di $V(\theta; \mathbf{x}_n)$ rispetto a θ .

Esempio 8.

Consideriamo come modello lo schema di “prove ripetute” in cui di considera come v.c. X il numero di replicazioni indipendenti, effettuate prima che si manifesti l’evento “A” a cui è associata una probabilità costante pari a p . La v.c. X è di tipo discreto ed è detta “geometrica o di Pascal” $X \sim Ge(p)$, e presenta la seguente f.d.p. con $0 < p < 1$, parametro ignoto

$$f(x; p) = (1 - p)^{x-1}p \quad \text{per } x = 1, 2, \dots$$

Consideriamo che si sia effettuata una sola osservazione di X ottenendo $x_1 = x$ e si voglia stimare p mediante il metodo della verosimiglianza.

La funzione di verosimiglianza è

$$V(p; x) = f(x; p) = (1 - p)^{x-1}p$$

Quella di log-verosimiglianza

$$v(p; x) = \ln V(p; x) = (x - 1) \ln(1 - p) + \ln p$$

Per individuare la stima di p deriviamo $V(p; x)$ (o $v(p; x)$)

$$\frac{d}{dp} V(p; x) = \frac{d}{dp} [(1 - p)^{x-1}p] = (1 - p)^{x-2}(1 - px)$$

da cui, essendo $0 < p < 1$ e quindi $(1 - p) > 0$

$$\frac{d}{dp} V(p; x) = 0 \implies (1 - px) = 0 \implies \hat{p} = \frac{1}{x}$$

A cui corrisponde la *v.c.* stimatore di massima verosimiglianza

$$\hat{p} = 1/X$$

La derivata seconda di $V(p; x)$

$$\begin{aligned} \frac{d^2}{dp^2} V(p; x) &= \frac{d}{dp} [(1 - p)^{x-2} (1 - px)] \\ &= (1 - p)^{x-3} [(1 - x)(1 - px) - (1 - p)x] \end{aligned}$$

che per $p = 1/x$ risulta

$$\left[\frac{d^2}{dp^2} V(p; x) \right]_{p=1/x} = -x \left(1 - \frac{1}{x} \right)^{x-3} < 0$$

confermando che si tratta di un punto di massimo.

Esempio 9.

Sia X una *v.c.* normale con media nota $\mu = 0$ e varianza σ^2 da stimare con il metodo della massima verosimiglianza. Si consideri di disporre un campione c.s. $\mathbf{x}_n = (x_1, x_2, \dots, x_n)$ di numerosità n .

La *f.d.d.* di X risulta $f(x; \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2} \frac{x^2}{\sigma^2}\right\}$ e la funzione di verosimiglianza, è data da

$$V(\sigma^2; \mathbf{x}_n) = f_{\mathbf{X}}(\sigma^2; \theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2} \frac{x_i^2}{\sigma^2}\right\}$$

e la funzione di log-verosimiglianza

$$\begin{aligned}
v(\sigma^2; \mathbf{x}_n) &= \ln V(\sigma^2; \mathbf{x}_n) \\
&= \sum_{i=1}^n \left[-\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln(\sigma^2) - \frac{1}{2} \frac{x_i^2}{\sigma^2} \right] = -\frac{n}{2} \ln(2\pi) \\
&\quad - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2} (\sigma^2)^{-1} \sum_{i=1}^n x_i^2
\end{aligned}$$

Derivando $v(\sigma^2; \mathbf{x}_n)$ rispetto a σ^2 , abbiamo

$$\frac{d}{d\sigma^2} v(\sigma^2; \mathbf{x}_n) = -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2} \frac{1}{(\sigma^2)^2} \sum_{i=1}^n x_i^2 = \frac{1}{2\sigma^2} \left[-n + \frac{1}{\sigma^2} \sum_{i=1}^n x_i^2 \right]$$

da cui

$$\frac{d}{d\sigma^2} v(\sigma^2; \mathbf{x}_n) = 0 \implies -n + \frac{1}{\sigma^2} \sum_{i=1}^n x_i^2 = 0 \implies \hat{\sigma}^2 = \frac{\sum_{i=1}^n x_i^2}{n}$$

In cui l'espressione al secondo membro può interpretarsi come varianza campionaria dei dati osservati essendo la media del modello assunta pari a zero. Anche in questo esempio, come nel precedente, il punto notevole si dimostra che è di massimo assoluto.

Alla stima del parametro σ^2 corrisponde una v.c. stimatore di massima verosimiglianza data da

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n X_i^2}{n}$$

Osservazioni

Gli stimatori di massima verosimiglianza, sotto particolari condizioni di regolarità della funzione di verosimiglianza e di esistenza delle derivate parziali fino al terzo ordine, condizioni che spesso sono verificate nei modelli probabilistici più frequentemente impiegati, godono di proprietà tra le quali:

- La soluzione del sistema di equazioni di verosimiglianza esiste, con probabilità uno, per $n \rightarrow \infty$, e converge all'insieme dei parametri ignoti.
- Se la soluzione è unica, per $n > n_0$, lo stimatore è “consistente”.
- Per uno stimatore unidimensionale $\hat{\theta}$ consistente la distribuzione di $\hat{\theta}(\mathbf{X}_n)$ tende asintoticamente alla legge normale, risultando inoltre asintoticamente efficiente (massimamente efficiente)

$$\sqrt{n}(\hat{\theta}(\mathbf{X}_n) - \theta) \rightarrow N(0, 1)$$

- In generale si tratta di uno stimatore “distorto”.

Per queste proprietà e loro generalizzazioni, di veda Lehmann (1983) “*Theory of Point Estimation*” pg. 409 e seg..

10. Il metodo di stima dei momenti

Abbiamo visto come molti dei parametri che definiscono le distribuzioni delle v.c. che abbiamo considerato nella Parte II del testo coincidono o sono strettamente legati ai momenti dall'origine o centrali della v.c. considerata.

Questo è il caso della:

- v.c. “bernulliana” $X \sim Be(p)$. In cui $\mu = M(X) = p$;
- v.c. “binomiale” $Y \sim Bi(n, p)$. In cui $\mu = M(X) = np$, che se n è noto p è proporzionale a μ ;
- v.c. di Poisson $W \sim Po(\lambda)$. In cui $\mu = M(X) = \lambda$
- v.c. “esponenziale negativa” $X \sim Ex(\theta)$. In cui $\mu = M(X) = \theta$;
- v.c. “gaussiana o normale” $X \sim N(\mu_X, \sigma_X^2)$. In cui $\mu = M(X) = \mu_X$ e $\sigma^2 = Var(X) = \sigma_X^2$.
- Se consideriamo la v.c. “geometrica o di Pascal” $X \sim Ge(p)$, presentata nell'esercizio 8., abbiamo, posto $q = 1 - p$

$$\begin{aligned}
M(X) &= \sum_{x=1}^{\infty} x(1-p)^{x-1}p = p \sum_{x=1}^{\infty} xq^{x-1} = p \sum_{x=1}^{\infty} \frac{d}{dq} q^x \\
&= p \frac{d}{dq} \left[\sum_{x=1}^{\infty} q^x \right] = p \frac{d}{dq} \left\{ q \left[\sum_{x=0}^{\infty} q^x \right] \right\} \\
&= p \frac{d}{dq} \left\{ q \left[\lim_{x \rightarrow \infty} \frac{1 - q^{x+1}}{1 - q} \right] \right\} = p \frac{d}{dq} [q(1 - q)^{-1}] \\
&= p [q(1 - q)^{-2} + (1 - q)^{-1}] = p \left[\frac{q}{p^2} + \frac{1}{p} \right] = p \frac{q + p}{p^2} = \frac{1}{p}
\end{aligned}$$

La media di X è pari a $\mu = M(X) = 1/p$.

Qualora si consideri un campione *c.s.* $\mathbf{x}_n = (x_1, x_2, \dots, x_n)$ analogamente a quanto fatto in statistica descrittiva, possiamo definire la media aritmetica (media campionaria) e i relativi momenti dall'origine e centrali (campionari)

$$m = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i; \quad m_r = \frac{1}{n} \sum_{i=1}^n x_i^r; \quad \bar{m}_r = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^r$$

Per $r = 0, 1, 2, \dots$. Tali grandezze si possono ritenere determinazioni di corrispondenti *v.c.* campionarie con distribuzione dipendente da quella della *v.c.* X ; per non appesantire la simbologia per le *v.c.* momenti si è utilizzato lo stesso indicatore della determinazione.

Le principali proprietà delle *v.c.* momenti campionari sono:

A. I momenti campionari dall'origine sono stimatori "corretti" dei corrispondenti momenti della *v.c.* X :

$$M(m_r) = \mu_r = M(X^r) \text{ da cui } M(\bar{X}) = \mu$$

B. I momenti campionari centrali sono stimatori "distorti" dei corrispondenti momenti della *v.c.* X :

$$M(\bar{m}_r) \neq \bar{\mu}_r = M((X - \mu)^r) \text{ da cui } M(S'^2) \neq \sigma^2$$

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

C. Si dimostra che i momenti campionari sono stimatori “consistenti” dei relativi momenti della v.c. X .

Si lascia al lettore di verificare o dimostrare le proprietà A. e B. che sono ottenibili da quelle generali dell'operatore media.

