

**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”.

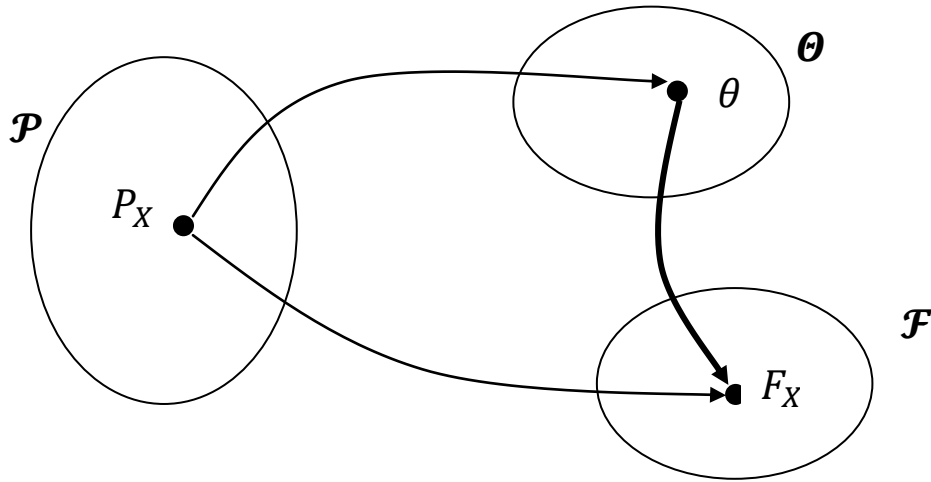
3. Inferenza e induzione statistica

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.*

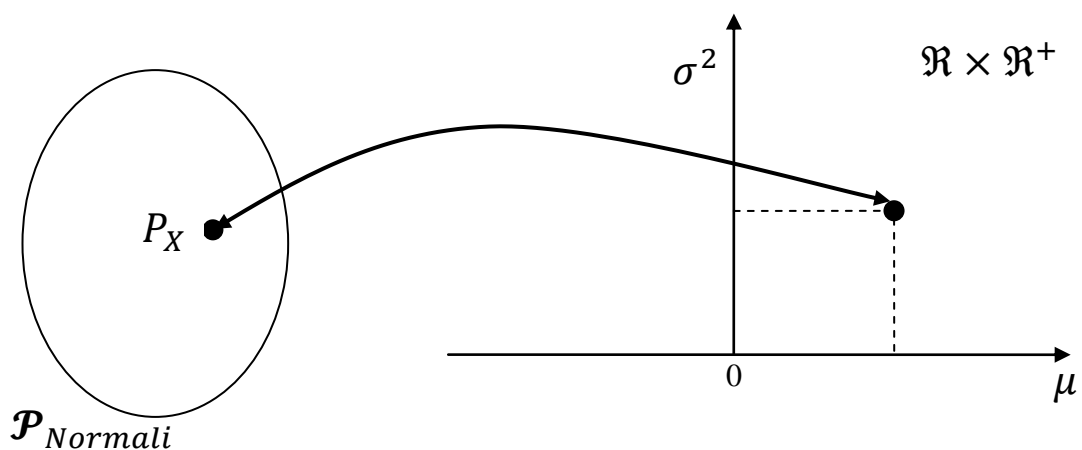


Esempio 1.

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 $\theta: \mathcal{P}_{\text{Normale}} \rightarrow (\mathcal{R} \times \mathcal{R}^+)$

