Stima puntuale

- *Stimare*: attribuire un valore plausibile ad una grandezza (parametro) non misurabile esattamente.

- *Stimatore del parametro* θ: ogni statistica

$$T = t(X_1, X_2, \ldots, X_n)$$

utilizzata per stimare θ.

NOTA BENE: è una grandezza aleatoria perchè è una funzione del campione che pure è aleatorio

- *Stima*: la singola determinazione dello stimatore,

$$t(x_1, x_2, \ldots, x_n).$$

cioè il valore che lo stimatore del parametro assume una volta che esso è valutato sul campione che ho effettivamente estratto

Esempio: supponiamo di voler stimare la **media** di una popolazione. A tal fine estraiamo un campione casuale di ampiezza 5 ed otteniamo i seguenti valori:

$$x_1 = 2, \quad x_2 = -1, \quad x_3 = 0, \quad x_4 = 4, \quad x_5 = 3.5$$

Come stimatore della media della popolazione (parametro di interesse) decidiamo di usare la media del campione, (il campione è casuale quindi mi aspetto che sia rappresentativo della popolazione da cui l'ho estratto). Abbiamo quindi che lo stimatore è

$$T_5 = t(X_1, X_2, \ldots, X_5) = \bar{X}$$

la stima è il valore che lo stimatore assume con riferimento al campione estratto:

$$t(x_1, x_2, \ldots, x_5) = \frac{2 - 1 + 0 + 4 + 3.5}{5} = 1.7$$

quindi la nostra stima è

$$\hat{\mu} = 1.7$$

Criteri di valutazione degli stimatori

- *Errore quadratico medio* dello stimatore

$$T = t(X_1, X_2, \ldots, X_n)$$

di θ:

$$\mathrm{MSE}_\theta(T) = \mathrm{E}_\theta[(T - \theta)^2].$$

NOTA BENE il MSE (o EQM) è una funzione del parametro θ

- *Confronto tra stimatori sulla base dell'errore quadratico medio*

- In generale non esiste uno stimatore con errore quadratico medio *uniformemente minimo* cioè più piccolo (rispetto agli altri stimatori) per tutti i valori del parametro θ.

Stimatori non distorti

- Lo stimatore $T = t(X_1, X_2, \ldots, X_n)$ di θ si dice *non distorto* se

$$\mathrm{E}_\theta(T) = \theta, \quad \forall\theta.$$

- *Distorsione (BIAS)*:

$$\mathrm{B}_\theta(T) = \mathrm{E}_\theta(T) - \theta.$$

può essere positiva (in media sovrastimo il parametro) o negativa (in media sottostimo)

- *Decomposizione dell'errore quadratico medio*: (da sapere la dimostrazione)

$$\mathrm{MSE}_\theta(T) = \mathrm{Var}_\theta(T) + \left[\mathrm{B}_\theta(T)\right]^2.$$

- Se T è uno stimatore *non distorto*,

$$\mathrm{MSE}_\theta(T) = \mathrm{Var}_\theta(T).$$

Esempio

Descriviamo il numero di visitatori settimanale di un negozio di arredamento con un numero aleatorio X. Il numero medio di visite è $\mu = E(X)$ e un indice della variabilità delle visite può essere la varianza $\sigma^2 = V(X)$.

Supponiamo di sapere che $\sigma^2 = 48.5$.

Si consideri la media campionaria come stimatore del numero medio μ di visite settimanali

Lo stimatore è non distorto perchè sappiamo che la media della media campionaria è pari alla media della popolazione

$$\mathrm{E}(\bar{X}) = \mu$$

Quanto vale l'errore quadratico medio dello stimatore per un campione di ampiezza $n = 30$?

Poichè lo stimatore è non distorto l'errore quadratico medio è pari alla varianza dello stimatore (la distorsione è nulla) quindi

$$\mathrm{MSE}_\mu(\bar{X}) = \mathrm{Var}_\mu(\bar{X}) = \frac{\sigma^2}{n} = \frac{48.5}{30} = 1.62$$

Notiamo che poichè al **denominatore** dell'errore quadratico medio (EQM) compare l'ampiezza del campione, una maggiore ampiezza campionaria darà un errore quadratico medio minore.