Dalle proprietà dei momenti campionari K. Pearson (1857-1936) ha proposto un metodo per determinare “buoni” stimatori dei parametri dei modelli distributivi denominato appunto “metodo dei momenti”.

Sia $\mathbf{X}_n = (X_1, X_2, \dots, X_n)$ un campione c.s. con componenti aventi *F.d.R.* $F_X(x; \boldsymbol{\theta})$ dove $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$ è un vettore parametrico k dimensionale, il metodo dei momenti ottiene la stima di $\boldsymbol{\theta}$ come soluzione di un sistema di k equazioni in k incognite ottenute dall'uguaglianza dei momenti dall'origine della v.c. X , espressi in funzione dei parametri, con i corrispondenti momenti campionari. Sia $\mu_j(\boldsymbol{\theta}) = M(X^j)$ e $m_j = [\sum_{i=1}^n x_i^j]/n$ per $j = 1, 2, \dots, k$, abbiamo

$$m_j = \mu_j(\boldsymbol{\theta}) \text{ per } j = 1, 2, \dots, k$$

Esempio 10.

Sia $X \sim N(\mu, \sigma^2)$ e $\mathbf{x}_n = (x_1, x_2, \dots, x_n)$ un campione c.s. con $n \geq 2$, per determinare le stime di entrambi i parametri μ e σ^2 mediante il metodo dei momenti si deve risolvere il seguente sistema

$$\begin{cases} m_1 = \mu_1 \\ m_2 = \mu_2 \end{cases} \begin{cases} [\sum_{i=1}^n x_i]/n = \mu \\ [\sum_{i=1}^n x_i^2]/n = \sigma^2 + \mu^2 \end{cases} \begin{cases} \hat{\mu} = \bar{x} \\ \hat{\sigma}^2 = \frac{[\sum_{i=1}^n (x_i - \bar{x})^2]}{n} \end{cases} \begin{cases} \hat{\mu} = \bar{x} \\ \hat{\sigma}^2 = s'^2 \end{cases}$$

Si osservi che lo stimatore S'^2 di σ^2 risulta “distorto”, infatti

$$\begin{aligned}
M(S'^2) &= M\left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right) \\
&= \frac{1}{n} \sum_{i=1}^n M([(X_i - \mu) - (\bar{X} - \mu)]^2) = \frac{1}{n} \sum M((X_i - \mu)^2) \\
&\quad + \frac{1}{n} \sum M((\bar{X} - \mu)^2) - \frac{2}{n} \sum M((X_i - \mu)(\bar{X} - \mu))
\end{aligned}$$

Al secondo membro abbiamo:

- 1° termine $\frac{1}{n} \sum M((X_i - \mu)^2) = \frac{n}{n} \sigma^2$ essendo $M((X - \mu)^2) = \text{Var}(X) = \sigma^2$;
- 2° termine $\frac{1}{n} \sum M((\bar{X} - \mu)^2) = \frac{n}{n} M((\bar{X} - \mu)^2) = \frac{\sigma^2}{n}$ essendo $M((\bar{X} - \mu)^2) = \text{Var}(\bar{X}) = \frac{\sigma^2}{n}$;
- 3° termine

$$\begin{aligned}
\frac{2}{n} \sum M((X_i - \mu)(\bar{X} - \mu)) &= \frac{2}{n} \sum M\left((X_i - \mu) \left(\frac{X_1 + X_2 + \dots + X_n}{n} - \mu\right)\right) = \\
&= \frac{2}{n^2} \sum M((X_i - \mu)[(X_1 - \mu) + (X_2 - \mu) + \dots + (X_n - \mu)]).
\end{aligned}$$

L'espressione entro la sommatoria è pari a $M((X_i - \mu)^2) = \sigma^2$ in quanto tutti gli altri addendi sono nulli essendo covarianze di v.c. indipendenti. Il 3° termine si riduce a: $\frac{2}{n^2} n \sigma^2 = 2 \frac{\sigma^2}{n}$.

Complessivamente abbiamo

$$M(S'^2) = \sigma^2 + \frac{\sigma^2}{n} - 2 \frac{\sigma^2}{n} = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n} \sigma^2 \neq \sigma^2$$

Lo stimatore distorto S'^2 può rendersi corretto mediante un fattore, infatti se si considera

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

abbiamo

$$M(S^2) = M\left(\frac{n}{n-1}S'^2\right) = \frac{n}{n-1}M(S'^2) = \frac{n}{n-1}\frac{n-1}{n}\sigma^2 = \sigma^2$$

Al crescere di n abbiamo $\frac{n-1}{n} \rightarrow 1$ e quindi i due stimatori si identificano e gli stimatori S'^2 e S^2 sono entrambi "consistenti". Questo risultato ha una valenza generale per quanto riguarda la varianza di una v.c. di cui non sia nota la media e quindi è valido non solo per la distribuzione normale.

Qualora si desideri uno stimatore dello s.q.m. σ si può assumere

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

si tratta di uno stimatore "distorto" ma "consistente".

Si osservi che $(n-1)$ sono chiamati "gradi di libertà" (*g.d.l.*) in quanto esprime il numero di "scarti" $(X_i - \bar{X})$ indipendenti funzionalmente dovendosi verificare la condizione $\sum_{i=1}^n (X_i - \bar{X}) = 0$, da cui deriva che uno scarto è funzione lineare degli altri $(n-1)$: ad es. $(X_n - \bar{X}) = -\sum_{i=1}^{n-1} (X_i - \bar{X})$, situazione che esprimiamo come "perdita di un grado di libertà".

11. La stima dei parametri di variabili casuali notevoli

Ci limitiamo a esporre i problemi di stima dei parametri delle v.c., ottenute con i metodi della massima verosimiglianza e dei momenti, della "bernoulliana o binomiale unitaria" $X \sim Be(p) = B(n=1, p)$ e della "normale o gaussiana" $X \sim N(\mu, \sigma^2)$, dato il particolare ruolo che tali distribuzioni hanno in statistica. Consideriamo sempre di disporre di un campione c.s. di numerosità n $\mathbf{x}_n = (x_1, x_2, \dots, x_n)$ dal quale partire per fare inferenza.

A. v.c. "bernoulliana" $X \sim Be(p)$

Il parametro p è la probabilità dell'evento considerato A : $P(A) = p$ ($0 < p < 1$); la *f.d.p.* di X è: $f(x) = p^x q^{1-x}$, con $q = 1 - p$ e $x = 0, 1$.

La funzione di verosimiglianza risulta

$$V(p; x_1, x_2, \dots, x_n) = p^{\sum_{i=1}^n x_i} (1 - p)^{n - \sum_{i=1}^n x_i}$$

da cui $v(p; x_1, x_2, \dots, x_n) = (\sum_{i=1}^n x_i) \ln p + (n - \sum_{i=1}^n x_i) \ln(1 - p)$
e

$$\frac{dv}{dp} = \frac{(\sum_{i=1}^n x_i)}{p} - \frac{(n - \sum_{i=1}^n x_i)}{1 - p} = 0$$

$$(1 - p) \left(\sum_{i=1}^n x_i \right) = p \left(n - \sum_{i=1}^n x_i \right) \Rightarrow \left(\sum_{i=1}^n x_i \right) = np$$

ottenendo

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

Come stima di massima verosimiglianza si ottiene la media campionaria che è anche la stima ottenuta con il metodo dei momenti.

Per le proprietà dello stimatore media campionaria abbiamo: \hat{p} è uno stimatore consistente e corretto, la sua distribuzione è legata alla binomiale infatti: $\hat{p} \sim Y/n$ dove $Y = \sum_{i=1}^n X_i \sim Bi(n, p)$, da cui $M(\hat{p}) = p$, $Var(\hat{p}) = p(1 - p)/n$ e la *f.d.p.* è

$$f_{\hat{p}}(u) = \binom{n}{y} p^y (1 - p)^{n-y} \quad \text{per } u = \frac{y}{n} \text{ e } y = 0, 1, 2, \dots, n$$

Al divergere di n , la distribuzione di \hat{p} si approssima alla legge normale $\hat{p} \cong N\left(p, \frac{p(1-p)}{n}\right)$, per il teorema del limite centrale.

B. v.c. "normale" $X \sim N(\mu, \sigma^2)$

Qualora sia ignoto il parametro valor medio μ di X la stima ottenuta sia con il metodo della massima verosimiglianza sia con il metodo dei momenti coincide con la "media campionaria" $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ e lo stimatore \bar{X} corrispondente è: "consistente", "non distorto", e "massimamente efficiente". Inoltre \bar{X} si distribuisce con legge normale $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ essendo $M(\bar{X}) = \mu$ e $Var(\bar{X}) = \sigma^2/n$.