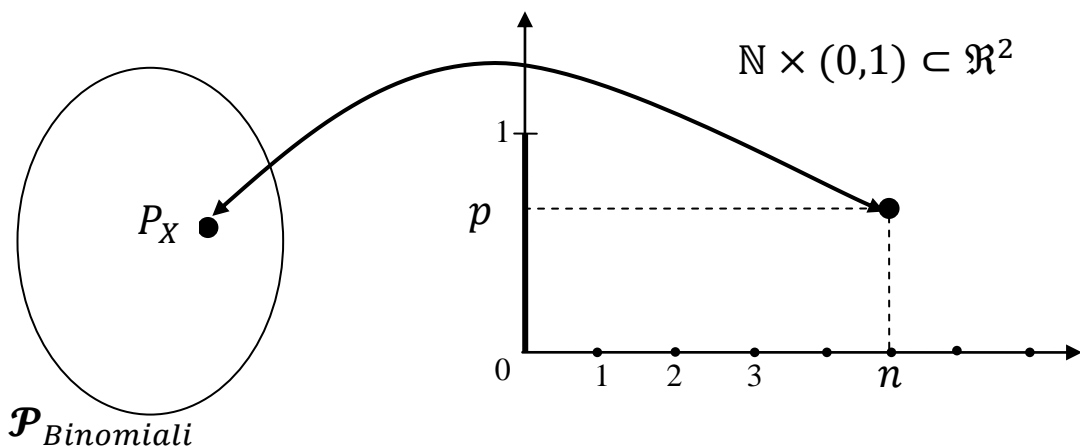


Esempio 2.

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

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

- 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 .

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

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

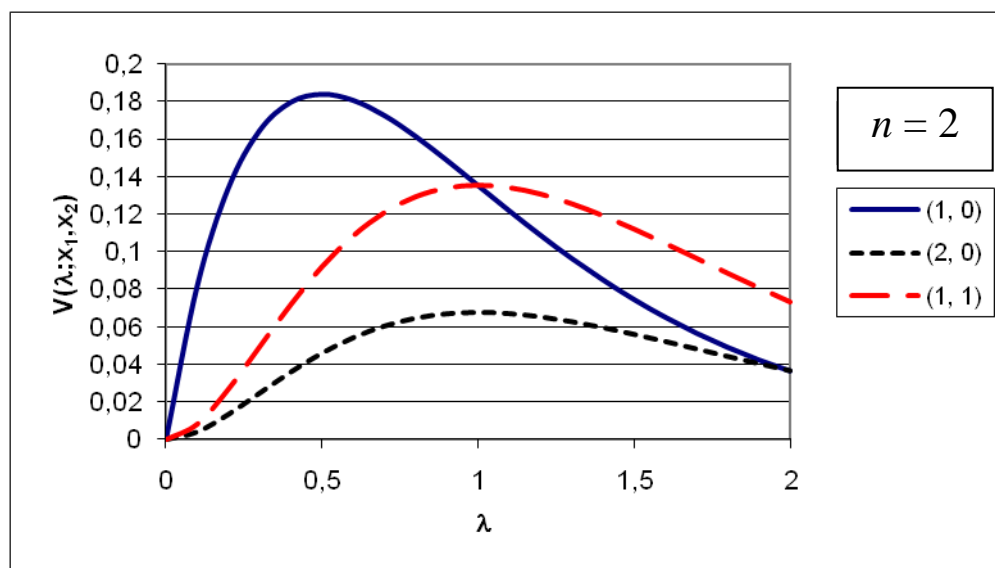
$$V(\lambda; (1, 0)) = e^{-2\lambda} \lambda$$

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

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. Verifica delle ipotesi statistiche

13. Test parametrici e non parametrici, test semplici o complessi

14. Sistema d'ipotesi dicotomico e regola decisionale

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

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

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.	Verifica delle ipotesi statistiche.....	34
13.	Test parametrici e non parametrici, test semplici o complessi.....	34
14.	Sistema d'ipotesi dicotomico e regola decisionale.....	34
15.	Errori e probabilità degli errori decisionali: livello di significatività di un test e funzione di potenza.....	34
16.	Prove d'ipotesi riguardanti la media di una o due variabili casuali normali con varianza nota o ignota	35
17.	Prove d'ipotesi riguardanti la varianza di una o due variabili casuali normali con media nota o ignota.....	35
18.	Prove d'ipotesi riguardanti il parametro p della variabile casuale indicatore	35