Inoltre l'EQM (o MSE) dipende dalla varianza della popolazione (che compare al **numeratore**): intuitivamente, se σ^2 è elevato, ossia se il numero di visite varia molto di settimana in settimana, sarà più difficile stimare il numero medio di visite! Infatti, a parità di n, se σ^2 è elevato l'errore quadratico medio sarà maggiore

Stimatori consistenti

La consistenza è una proprietà **asintotica** di uno stimatore cioè descrive come lo stimatore si comporta all'aumentare dell'ampiezza del campione. In generale mi aspetto che se l'ampiezza del campione aumenta ($n \to \infty$) lo stimatore migliori

- *Concentrazione*: uno stimatore $T = t(X_1, X_2, \ldots, X_n)$ di θ è *tanto più concentrato attorno a* θ quanto maggiore è la probabilità

$$P_\theta(|T - \theta| < \varepsilon), \quad \forall \theta$$

per una costante fissata $\varepsilon > 0$ e piccola a piacere

- *Consistenza*: uno stimatore $T_n = t_n(X_1, X_2, \ldots, X_n)$ di θ è *consistente in probabilità* (o in senso debole) se per ogni $\varepsilon > 0$ si ha

$$\lim_{n \to \infty} P_\theta(|T_n - \theta| < \varepsilon) = 1, \quad \forall \theta.$$

- *Consistenza in media quadratica*: uno stimatore T_n di θ è *consistente in media quadratica* (o in senso forte) se

$$\lim_{\mathrm{n}\to\infty} E_\theta\left[(T_n-\theta)^2\right]=0, \quad \forall\theta.$$

  Quindi uno stimatore è consistente in senso forte se l'errore quadratico medio tende a zero all'aumentare dell'ampiezza del campione estratto

- **Teorema**: La consistenza in senso forte implica la consistenza in senso debole

- **Esempio**: la media campionaria è uno stimatore, per la media della popolazione, consistente in senso forte (e quindi anche in senso debole) perchè l'errore quadratico medio tende a zero al crescere dell'ampiezza del campione:

$$\lim_{\mathrm{n}\to\infty} \mathrm{MSE}_\mu(\bar{X})=\lim_{\mathrm{n}\to\infty}\frac{\sigma^2}{n}=0$$

- **Esempio**: anche la varianza campionaria (T) è uno stimatore consistente in senso forte per la varianza della popolazione ($\theta = \sigma^2$) in quanto abbiamo che

$$\mathrm{E}(S^2) = \sigma^2$$

cioè lo stimatore è non distorto, quindi l'errore quadratico medio coincide con la varianza dello stimatore:

$$\mathrm{Var}(S^2) = \frac{\sigma^4}{n}\left(\beta_2 + 2\frac{n}{n-1}\right)$$

e questa quantità tende a zero al tendere di n all'infinito

- **Esempio**: Voglio stimare la media di una popolazione e come stimatore uso

$$T = 10$$

cioè la stima è sempre 10 indipendentemente dal campione estratto. Questo stimatore non è consistente nè in senso forte nè in senso debole perchè la stima di fatto non dipende dal campione estratto. Il suo MSE sarà nullo solo quando $\theta = 10$.

- Uno stimatore T_n di θ è *asintoticamente non distorto* se

$$\lim_{n\to\infty} \mathrm{B}_\theta(T_n) = 0, \quad \forall\theta.$$

- Se uno stimatore è non distorto a maggior ragione è asintoticamente non distorto

- **Teorema** Condizione necessaria e suff. per la consistenza in media quadratica è che uno stimatore sia asintoticamente non distorto e che la varianza tenda a zero al crescere dell'ampiezza del campione:

$$\lim_{n\to\infty} \mathrm{B}_\theta(T_n) = 0, \quad \forall\theta$$

e

$$\lim_{n\to\infty} V(T_n) = 0$$

- **Dimostrazione** segue dalla scomposizione dell'errore quadratico medio come somma di varianza e distorsione al quadrato

Stimatori efficienti

- Dati due stimatori T_1 e T_2 di θ,

  T_1 è *più efficiente* di T_2 se

  $$\mathrm{MSE}_\theta(T_1) \leq \mathrm{MSE}_\theta(T_2), \quad \forall\theta$$

  e la disuguaglianza vale in senso stretto per almeno un valore di θ.