Qualora sia ignoto il solo parametro varianza σ^2 di X la stima ottenuta sia con il metodo della massima verosimiglianza sia con il metodo dei momenti coincide con la "media dei quadrati degli scarti dei dati rilevati dalla media μ " $s''^2 = \sum_{i=1}^n (x_i - \mu)^2/n$ e lo stimatore S''^2 corrispondente è: "consistente", "non distorto". Si distribuisce con legge collegata alla v.c. Chi-quadrato, infatti

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

Essendo la v.c. $Z_i = \left(\frac{X_i - \mu}{\sigma} \right)$ distribuita come una normale standardizza $Z_i \sim N(0,1)$ e le n v.c. Z_i sono mutualmente indipendenti, trattandosi di campionamento c.s., abbiamo, v. Parte II parag. 26.7.

$$S''^2 = \frac{\sigma^2}{n} \sum_{i=1}^n Z_i^2 \sim \sigma^2 \frac{\chi^2(n)}{n}$$

Con valor medio e varianza pari a: $M(S''^2) = \sigma^2$, $Var(S''^2) = 2\sigma^4/n$. Ne consegue che lo stimatore è "corretto" e "consistente".

Se μ è ignoto come stima di massima verosimiglianza e dei momenti si ricorre alla "varianza campionaria"

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

dove al posto di μ , nella relazione che dà S'^2 , di è posto lo stimatore \bar{X} . Tale stimatore S'^2 , come abbiamo visto nel parag. precedente, è “consistente” ma “distorto” e si dimostra che si distribuisce

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

con media $M(S'^2) = \frac{n-1}{n} \sigma^2$ e varianza $Var(S'^2) = 2\sigma^4(n-1)/n^2$.

Come stimatore “corretto” di σ^2 si ricorre a

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

che si distribuisce come

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

con media $M(S^2) = \sigma^2$ e varianza $Var(S^2) = 2\sigma^4/(n-1)$.

12. Statistica e stimatore sufficiente

Data la v.c. campionaria $\mathbf{X}_n = (X_1, X_2, \dots, X_n)$ definita in corrispondenza dello spazio probabilistico $\{\mathfrak{R}^n, \mathcal{B}_n, \mathcal{P}_\theta\}$ dove \mathcal{P}_θ è la distribuzione di probabilità individuata dal parametro θ , (unidimensionale o multidimensionale), indichiamo con $\mathbf{t}_r = \mathbf{t}(x_1, x_2, \dots, x_n)$ la trasformazione $\mathbf{t}_r: \mathfrak{R}^n \rightarrow \mathfrak{R}^r$ che individua la “statistica” (informatore statistico) \mathbf{t}_r funzione delle osservazioni che può impiegarsi come stima del parametro θ ritenuto ignoto. Diremo che \mathbf{t}_r è una “statistica sufficiente” per θ se per ogni insieme $B \in \mathcal{B}_n$ si ha

$$P_\theta(B|\mathbf{t}_r) \equiv P(B|\mathbf{t}_r) \quad \forall \theta \quad (q.c. = \text{quasi certamente})$$

Condizione che esprime la probabilità condizionata $P_\theta(B|\mathbf{t}_r)$ come “indipendente” dal valore del parametro θ .

Una statistica (e quindi anche una stima) risulta migliore a un'altra che non ha tale proprietà in quanto

L'interesse per una statistica sufficiente, nell'ambito dei problemi di inferenza parametrica, consiste nel fatto che essa è migliore, rispetto a uno che non gode di tale proprietà, in quanto tale statistica contiene tutto quanto è disponibile di “informazione” nel campione relativamente a θ . Si può dire che, al fine di determinare l'influenza di θ sul campione \mathbf{x}_n è sufficiente conoscere l'influenza di θ sull'informatore \mathbf{t}_r . Convien quindi studiare le proprietà della v.c. $\mathbf{T} = \mathbf{t}(\mathbf{X}_n)$ al posto di quelle di \mathbf{X}_n , in quanto l'informatore risulta più “parsimonioso” rispetto all'intero campione essendo $r < n$.

Esempio 11.

Sia X una v.c. discreta appartenente alla famiglia di distribuzioni di Poisson con parametro λ ignoto.

$$X \sim Poi(\lambda) \quad \text{con } f.d.p. \quad f_X(x; \lambda) = e^{-\lambda} \frac{\lambda^x}{x!} \quad \text{per } x = 1, 2, \dots$$

Si consideri un campione c.s. di numerosità n : \mathbf{x}_n e l'informatore statistico “somma” $T = \sum_{i=1}^n x_i$. La f.d.p. della v.c. $\mathbf{X}_n|(T = t')$, cioè condizionata a un particolare valore di $t = t'$, si ottiene mediante la relazione

$$f(\mathbf{x}_n|t'; \lambda) = \frac{f_{\mathbf{X}}(\mathbf{x}_n; \lambda)}{f_T(t = t'; \lambda)} \quad \text{per } \forall \mathbf{x}_n \in t^{-1}(t') \quad (**)$$

La f.d.p. presente al numeratore è ottenuta dall'espressione generale congiunta, ricordando che i valori di \mathbf{x}_n debbono rispettare la condizione $\sum_{i=1}^n x_i = t'$.

$$\begin{aligned}
 f_{\mathbf{X}}(\mathbf{x}_n; \lambda) &= \prod_{i=1}^n e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda} \frac{\lambda^{\sum x_i}}{\prod x_i!} \\
 &= e^{-n\lambda} \frac{\lambda^{t'}}{x_1! x_2! \cdots x_n!} \quad \text{per } \sum_{i=1}^n x_i = t'
 \end{aligned}$$

Per ottenere la *f.d.p.* presente al denominatore, ossia la *f.d.p.* della v.c. T , si sommano tutti i termini della $f_{\mathbf{X}}(\mathbf{x}_n; \lambda)$ relativi allo spazio campionario $\mathbf{x}_n = t^{-1}(t') \in \mathfrak{R}^n$, cioè tali che $\sum_{i=1}^n x_i = t'$

$$f_T(t'; \lambda) = \sum_{\mathbf{x}_n: \sum x_i = t'} f_{\mathbf{X}}(\mathbf{x}_n; \lambda) = e^{-n\lambda} \lambda^{t'} \sum_{\mathbf{x}_n: \sum x_i = t'} \frac{1}{x_1! x_2! \cdots x_n!}$$

moltiplicando e dividendo per $t'!$ l'ultima espressione, si ottiene

$$\begin{aligned}
 f_T(t'; \lambda) &= \frac{e^{-n\lambda} \lambda^{t'}}{t'!} \sum_{\mathbf{x}_n: \sum x_i = t'} \frac{t'!}{x_1! x_2! \cdots x_n!} \\
 &= \frac{e^{-n\lambda} \lambda^{t'}}{t'!} \sum_{\mathbf{x}_n: \sum x_i = t'} \binom{t'}{x_1, x_2, \dots, x_n}
 \end{aligned}$$

dove $\binom{t'}{x_1, x_2, \dots, x_n}$ è il coefficiente multinomiale, la cui somma, estesa ai valori di x_i tali che $\sum_{i=1}^n x_i = t'$ si dimostra che vale $\left(\underbrace{1, 1, \dots, 1}_{n \text{ volte}}\right)^{t'} = n^{t'}$.

Si tratta di un'espressione generalizzata dello sviluppo di potenza del binomio di Newton.

Si ottiene così la *f.d.p.* della v.c. informatore T che si distribuisce con legge di Poisson $X \sim Poi(n\lambda)$

$$f_T(t'; \lambda) = e^{-n\lambda} \frac{(n\lambda)^{t'}}{t'!}$$

e si ottiene la distribuzione di $\mathbf{X}_n | (T = t')$ come:

$$f(\mathbf{x}_n | T = t'; \lambda) = \frac{f_{\mathbf{X}}(\mathbf{x}_n; \lambda)}{f_T(t = t'; \lambda)} = \frac{e^{-n\lambda} \frac{\lambda^{t'}}{x_1! x_2! \cdots x_n!}}{e^{-n\lambda} \frac{(n\lambda)^{t'}}{t'!}} = \frac{1}{n^{t'}} \frac{t'!}{x_1! x_2! \cdots x_n!}$$

$$= \frac{1}{n^{t'}} \binom{t'}{x_1, x_2, \dots, x_n}$$

espressione che “non dipende” dal parametro λ . Si osservi che la distribuzione di $\mathbf{X}_n | t'$ è una legge multinomiale $\mathbf{X}_n | t' \sim \text{Mun}(\mathbf{p}, n)$ con $\mathbf{p} = \left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)$. Si ricordi, poi, che la generica v.c. k - multinomiale $\mathbf{Y}_k \sim \text{Mun}(\mathbf{p}_k, m)$ ha la seguente *f.d.p.*:

$$f_{\mathbf{Y}}(\mathbf{y}; \mathbf{p}, m) = \binom{m}{y_1, y_2, \dots, y_k} p_1^{y_1} p_2^{y_2} \cdots p_k^{y_k}$$

con

$$y_h \in \{0, 1, \dots, m\} \text{ per } h = 1, 2, \dots, k \text{ e } \sum_{h=1}^m y_h = m$$

$$p_h \in (0, 1) \text{ per } h = 1, 2, \dots, k \text{ e } \sum_{h=1}^m p_h = 1$$

che per $m = t'$, $k = n$, $\mathbf{p} = \left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)$, $\mathbf{y} = (y_1, y_1, \dots, y_k) = \mathbf{x}_n$, risulta

$$\binom{t'}{x_1, x_2, \dots, x_n} \left(\frac{1}{n}\right)^{x_1} \left(\frac{1}{n}\right)^{x_2} \cdots \left(\frac{1}{n}\right)^{x_n}$$

$$= \frac{1}{n^{t'}} \frac{t'!}{x_1! x_2! \cdots x_n!} \text{ con } x_1 + x_2 + \cdots + x_n = t'$$

coincidente con la *f.d.p.* $\mathbf{X}_n | t'$.