- La scelta fra due stimatori dello stesso parametro può essere fatta sulla base del rapporto

  $$\mathrm{eff}_\theta(T_1, T_2) = \frac{\mathrm{MSE}_\theta(T_2)}{\mathrm{MSE}_\theta(T_1)}$$

  T_1 si dice più efficiente di T_2 se $\mathrm{eff}_\theta(T_1, T_2) \geq 1$ con la disuguaglianza in senso stretto per almeno un valore di θ

- Se T_1 e T_2 sono non distorti

  $$\mathrm{eff}_\theta(T_1, T_2) = \frac{\mathrm{Var}_\theta(T_2)}{\mathrm{Var}_\theta(T_1)}$$

ESEMPIO: State stimando un parametro θ. Avete a disposizione due stimatori, T_1 e T_2. Sapete che $E(T_1) = \theta - 3$ e $E(T_2) = \theta$. Sapete inoltre che lo scarto quadratico medio di T_1 è pari a 2 mentre lo scarto quadratico medio di T_2 è pari a 7. Quale stimatore scegliereste?

Lo stimatore T_1 è distorto per θ, mentre T_2 è non distorto. Questo non significa che T_2 sia più efficiente!

Il confronto fra due stimatori si basa sull'errore quadratico medio: preferisco lo stimatore con MSE più piccolo

Poiche' l'errore quadratico medio dipende dalla distorsione ma anche dalla varianza dello stimatore, uno stimatore distorto ma con varianza piccola può essere preferibile ad uno stimatore non distorto ma con varianza elevata. E' quello che succede in questo caso.

L'errore quadratico medio di T_1 è

$$MSE_\theta(T_1) = V_\theta(T_1)+(E(T_1)-\theta)^2 = 4+(\theta-3-\theta)^2 = 4+9 = 13\,,$$

mentre quello di T_2 è

$$MSE_\theta(T_2) = V_\theta(T_2) = 49.$$

Poichè $MSE_\theta(T_1) < MSE_\theta(T_2)$ per ogni θ, T_1 è più efficiente di T_2.

Abbiamo inoltre che

$$\mathrm{eff}_\theta(T_1, T_2) = \frac{\mathrm{MSE}_\theta(T_2)}{\mathrm{MSE}_\theta(T_1)} = \frac{49}{13} = 3.7 > 1$$

quindi ancora si vede che T_1 è più efficiente di T_2 e di conseguenza preferirò usare T_1 nonostante sia distorto

ESEMPIO: Per valutare la spesa media mensile μ delle famiglie italiane, si rileva la spesa mensile su un campione casuale di 5 famiglie (solo 5 per semplicità nell'esercizio..). Sia $(X_1, \ldots, X_5)$ il campione, e consideriamo due possibili stimatori per μ,

$$\bar{X} = \frac{\Sigma_{i=1}^{5} X_i}{5}$$

e

$$T = 0.1X_1 + 0.2X_2 + 0.4X_3 + 0.2X_4 + 0.1X_5 \,.$$

Quale dei due stimatori è più efficiente?

Calcoliamo l'errore quadratico medio per i due stimatori.

$$MSE(\bar{X}) = \frac{\sigma^2}{5} = 0.2\,\sigma^2,$$

dove σ^2 indica la varianza della popolazione.

Per il secondo stimatore, poichè

$$\begin{aligned} E(T) &= 0.1E(X_1) + 0.2E(X_2) + \\ &\quad +0.4E(X_3) + 0.2E(X_4) + 0.1E(X_5) \\ &= \mu(0.1 + 0.2 + 0.4 + 0.2 + 0.1) = \mu \ , \end{aligned}$$

si ha

$$\begin{aligned} MSE(T) &= V(T) \\ &= 0.1^2 V(X_1) + 0.2^2 V(X_2) + \\ &\quad +0.4^2 V(X_3) + 0.2^2 V(X_4) + 0.1^2 V(X_5) \\ &= 0.26\sigma^2 \, . \end{aligned}$$

Poichè

$$EQM(\bar{X}) = 0.2\sigma^2 < 0.26\sigma^2 = EQM(T) \, ,$$

la media campionaria $\bar{X}$ è più efficiente di T. (In effetti, potreste dimostrare che la media campionaria è più efficiente di ogni altra combinazione lineare convessa di $X_1, \ldots, X_n$).