Si è verificato, dunque, che la statistica T , somma delle osservazioni, è un informatore sufficiente per il parametro λ della v.c. di Poisson. Ovviamente T non ha le proprietà richieste agli stimatori di λ , consistenza, non distorsione, ecc., ma da T è immediato ottenere uno stimatore, che

pure è sufficiente, e gode di tali proprietà infatti la media campionaria \bar{X} , ottenendosi come $\bar{X} = T/n$ mediante una trasformazione di scala di T gode delle proprietà proprie degli stimatori.

La condizione di “sufficienza” di un informatore statistico T (o, in particolare stimatore) data nell’esempio 11. – equazione (***) – può scriversi in forma di prodotto delle *f.d.p.* o *f.d.d.* come, essendo θ il generico parametro che caratterizza la *v.c.* X

$$f_X(\mathbf{x}_n; \theta) = f(\mathbf{x}_n|t)f_T(t; \theta)$$

Jerzy Neyman (1894-1981) (uno dei “padri” della Statistica Inferenziale, insieme a Ronald Aylmer Fisher (1890-1962)), ha dato un teorema che generalizza la precedente relazione.

Teorema – Sia $\{\mathfrak{R}^n, \mathcal{B}_n, \mathcal{P}_\theta\}$ lo spazio probabilistico relativo alla *v.c.* \mathbf{X}_n e sia $f_X(\mathbf{x}_n; \theta)$ la sua *f.d.p.* o *f.d.d.i.*, condizione “necessaria e sufficiente” perché un informatore statistico $\mathbf{t}: \mathfrak{R}^n \rightarrow \mathfrak{R}^r$ sia “un informatore sufficiente” è che $f_X(\mathbf{x}_n; \theta)$ possa fattorizzarsi nel seguente modo

$$f_X(\mathbf{x}_n; \theta) = k(\mathbf{x}_n)g(\mathbf{t}; \theta)$$

dove $k(\mathbf{x}_n) \geq 0$ è una funzione \mathcal{B}_n -misurabile indipendente da θ e $g(\mathbf{t}; \theta) \geq 0$ è una funzione \mathcal{B}_r -misurabile.

Esempio 12.

Consideriamo una *v.c.* $X \sim N(\mu, \sigma^2)$ e sia \mathbf{X}_n la *v.c.* ottenuta da un campionamento *c.s.* di numerosità n . Il vettore dei parametri sia $\theta = (\mu, \sigma^2) \in \mathfrak{R} \times \mathfrak{R}^+$.

La funzione di verosimiglianza risulta

$$\begin{aligned} V_X(\theta, \mathbf{x}_n) &= f_X(\mathbf{x}_n; \theta) = \prod_{i=1}^n \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(x_i - \mu)^2}{2\sigma^2} \right\} \right] \\ &= \left(\frac{1}{\sqrt{2\pi}} \right)^n \left(\frac{1}{\sqrt{\sigma^2}} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right\} \end{aligned}$$

Indicato con $\mathbf{t} = (t_1 = \sum_{i=1}^n x_i, t_2 = \sum_{i=1}^n x_i^2)$ possiamo scrivere la relazione precedente come

$$\begin{aligned} V_X(\theta, \mathbf{x}_n) &= (2\pi)^{-n/2} (\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i^2 - 2\mu x_i + \mu^2) \right\} \\ &= (2\pi)^{-n/2} (\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} (t_2 - 2\mu t_1 + n\mu^2) \right\} \\ &= (2\pi)^{-n/2} (\sigma^2)^{-n/2} \exp \left\{ -\frac{n\mu^2}{2\sigma^2} \right\} \exp \left\{ -\frac{t_2 - 2\mu t_1}{2\sigma^2} \right\} \end{aligned}$$

Espressione che può fattorizzarsi come

$$V_X(\theta, \mathbf{x}_n) = k(\mathbf{x}_n) g(\mathbf{t} = (t_1, t_2); \theta)$$

con

$$k(\mathbf{x}_n) = (2\pi)^{-n/2}$$

$$g(t_1, t_2; \theta) = (\sigma^2)^{-n/2} \exp \left\{ -\frac{n\mu^2}{2\sigma^2} \right\} \exp \left\{ -\frac{t_2 - 2\mu t_1}{2\sigma^2} \right\}$$

Quindi l'informatore statistico $\mathbf{t} = (t_1 = \sum_{i=1}^n x_i, t_2 = \sum_{i=1}^n x_i^2)$ è un informatore "sufficiente" per i parametri della famiglia di distribuzioni normali. Da tali informatori si possono ottenere stimatori "sufficienti" dei parametri μ, σ^2 quali la media e la varianza campionaria.

13. Verifica delle ipotesi statistiche

La "verifica delle ipotesi statistiche" si affianca ai "problemi di stima" nell'ambito dell'inferenza e consente di operare una scelta decisionale su due o più ipotesi riguardanti il tipo di distribuzione del carattere in oggetto o, più frequentemente, su valori alternativi dei parametri che identificano la distribuzione, accettando una di tali ipotesi sulla base delle informazioni campionarie disponibili.

L'esigenza di assumere decisioni in condizioni di incertezza, data la componente aleatoria presente nei fenomeni, è propria di molte situazioni, da quelle scientifiche (a cui noi ci riferiamo specificatamente) a quelle personali e di vita di ciascuno.

Possiamo pensare che il problema di una verifica d'ipotesi sorga da una specifica domanda che si fa lo studioso, quale ad esempio:

1. La vicinanza a centrali nucleari aumenta il livello di radiazioni presenti nell'atmosfera?
2. Fumare abitualmente può comportare un aumento dell'insorgenza di fenomeni tumorali alle vie respiratorie e polmonari?
3. L'aggiunta di un particolare catalizzatore in processo chimico può aumentare o mantenere inalterato il rendimento e il tempo di reazione?
4. L'impiego di un diserbante influisce positivamente sulla resa e sulla qualità di produzione di un tipo di coltivazione agricola?
5. Ecc.

Come astrazione, relativamente ai molteplici quesiti che si possono presentare concretamente in ambiti diversi, possiamo considerare che il fenomeno allo studio sia caratterizzato da un legge di distribuzione $F(x; \theta)$ ignota nella sua legge oppure, più frequentemente, per quanto riguarda il parametro θ che la individua.

14. Sistema d'ipotesi dicotomico e regola decisionale

Conviene affrontare in modo formale il problema di verifica d'ipotesi stabilendo alcune definizioni e assunzioni. Sia X una v.c. con *F.d.R.* $F_X(x; \theta)$ dove $\theta \in \Theta$ è ignoto, essendo $\Theta \equiv \mathfrak{R}$ lo spazio parametrico di θ (X e θ possono essere unidimensionali o multidimensionali, per semplicità si sono qui considerati entrambi unidimensionali).

Lo spazio parametrico Θ sia partizionato in due sottospazi Θ_0 e Θ_1 , tali che $\Theta_0 \cup \Theta_1 = \Theta$ e $\Theta_0 \cap \Theta_1 = \emptyset$, così da poterli mettere in corrispondenza

dicotomica con affermazioni alternative (complementari) sullo stato di natura del parametro θ .

Si costituisce un sistema di affermazioni che vengono sottoposte a verifica mediante l'inferenza statistica, che viene indicato come "sistema d'ipotesi"

- **Ipotesi nulla** $H_0: \theta \in \Theta_0$.
- **Ipotesi alternativa** $H_1: \theta \in \Theta_1 \rightarrow \theta \in (\Theta - \Theta_0)$ oppure $H_1: \theta \notin \Theta_0$.

Esempio 13.

Con riferimento alla situazione 2. Consideriamo che la v.c. X misura il livello di radiazione per persona-anno a una distanza pari a 30 km dalla centrale. Assunto che X si distribuisca con legge normale $X \sim N(\theta, \sigma^2 = 0,09)$ con parametro ignoto il valor medio θ , e che lo spazio parametrico di θ sia $\Theta \equiv \mathcal{R}^+ = (0, \infty)$.