- T_{1n} è *asintoticamente più efficiente* di e T_{2n} se

$$\text{eff}_\theta(T_{1n}, T_{2n}) = \lim_{n\to\infty} \frac{\text{MSE}_\theta(T_{2n})}{\text{MSE}_\theta(T_{1n})} \geq 1, \quad \forall\theta$$

  e la disuguaglianza vale in senso stretto per almeno un valore di θ.

Metodi per la determinazione di uno stimatore

- Metodo analogico stima per es. la media della popolazione con la media del campione o la proporzione nella popolazione con la proporzione nel campione). Per la varianza della popolazione lo stimatore analogico è

$$\hat{\sigma}^2 = \frac{1}{n} \sum_i (X_i - \bar{X})^2$$

  che è distorto, per questo abbiamo introdotto lo stimatore

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

  che invece è non distorto

- Metodo della massima verosimiglianza

- Metodo dei momenti

- Metodo dei minimi quadrati

  si usa per stimare i parametri della retta di regressione

- Metodo Bayesiano: tiene conto dell'opinione iniziale oltre che dei dati campionari

ESEMPIO: Sia $X_1 = 0, X_2 = 0, X_3 = 0$ un campione casuale estratto da una popolazione Bernoulliana con probabilità di successo incognita p. Quale è la stima più **verosimile** di p?

$$\hat{p} = 0$$

Se invece il campione fosse stato $X_1 = 1, X_2 = 1, X_3 = 1$ la stima più **verosimile** di p sarebbe:

$$\hat{p} = 1$$

Se invece il campione fosse stato $X_1 = 1, X_2 = 0, X_3 = 1$ quale è la stima più **verosimile** di p?

STIMA DI MASSIMA VEROSIMIGLIANZA

Metodo della massima verosimiglianza

- *Funzione di verosimiglianza*: sia $(x_1, x_2, \ldots, x_n)$ un campione osservato proveniente da una popolazione descritta dal modello $f(x; \theta)$. La funzione di verosimiglianza è la probabilità (densità) di $(x_1, x_2, \ldots, x_n)$ interpretata come funzione del parametro θ,

$$L(\theta) = f(x_1, \ldots, x_n; \theta) =$$

$$= f(x_1; \theta) f(x_2; \theta) \cdots f(x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

nel caso di un campione casuale (quindi con osservazioni iid)

La funzione di verosimiglianza consente di ordinare i diversi valori di θ secondo il "*grado di verosimiglianza*" che ricevono dai dati campionari. La stima di massima verosimiglianza consiste nel proporre come stima di θ (se esiste) il valore "più verosimile", ossia il valore di $\theta \in \Theta$ che rende massima la funzione di verosimiglianza $L(\theta; x_1, \ldots, x_n)$.

- *Stima di massima verosimiglianza*: il valore di θ,

$$\hat{\theta} = t(x_1, x_2, \ldots, x_n),$$

che massimizza la funzione di verosimiglianza,

$$L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta).$$

- *Stimatore di massima verosimiglianza*: $t(X_1, X_2, \ldots, X_n)$

- Generalmente risulta conveniente massimizzare $l(\theta) = \log[L(\theta)]$ risolvendo l'equazione

$$\frac{dl(\theta)}{d\theta} = 0$$

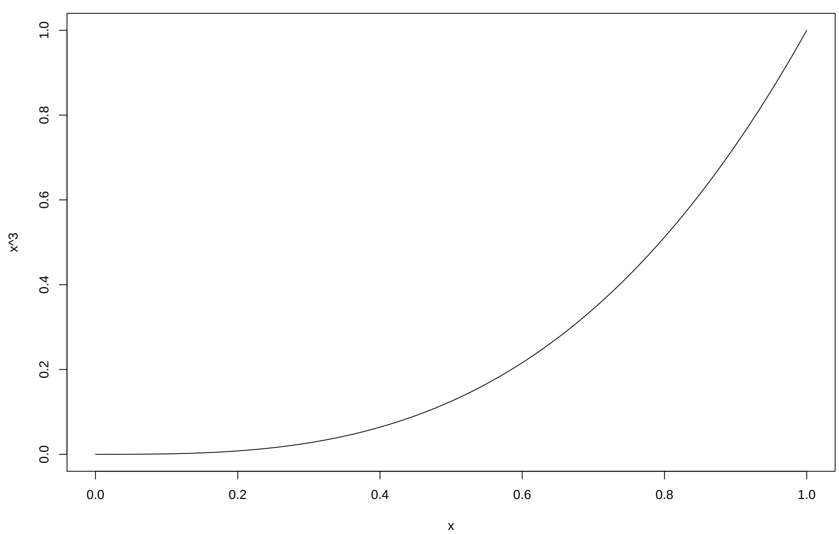
e controllando che in corrispondenza del valore trovato di θ la derivata seconda di $l(\theta)$ sia negativa.

Esempio Bernoulliana cont.: per il campione casuale (iid) $X_1 = 1, X_2 = 1, X_3 = 1$ estratto da una popolazione Bernoulliana la funzione di verosimiglianza è

$$L(X_1 = 1, X_2 = 1, X_3 = 1, p) =$$

$$= P(X_1 = 1)P(X_2 = 1)P(X_3 = 1) = (p)(p)(p) = p^3 \quad 0 \leq p \leq 1$$

quale valore di p massimizza questa funzione? Poichè la funzione di verosimiglianza è una funzione sempre crescente:



il massimo di ha per $\hat{p} = 1$.

Notate che se cercate di fare la derivata della funzione di verosimiglianza ed uguagliarla a zero

$$\frac{dL(p)}{dp} = 3p^2 = 0$$

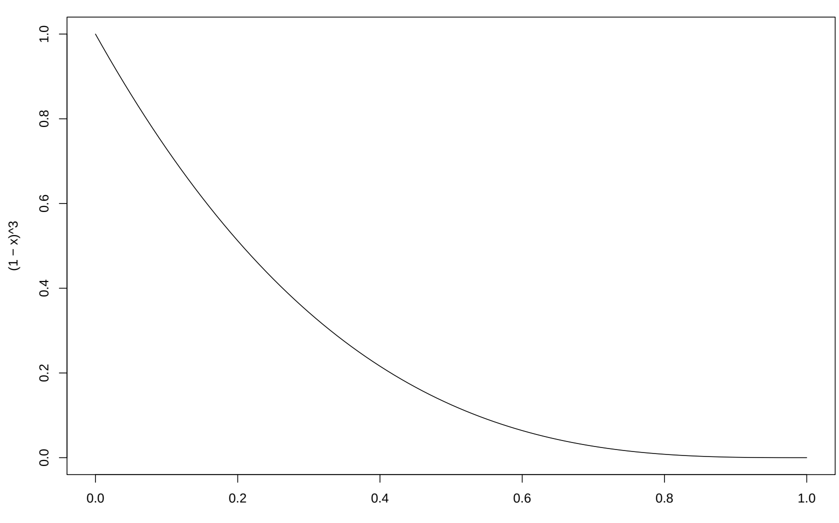
ottengo un risultato diverso e sbagliato perchè il massimo si ha sulla frontiera!

Se il campione estratto fosse stato $X_1 = 0, X_2 = 0, X_3 = 0$ la funzione di verosimiglainza sarebbe:

$$L(X_1 = 0, X_2 = 0, X_3 = 0, p) =$$

$$= P(X_1 = 0)P(X_2 = 0)P(X_3 = 0) = (1-p)^3 \quad 0 \leq p \leq 1$$

che è una funzione sempre decrescente



e quindi il massimo della funzione si ha per $\hat{p} = 0$

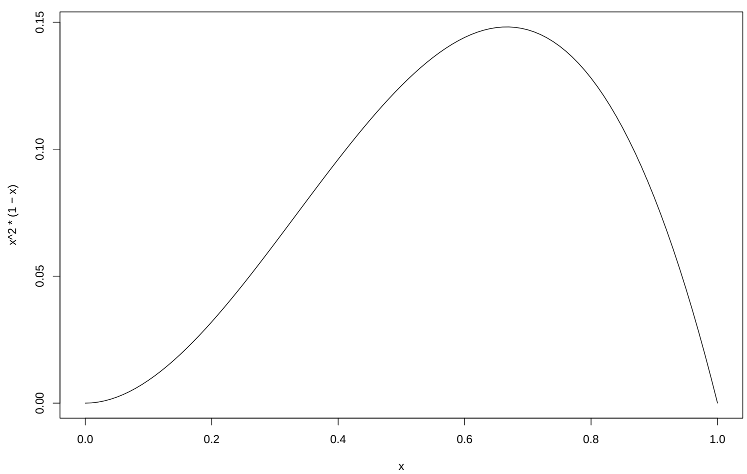
Se infine il campione estratto fosse stato