Si stabilisca un valore θ_0 come livello, in una scala di misura standardizzata, delle radiazioni per persona-anno se la centrale non è funzionante pari a $\theta_0 = 1,2$. Tale valore viene a costituire il sottospazio $\Theta_0 = \{\theta_0\}$ e quindi il sistema d'ipotesi corrisponde alle seguenti due alternative

$$\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta \neq \theta_0 \end{cases}$$

Oppure

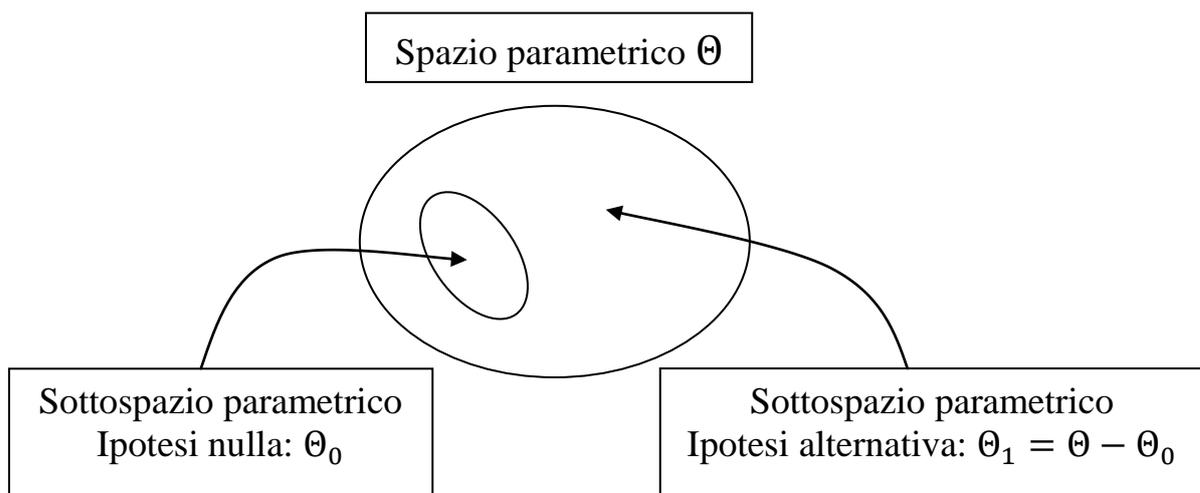
$$\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta > \theta_0 \end{cases}$$

Nell'esempio considerato il secondo sistema d'ipotesi è più interessante in quanto ritiene come alternativa, durante in funzionamento della centrale, che di possa manifestare un livello medio di radiazioni superiore a quello ipotizzato durante il fermo della centrale.

Abbiamo visto da questo esempio come le ipotesi parametriche possano essere:

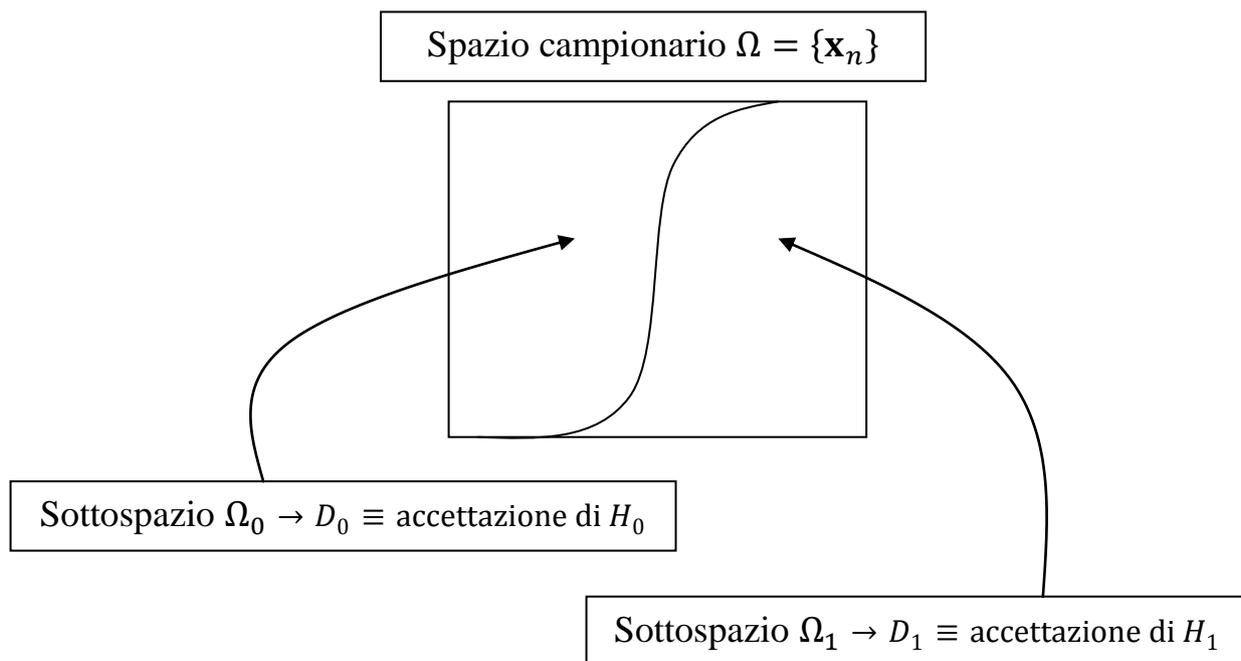
- Unilaterali (ad es. $\theta > \theta_0$);
- Bilaterali (ad es. $\theta \neq \theta_0$);
- Semplici $\Theta_0 = \{\theta_0\}$ se è costituita da un solo elemento;
- Complesse $\Theta_1 = \{\theta > \theta_0\}$ se è costituita da più di un elemento.

La verifica d'ipotesi consiste nel validare l'ipotesi nulla H_0 o quella alternativa H_1 (cioè decidendo: a favore di H_0 - decisione - D_0 oppure a favore di H_1 - decisione - D_1) ricorrendo a una serie di osservazioni del fenomeno (campionamento) ottenute predisponendo una opportuna sperimentazione, programmata in modo da rispettare i vincoli umani, tecnologici e di costo, al fine di raggiungere risultati decisionali il più aderenti possibile alla realtà, limitando gli errori di decisione e favorendo la precisione, tenendo conto dell'aleatorietà insita nel modello probabilistico considerato.



A differenza della situazione di stima, l'oggetto della verifica d'ipotesi è una "congettura" sull'appartenenza di θ a un sottospazio $\Theta_0 \subset \Theta$ stabilendo una regola decisionale che permetta di accettare o rifiutare l'ipotesi H_0 formulata.

Il test statistico è una procedura che comporta l'accettazione dell'ipotesi H_0 o H_1 in relazione ai risultati campionari $\mathbf{x}_n = (x_i; i = 1, 2, \dots, n)$ suddividendo lo spazio campionario $\Omega = \{\mathbf{x}_n\}$ in due sottospazi Ω_0 e Ω_1 associati rispettivamente alle decisioni D_0 (accettazione dell'ipotesi nulla H_0) e D_1 (accettazione dell'ipotesi alternativa H_1).

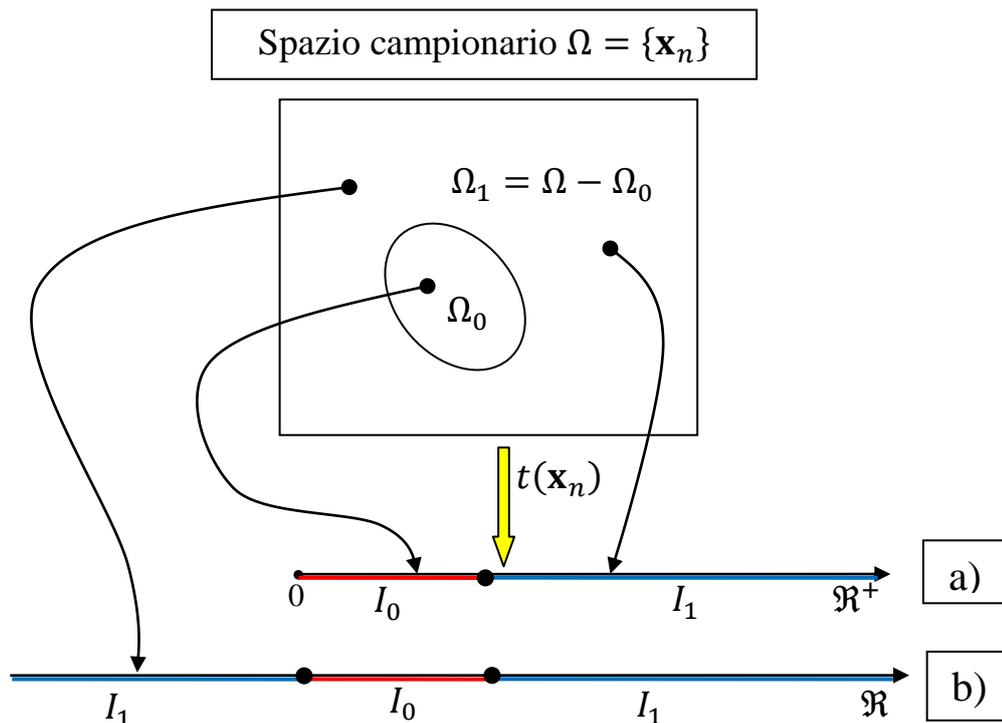


Il test statistico proposto viene detto “non casualizzato” per distinguerlo da quello in cui alcuni esiti campionari \mathbf{x}_n vengono assegnati alla decisione D_0 oppure a D_1 con una probabilità diversa da 0 o da 1.

La partizione dello spazio campionario $\Omega = \{\mathbf{x}_n\}$ nei due sottospazi Ω_0 e Ω_1 può opportunamente risolversi trasferendo dallo spazio in \mathfrak{R}^n a uno spazio di dimensione ridotta \mathfrak{R}^r con $r < n$, ad esempio in \mathfrak{R} , mediante un informatore statistico $t(\mathbf{x}_n) = t(x_1, x_2, \dots, x_n)$, che viene denominata “statistica test”. È possibile allora stabilire in \mathfrak{R} due insiemi (intervalli) I_0 e I_1 , con $I_0 \cap I_1 = \emptyset$, a cui associare le decisioni D_0 e D_1 .

$$\begin{cases} D_0 - \text{se } t(\mathbf{x}_n) \in I_0 \text{ si accetta come vera l'ipotesi } H_0 \\ D_1 - \text{se } t(\mathbf{x}_n) \in I_1 \text{ si accetta come vera l'ipotesi } H_1 \end{cases}$$

I valori estremi dell'intervallo I_0 sono detti “valori critici” della statistica test. La regola decisionale può presentarsi in forma “unilaterale” o “bilaterale” a seconda se I_0 è delimitato (superiormente o inferiormente) da I_1 o se I_0 è delimitato sia a destra che a sinistra da I_1 (v. in figura seguente i grafici a) e b) rispettivamente).



15. Errori e probabilità decisionali: livello di significatività di un test e funzione di potenza

Consideriamo due valori del parametro $\theta_0 \in \Theta_0$ e $\theta_1 \in \Theta_0$ che definiscono due *F.d.R.* della v.c. X rispettivamente relative alle due ipotesi H_0 e H_1 , una stessa determinazione campionaria \mathbf{x}_n può provenire da entrambe le ipotesi con probabilità diverse e comportare decisioni alternative a seconda che $\mathbf{x}_n \in \Omega_0$ e $\mathbf{x}_n \in \Omega_1$. Di conseguenza si possono trarre conferme o errori di decisioni dalla procedura impiegata.

In particolare si potrà verificare:

- un “errore del 1° tipo” ossia accettare l’ipotesi H_1 quando è vera l’ipotesi $H_0 \Rightarrow D_1|H_0$;
- un “errore del 2° tipo” ossia accettare l’ipotesi H_0 quando è vera l’ipotesi $H_1 \Rightarrow D_0|H_1$.

<i>Condizioni</i>	Decisione	
	D_0	D_1
Stato di natura H_0	---	Errore di 1° tipo
Stato di natura H_1	Errore di 1° tipo	---

Stabiliti i valori del parametro θ , rispettivamente θ_0 e θ_1 è possibile determinare le probabilità degli errori, indicati con α e β .

<i>Probabilità</i>	Decisione	
	D_0	D_1
Stato di natura H_0	$1 - \alpha$	$\alpha = P(D_1 H_0)$
Stato di natura H_1	$\beta = P(D_0 H_1)$	$1 - \beta$

La procedura decisionale (D_0, D_1) verrà scelta in modo da rendere piccoli le probabilità degli errori di 1° e 2° tipo, in particolare spesso si assegna il valore di α , relativo all'ipotesi nulla che è di riferimento della verifica d'ipotesi. Il valore $\alpha = P(D_1|H_0)$ è detto "livello di significatività" o, anche, "dimensione del test". Per α normalmente si assumono valori quali: 0,01; 0,05; 0,10.

Se l'ipotesi H_0 è costituita da più di un elemento il valore della dimensione del test α è definito come

$$\alpha = \sup_{\theta \in \Theta_0} P(D_1|\theta)$$

Per quanto riguarda l'errore del 2° tipo la sua probabilità è funzione, oltre della procedura decisionale e della scelta di α , dal valore del parametro $\theta \in \Theta_1$, viene quindi espressa come funzione di θ

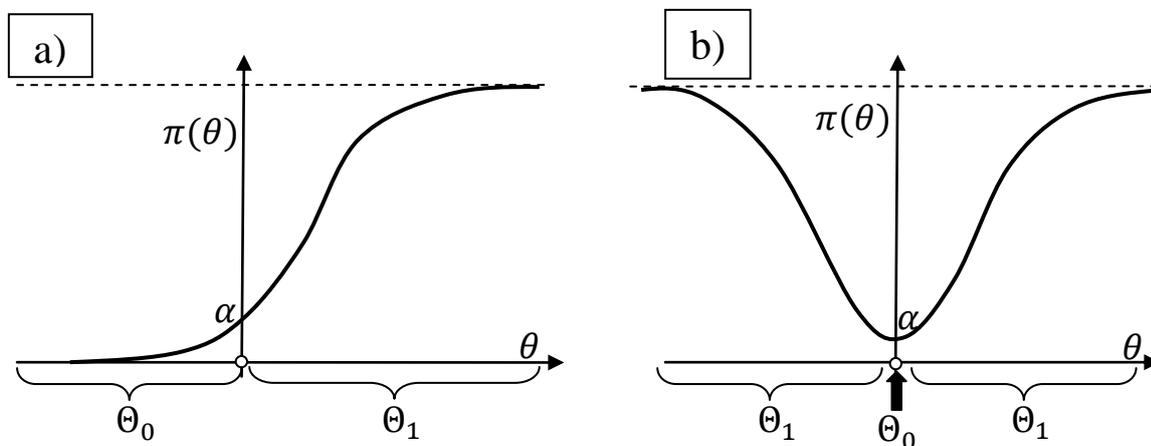
$$\beta(\theta) = P(D_0; \theta \in \Theta_1)$$

La funzione $\beta(\theta)$ è detta “funzione operativa caratteristica” della procedura test considerata e fornisce la probabilità di accettazione dell’ipotesi nulla quando $\theta \in \Theta_1$, ossia quando è vera l’ipotesi alternativa.

Dalla funzione $\beta(\theta)$ si ottiene una funzione che è detta “funzione di potenza del test” come la probabilità di rifiuto dell’ipotesi nulla H_0 quando è vera l’ipotesi alternativa H_1 .

$$\pi(\theta) = 1 - \beta(\theta) \quad \text{per } \theta \in \Theta_1$$

La ricerca tra diverse procedure di test viene condotta proprio sull’andamento della funzione di potenza ricercando quel test che comporta il massimo valore di $\pi(\theta)$ per ogni $\theta \in \Theta_1$, avendo prefissato un livello di significatività α . Qualora esista una tale procedura essa è considerata “ottimale” e indicata come procedura “uniformemente massimamente potente” al livello di significatività α . Non sempre è possibile ottenere una tale procedura.



La funzione di potenza $\pi(\theta)$ di un test statistico costituisce uno strumento di confronto tra procedure diverse al fine di facilitarne la scelta.

Ricordiamo alcune proprietà che possono caratterizzare la funzione di potenza $\pi(\theta)$, a pari livello di significatività α del test statistico.

- **Test più potente** (confronto tra due test)

Un test è migliore di un altro se la sua funzione di potenza, in corrispondenza di una specifica alternativa $\theta_1 \in \Theta_1$, presenta una funzione di potenza maggiore di quella posseduta dall'altro test.

- **Test uniformemente più potente** (confronto tra due test)

Se quello che avviene nel caso precedente è esteso a ogni valore di θ contenuto in Θ_1 allora tale test è detto “uniformemente più potente”.

- **Test uniformemente massimamente potente**

Un test che presenti rispetto ad ogni altro test (sempre a pari livello di significatività α , una funzione di potenza uniformemente superiore (o meglio non inferiore) di quella di ogni altro test (per $\theta \in \Theta_1$) è detto “uniformemente massimamente potente. Una tale condizione di “ottimalità” non è sempre raggiungibile. Qualora si consideri un sistema d'ipotesi entrambe “semplici” si dimostra la possibilità di ottenere un test ottimale avente tale proprietà.

Diamo l'enunciato del seguente “Lemma di Neyman-Pearson” che ci assicura un test “uniformemente massimamente potente”.