$X_1 = 1, X_2 = 0, X_3 = 1$ la funzione di verosimiglianza sarebbe:

$$L(X_1 = 1, X_2 = 0, X_3 = 1, p) =$$

$$= P(X_1 = 1)P(X_2 = 0)P(X_3 = 1) = p^2(1-p) \quad 0 \leq p \leq 1$$

il cui grafico è



il massimo non è sulla frontiera e lo si può trovare facendo la derivata prima e seconda della funzione di log-verosimiglianza:

$$l(p) = logL(p) = 2log(p) + log(1-p)$$

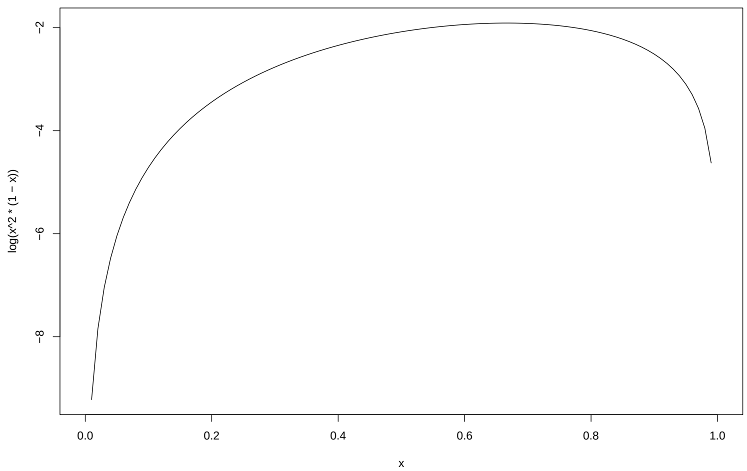
$$\frac{dl(p)}{dp} = 2/p - 1/(1-p) = 0$$

$$\frac{2(1-p) - p}{p(1-p)} = 0$$

$$2 - 2p - p = 0$$

$$p = 2/3$$

Notiamo che prendere il logaritmo di una funzione non sposta il punto di massimo (che è quello che a noi interessa), come si vede dal grafico in cui è riportato il logaritmo della funzione di verosimiglianza di prima



In generale per un campione $X_1, \cdots, X_n$ casuale estratto da una popolazione Bernoulliana, la funzione di verosimiglianza è

$$L(X_1, \cdots X_n, p) = p^{\sum X_i}(1-p)^{n-\sum X_i}$$

$$l(p) = log(p)(\sum X_i) + log(1-p)(n - \sum X_i)$$

$$\frac{dl(p)}{dp} = \frac{1}{p}(\sum X_i) - \frac{1}{1-p}(n - \sum X_i) = 0$$

e risolvendo per p otteniamo

$$\hat{p} = \frac{\sum_i X_i}{n}$$

la derivata seconda risulta sempre negativa quindi si può concludere che questo valore è effettivamente un massimo ed è quindi lo stimatore di massima verosimiglianza

Proprietà dello stimatore di massima verosimiglianza

- **Proprietà di invarianza**: se $\hat{\theta}$ è lo stimatore di massima verosimiglianza di θ, allora $g(\hat{\theta})$ è lo stimatore di massima verosimiglianza di $g(\theta)$.

- Sotto certe condizioni di regolarità, lo stimatore di massima verosimiglianza è:
  - *consistente*;
  - *asintoticamente efficiente; e non distorto*
  - ha *distribuzione limite normale*,

  Lo stimatore di Max Vero non è necessariamente unico

ESEMPIO: Vogliamo studiare il tempo di attesa prima che un macchinario si rompa

(a) Proponete un modello statistico, fra quelli studiati, per descrivere il tempo d'attesa

(b) Scrivete la funzione di verosimiglianza in corrispondenza ad un campione casuale $(x_1, \ldots, x_n)$.

(c) Immaginate ora che $(x_1, \ldots, x_n) = (20, 10, 5, 6, 4, 15)$. Calcolate la stima di massima verosimiglianza del tempo medio di attesa.