Sia X una v.c., per esempio continua con f.d.d. $f_x(x; \theta)$, dove l'ignoto parametro θ si vuole saggiare rispetto al seguente sistema d'ipotesi semplici

$$\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta = \theta_1 \end{cases}$$

dove $\theta_0 \neq \theta_1$ sono valori assegnati. Si disponga di un campionamento c.s. $\mathbf{x}_n = (x_i; i = 1, 2, \dots, n)$ e come statistica test si utilizzi il “rapporto di verosimiglianza” dato da

$$r(\mathbf{x}_n) = \frac{V(\theta_1, \mathbf{x}_n)}{V(\theta_0, \mathbf{x}_n)} = \frac{\prod_{i=1}^n f_x(x_i; \theta_1)}{\prod_{i=1}^n f_x(x_i; \theta_0)}$$

Più tale rapporto è elevato maggiore è la coerenza dei dati con l'ipotesi H_1 , è logico quindi come regola decisionale stabilire

$$\begin{cases} D_0 \rightarrow \Omega_0 \equiv \{\mathbf{x}_n: r(\mathbf{x}_n) \leq \lambda_c\} \text{ si accetta come vera l'ipotesi } H_0 \\ D_1 \rightarrow \Omega_1 \equiv \{\mathbf{x}_n: r(\mathbf{x}_n) > \lambda_c\} \text{ si accetta come vera l'ipotesi } H_1 \end{cases}$$

dove λ_c è il valore critico della procedura ottenibile in funzione del valore di α prefissato mediante la condizione (valida sotto H_1)

$$P(\mathbf{x}_n \in \Omega_1 | H_0) = P(\mathbf{x}_n \in \Omega_1 | \theta_0) = \int_{\mathbf{x}_n \in \Omega_1} V(\theta_0, \mathbf{x}_n) d\mathbf{x}_n = \alpha$$

La potenza del test, cioè la probabilità di accettare H_1 quando è vera H_1 , risulta

$$\begin{aligned} \pi = 1 - \beta &= P(\mathbf{x}_n \in \Omega_1 | \theta_1) = \int_{\mathbf{x}_n \in \Omega_1} V(\theta_1, \mathbf{x}_n) d\mathbf{x}_n \\ &= \int_{\mathbf{x}_n \in \Omega_1} \left[\frac{V(\theta_1, \mathbf{x}_n)}{V(\theta_0, \mathbf{x}_n)} \right] V(\theta_0, \mathbf{x}_n) d\mathbf{x}_n \\ &= \int_{\mathbf{x}_n \in \Omega_1} r(\mathbf{x}_n) V(\theta_0, \mathbf{x}_n) d\mathbf{x}_n \end{aligned}$$

Per massimizzare $\pi = 1 - \beta$ occorre considerare nel dominio Ω_1 tutti quei punti \mathbf{x}_n che presentano i più elevati valori di $r(\mathbf{x}_n)$ e questo corrisponde alla regola decisionale adottata: $\Omega_1 \equiv \{\mathbf{x}_n: r(\mathbf{x}_n) > \lambda_\alpha\}$ per accettare H_1 .

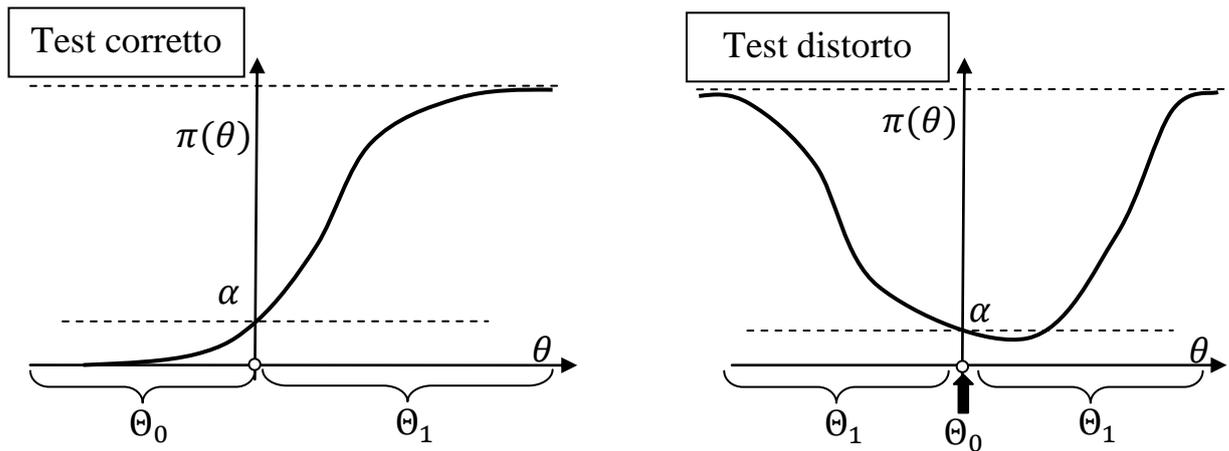
Questo Lemma costituisce uno dei pilastri dell'inferenza statistica ed è stato formalizzato quasi contemporaneamente da Neyman e Pearson attorno agli anni '30 del secolo scorso.

Il Lemma di Neyman-Pearson è generalizzabile anche nel caso di verifica di H_0 semplice contro una verifica H_1 composta se è possibile ottenere un test, dato dal rapporto di verosimiglianza a livello α , per ogni ipotesi alternativa semplice contenuta in H_1 .

- **Test non distorto**

Un test si dice “non distorto” o “corretto” se la funzione di potenza è tale che

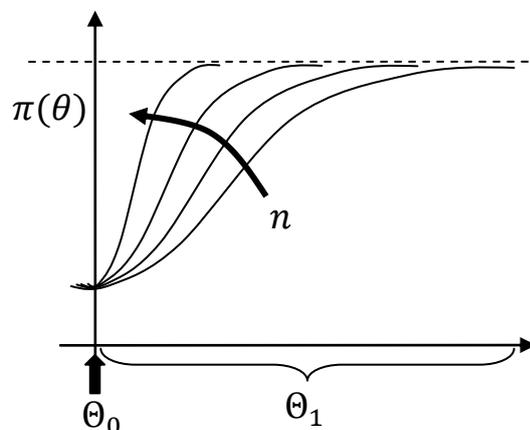
$$\pi(\theta) \geq \alpha \text{ per } \forall \theta \in \Theta_1$$



- **Test consistente**

Un test, avente livello di significatività α , si dice “consistente” se la funzione di potenza è tale che

$$\lim_{n \rightarrow \infty} \pi(\theta; n) = 1 \text{ per } \forall \theta \in \Theta_1$$



Un test consistente è almeno “asintoticamente non distorto”. Un test non distorto può non essere “consistente”.

16. Prove d’ipotesi riguardanti la media di una o due variabili casuali normali con varianza nota o ignota

A. Test sulla media con varianza nota e ipotesi entrambe semplici

Sia $X \sim N(\mu, \sigma^2)$ con $\sigma^2 > 0$ noto e $\mathbf{x}_n = (x_i; i = 1, 2, \dots, n)$ un campione c.s. si voglia verificare il seguente sistema d’ipotesi

$$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu = \mu_1 \end{cases} \text{ per } \mu_0 < \mu_1$$

Siamo nelle condizioni di applicare il Lemma di Neyman-Pearson utilizziamo come statistica test il rapporto di verosimiglianza che è

$$\begin{aligned} r(\mathbf{x}_n) &= \frac{\prod_{i=1}^n f_x(x_i; \mu_1)}{\prod_{i=1}^n f_x(x_i; \mu_0)} = \prod_{i=1}^n \frac{f_x(x_i; \mu_1)}{f_x(x_i; \mu_0)} \\ &= \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n [(x_i - \mu_1)^2 - (x_i - \mu_0)^2] \right\} \\ &= \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n [(x_i - \mu_1)^2 - (x_i - \mu_0)^2] \right\} \end{aligned}$$

semplificando l’espressione entro parentesi quadra abbiamo

$$r(\mathbf{x}_n) = \exp \left\{ \frac{(\mu_1 - \mu_0)[2 \sum_{i=1}^n x_i - (\mu_1 + \mu_0)]}{2\sigma^2} \right\}$$

funzione monotona crescente della media campionaria $\bar{x} = \sum_{i=1}^n x_i/n$, consegue che la regola decisionale è esprimibile in termini della statistica test \bar{x}

$$\begin{cases} D_0 \equiv \{\mathbf{x}_n: r(\mathbf{x}_n) \leq \lambda_c\} \\ D_1 \equiv \{\mathbf{x}_n: r(\mathbf{x}_n) > \lambda_c\} \end{cases} \Rightarrow \begin{cases} D_0 \equiv \bar{x} \leq v_c \\ D_1 \equiv \bar{x} > v_c \end{cases}$$

dove v_c è il valore critico.

Per determinare il valore v_c , assegnato α si ricordi che la v.c. $\bar{X} \sim N(\mu, \sigma^2/n)$ quindi abbiamo per $\mu = \mu_0$

$$\begin{aligned}\alpha &= P(\bar{x} > v_c; \mu_0) = 1 - F_{\bar{X}}(v_c; \mu_0) = 1 - \Phi\left(\frac{v_c - \mu_0}{\sigma/\sqrt{n}}\right) \Rightarrow \Phi\left(\frac{v_c - \mu_0}{\sigma/\sqrt{n}}\right) \\ &= 1 - \alpha\end{aligned}$$

Indicato con $z_{1-\alpha}$ il punto quantile della distribuzione normale standardizzata $\Phi(z_{1-\alpha}) = 1 - \alpha$, abbiamo il valore critico v_c pari a

$$v_c = \mu_0 + z_{1-\alpha}\sigma/\sqrt{n}$$

La probabilità di commettere un errore di 2° tipo risulta

$$\begin{aligned}\beta &= P(\bar{x} \leq v_c; \mu_1) = F_{\bar{X}}(v_c; \mu_1) = \Phi\left(\frac{v_c - \mu_1}{\sigma/\sqrt{n}} \pm \frac{\mu_0}{\sigma/\sqrt{n}}\right) \\ &= \Phi\left(z_{1-\alpha} - \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right) = \Phi(z_{1-\alpha} - \delta)\end{aligned}$$

Essendo $\delta = \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}$ una misura standardizzata della distanza tra le due medie ipotizzate.

La procedura ottenuta è “corretta”, “consistente” e ”uniformemente massimamente potente”.

B. Test sulla media con varianza nota e ipotesi alternativa unilaterale

Si consideri le stesse assunzioni fatte in A con però il seguente sistema d'ipotesi

$$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$$

Per ogni singola ipotesi alternativa ($\mu = \mu_1 > \mu_0$) si può procedere come in A ottenendo un test basato sulla media campionaria, che ha la stessa

regione di rifiuto ed è “corretto”, “consistente” e ”uniformemente massimamente potente”.

La regola decisionale e il valore critico sono dati ancora, come nel caso A da

$$\begin{cases} D_0 \equiv \bar{x} \leq v_c \\ D_1 \equiv \bar{x} > v_c \end{cases}$$

$$v_c = \mu_0 + z_{1-\alpha} \sigma / \sqrt{n}$$

La funzione di potenza, per $\mu > \mu_0$ è

$$\begin{aligned} \pi(\mu) &= 1 - \beta = P(\bar{x} > v_c; \mu) = 1 - F_{\bar{X}}(v_c; \mu) = 1 - \Phi\left(\frac{v_c - \mu}{\sigma/\sqrt{n}}\right) \\ &= 1 - \Phi\left(z_{1-\alpha} - \frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right) = 1 - \Phi(z_{1-\alpha} - \delta) \end{aligned}$$

Al tendere di $\delta = \frac{\mu - \mu_0}{\sigma/\sqrt{n}}$ all'infinito, abbiamo $\Phi(z_{1-\alpha} - \delta) \rightarrow \Phi(-\infty) = 0$ e quindi $\pi(\mu) \rightarrow 1$.

C. Test sulla media con varianza nota e ipotesi alternativa bilaterale

Si consideri le stesse assunzioni fatte in A con però il seguente sistema d'ipotesi in cui l'ipotesi alternativa è bilaterale

$$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$$

In questa situazione non è possibile ottenere una procedura “ottimale” avvalendosi del Lemma di Neyman-Pearson, la regola decisionale, per analogia con il sistema d'ipotesi considerato, è esprimibile in termini della statistica test \bar{x}

$$\begin{cases} D_0 \equiv v'_c \leq \bar{x} \leq v''_c \\ D_1 \equiv (\bar{x} < v'_c) \cup (\bar{x} > v''_c) \end{cases}$$

I valori critici v'_c e v''_c sono determinati, sotto l'ipotesi nulla, in modo da assicurare il livello di significatività $\alpha = \alpha' + \alpha''$ così d'avarsi

$$\alpha' = P(\bar{x} < v'_c; \mu_0) = F_{\bar{X}}(v'_c; \mu_0) = \Phi\left(\frac{v'_c - \mu_0}{\sigma/\sqrt{n}}\right)$$

$$\alpha'' = P(\bar{x} > v''_c; \mu_0) = 1 - F_{\bar{X}}(v''_c; \mu_0) = 1 - \Phi\left(\frac{v''_c - \mu_0}{\sigma/\sqrt{n}}\right)$$

Data la simmetria della legge normale si può assumere $\alpha' = \alpha'' = \alpha/2$, da cui si ottengono i valori critici

$$\Phi\left(\frac{v'_c - \mu_0}{\sigma/\sqrt{n}}\right) = \alpha/2 = \Phi(z_{\alpha/2}) \Rightarrow v'_c = \mu_0 + z_{\alpha/2}\sigma/\sqrt{n}$$

$$1 - \Phi\left(\frac{v''_c - \mu_0}{\sigma/\sqrt{n}}\right) = \frac{\alpha}{2} = \Phi(z_{1-\alpha/2}) \Rightarrow v''_c = \mu_0 + z_{1-\alpha/2}\sigma/\sqrt{n}$$

Essendo $z_{\alpha/2} = -z_{1-\alpha/2}$ l'intervallo di accettazione di H_0 è $\mu_0 \pm z_{1-\alpha/2}\sigma/\sqrt{n}$ mentre la regione di rifiuto è data da: $|\bar{x} - \mu_0| > z_{1-\alpha/2}\sigma/\sqrt{n}$.

D. Test sulle medie di popolazioni con varianze note e ipotesi entrambe semplici

Si consideri due popolazioni aventi una caratteristica distribuita normalmente con media ignota e varianza nota:

$$X_1 \sim N(\mu_1, \sigma_1^2); \quad X_2 \sim N(\mu_2, \sigma_2^2)$$

Dati due campioni *c.s.* mutualmente indipendenti, di numerosità n e m estratti dalle due popolazioni

$$\mathbf{x}_1 = (x_{1i}; i = 1, 2, \dots, n), \quad \mathbf{x}_2 = (x_{2i}; i = 1, 2, \dots, m)$$

si voglia congetturare sulle medie considerando il seguente sistema d'ipotesi:

$$\begin{cases} H_0: \mu_2 = \mu_1 \\ H_1: \mu_2 = \mu_1 + d \end{cases}$$

Si consideri la v.c. $X_D = X_2 - X_1$, avente media $\mu_D = M(X_D) = M(X_2 - X_1) = \mu_2 - \mu_1 = d$ e varianza $\sigma_D^2 = Var(X_D) = Var(X_1) + Var(X_2) = \sigma_1^2 + \sigma_2^2$, il sistema d'ipotesi precedentemente indicato può scriversi come equivalente nei riguardi di X_D come

$$\begin{cases} H_0: \mu_2 - \mu_1 = 0 \\ H_1: \mu_2 - \mu_1 = d \end{cases} \Rightarrow \begin{cases} H_0: \mu_D = 0 \\ H_1: \mu_D = d \end{cases}$$

Come funzione test consideriamo la stima $\bar{x}_D = \bar{x}_2 - \bar{x}_1$ con $\bar{x}_1 = \sum_{i=1}^n x_{1i}/n$ e $\bar{x}_2 = \sum_{i=1}^m x_{2i}/m$ avente come s.q.m. $\sigma_{\bar{x}_D} = \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}$.

Impiegando tali valori nella analoga procedura decisionale nella situazione C precedentemente considerata.

L'accettazione dell'ipotesi nulla si ha per

$$D_0 \equiv \frac{|\bar{x}_D - \mu_D|}{\sigma_{\bar{x}_D}} = \frac{|\bar{x}_D|}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \leq z_{1-\alpha/2}$$

17. Prove d'ipotesi riguardanti la varianza di una o due variabili casuali normali con media nota o ignota

18. Prove d'ipotesi riguardanti il parametro p della variabile casuale indicatore

Riferimenti bibliografici

Cox D.R., (2006) *Principles of Statistical Inference*, Cambridge University Press, Cambridge.

Dudewicz S.N., Mishra S.N., (1963) *Modern Mathematical Statistics*, J. Wiley, New York.

Fisz M., (1963) *Probability Theory and Mathematical Statistics*, 2nd edition, J. Wiley, New York.

Lehmann E.L., (1959) *Testing Statistical Hypotheses*, J. Wiley, New York.

Lehmann E.L., Casella G., (1998) *Theory of Point Estimation*, 2nd edition, Springer-Verlag, New York.

Mood A.M., Graybill F.A., Boes D.C., (1988) *Introduzione alla Statistica*, McGraw-Hill Italia, Milano.

Ross S.M., (2008) *Probabilità e statistica per l'ingegneria e le scienze*, II edizione, Apogeo, Milano.

Wilks S.S., (1963) *Mathematical Statistics*, J. Wiley, New York.

Sommario

1.	Il ruolo della variabile casuale normale nell'inferenza statistica	1
2.	I teoremi del limite centrale	1
3.	Inferenza e induzione statistica.....	7
4.	Campionamento casuale da una variabile casuale.....	9
5.	Il campionamento casuale semplice da una v.c. unidimensionale X	10
6.	Campione casuale semplice, stima e stimatore di un parametro	12
7.	Criteri di qualificazione degli stimatori.....	16
8.	Metodi di ottenimento degli stimatori	21
9.	Il metodo di stima della massima verosimiglianza.....	21
10.	Il metodo di stima dei momenti	27
11.	La stima dei parametri di variabili casuali notevoli	31
12.	Statistica e stimatore sufficiente	34
13.	Verifica delle ipotesi statistiche.....	39
14.	Sistema d'ipotesi dicotomico e regola decisionale.....	40
15.	Errori e probabilità decisionali: livello di significatività di un test e funzione di potenza	44
16.	Prove d'ipotesi riguardanti la media di una o due variabili casuali normali con varianza nota o ignota	50
17.	Prove d'ipotesi riguardanti la varianza di una o due variabili casuali normali con media nota o ignota.....	55
18.	Prove d'ipotesi riguardanti il parametro p della variabile casuale indicatore	55