RISPOSTA

(a) Possiamo ipotizzare che il tempo d'attesa X abbia distribuzione esponenziale negativa,

$$f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{altrove} \end{cases}$$

Il tempo medio d'attesa è $E(X) = 1/\lambda$ e la varianza $V(X) = 1/\lambda^2$.

(b) La funzione di verosimiglianza in corrispondenza di un generico campione $x = (x_1, \ldots, x_n)$ è

$$L(\lambda; x) = \prod_{i=1}^{n} f(x_i; \lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda\, x_i} = \lambda^n e^{-\lambda\, \sum_{i=1}^{n} x_i}$$

(c) Massimizziamo la log-verosimiglianza

$$l(\lambda; x) = \log L(\lambda; x) = n \log \lambda - \lambda \sum_{i=1}^{n} x_i \; .$$

$$\frac{d}{d\lambda} l(\lambda, x) = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i \geq 0$$

per $\lambda \leq n / \sum_{i=1}^{n} x_i$. Dunque $\hat{\lambda} = 1/\bar{x} = 1/10$.

Per la proprietà di invarianza, la stima del valore atteso è $E(\hat{X}) = 1/\hat{\lambda} = \bar{x} = 10$.

Stimiamo ora la deviazione standard e la probabilità di aspettare fra 5 e 8 ore

$$\hat{P}(5 \leq X \leq 8) = \int_5^8 1/10 e^{-x/10} dx$$

ESEMPIO

In un processo di controllo di qualità , siamo interessati al numero mensile di guasti di un certo macchinario.

(a) Quale fra le distribuzioni di probabilità che avete studiato potreste proporre come modello statistico per descrivere il numero mensile di guasti? Spiegate cosa rappresenta il parametro del modello proposto.

(b) Supponete ora che in un campione casuale di 5 mesi si siano osservati $(3, 2, 4, 5, 1)$ guasti. Scrivete la funzione di verosimiglianza.

(c) Determinate infine la stima di massima verosimiglianza della probabilità di non avere nessun guasto.

RISPOSTA

(a) Possiamo proporre un modello di Poisson: $X \sim Pois(\lambda)$, ovvero

$$f(x; \lambda) = \begin{cases} \frac{e^{-\lambda}\,\lambda^x}{x!} & x = 0, 1, 2, \ldots \\ 0 & \text{altrove} \end{cases}$$

Il parametro λ rappresenta il numero medio di guasti in un mese, $\lambda = E(X)$ e anche la varianza $\lambda = V(X)$.

(b) La funzione di verosimiglianza è

$$L(\lambda; x_1, \dots, x_5) = \prod_{i=1}^{5} \frac{e^{-\lambda}\,\lambda^{x_i}}{x_i!} =$$

$$= \frac{1}{x_1! \cdots x_5!} e^{-5\lambda}\,\lambda^{\Sigma_{i=1}^{5} x_i} = \frac{1}{3!2!4!5!} e^{-5\lambda}\,\lambda^{15}\ .$$

(c) La probabilità di non avere guasti nel mese è data da

$$p = P(X = 0, \lambda) = p(0; \lambda) = e^{-\lambda}\ .$$

Possiamo determinare la stima di massima verosimiglianza $\hat{\lambda}$ di λ e poi utilizzare la proprietà di invarianza degli stimatori di massima verosimiglianza in base alla quale la stima di massima verosimiglianza di p è

$$\hat{p} = (\widehat{e^{-\lambda}}) = e^{-\hat{\lambda}}\ .$$

Facendo i calcoli, si trova che il massimo di $L(\lambda; x_1, \dots, x_5)$ si ha per $\hat{\lambda} = \frac{\Sigma_{i=1}^{5} x_i}{5} = 3$, da cui

$$\hat{p} = e^{-3} = 0.05$$

NOTA BENE:

- Lo stimatore di massima verosimiglianza non è necessariamente non distorto ma è asintoticamente non distorto

- Se devo stimare 2 (o più parametri) come nel caso di popolazione normale (μ e σ) faccio la derivata della funzione di verosimiglianza rispetto a ciascuno dei parametri ed uguaglio a zero. Ricavo così un sistema con 2 equazioni a 2 incognite (i 2 parametri)

- per popolazioni normali lo stimatore di massima verosimiglianza della media è la media campionaria (**non distorto**) mentre lo stimatore di massima verosimiglianza per la varianza è
$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$$
che risulta distorto ma **asintoticamente non distorto**

Metodo dei momenti

- Nel caso in cui $\boldsymbol{\theta}$ sia un parametro vettoriale di dimensione k, il metodo dei momenti consiste nel risolvere rispetto a $\boldsymbol{\theta}$ il sistema di equazioni

$$m_r = \mu_r(\boldsymbol{\theta}), \quad r = 1, \ldots, k,$$

dove

  - $\mu_r(\boldsymbol{\theta}) = E(X^r)$

    è il momento di ordine r della popolazione

  - $m_r = \frac{1}{n} \Sigma_i^n X_i^r$

    è il corrispondente momento campionario

Esempio

Ho un campione estratto da una popolazione Normale con media μ e varianza σ^2 (entrambi incogniti). Voglio trovare con il metodo dei momenti gli stimatori dei due parametri: μ e σ^2.

Risolvo allora il seguente sistema con due equazioni e due incognite:

$$\begin{cases} E(X) & = \frac{1}{n} \Sigma_i X_i \\ E(X^2) & = \frac{1}{n} \Sigma_i X_i^2 \end{cases}$$

cioè

$$\begin{cases} \mu & = \frac{1}{n} \Sigma_i X_i \\ \mu^2 + \sigma^2 & = \frac{1}{n} \Sigma_i X_i^2 \end{cases}$$

risolvendo il sistema si trova che gli stimatori ottenuti con il metodo dei momenti sono

$$\hat{\mu} = \bar{X}$$

e

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

cioè la media e la varianza del campione rispettivamente

Notiamo che lo stimatore ottenuto con il metodo dei momenti per la varianza è distorto (in generale il metodo dei momenti non garantisce che gli stimatori ottenuti siano non distorti! così come non lo garantisce il metodo della massima verosimiglianza) infatti

$$E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2$$

la distorsione è pari a

$$\mathrm{B}_\theta(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{\sigma^2}{n}$$

quindi la distorsione è sempre negativa cioè , in media, **sottostimo** il parametro di interesse

Abbiamo invece già visto che uno stimatore non distorto della varianza della popolazione è

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

Per un carattere X con distribuzione di Poisson di parametro λ lo stimatore per λ ottenuto con il metodo dei momenti $\hat{\lambda}_{MM}$ si ottiene nel modo seguente:

$$E(X) = \frac{1}{n} \sum_i X_1$$

ovvero, ricordando che, per questa distribuzione, $E(X) = \lambda$,

$$\lambda = \bar{X}$$

quindi

$$\hat{\lambda}_{MM} = \bar{X}$$

Notiamo che in questo caso $k = 1$, dove k è il numero di parametri da stimare, quindi per ottenere lo stimatore dei momenti basta una sola equazione (perchè ho una solo incognita, che è appunto λ)

Notiamo che in questo caso lo stimatore di massima verosimiglianza per λ coincide con lo stimatore ottenuto con il metodo dei momenti

Per un carattere X con distribuzione di Esponenziale di parametro λ lo stimatore per λ ottenuto con il metodo dei momenti $\hat{\lambda}_{MM}$ si ottiene nel modo seguente:

$$E(X) = \frac{1}{n} \sum_i X_1$$

ovvero

$$\frac{1}{\lambda} = \bar{X}$$

quindi

$$\hat{\lambda}_{MM} = \frac{1}{\bar{X}}$$

Notiamo che, anche in questo caso $k = 1$, dove k è il numero di parametri da stimare, quindi per ottenere lo stimatore dei momenti basta una sola equazione (perchè ho una solo incognita, che è appunto λ)

Notiamo che, anche in questo caso, lo stimatore di massima verosimiglianza per λ coincide con lo stimatore ottenuto con il metodo dei momenti

Per un carattere X con distribuzione di Bernoulliana di parametro p lo stimatore per p ottenuto con il metodo dei momenti $\hat{\lambda}_{MM}$ si ricava nel modo seguente:

$$E(X) = \frac{1}{n}\sum_i X_1$$

ovvero

$$p = \bar{X}$$

quindi

$$\hat{p}_{MM} = \bar{X}$$

Anche in questo caso $k = 1$ e, anche in questo caso, lo stimatore di massima verosimiglianza per p coincide con lo stimatore ottenuto con il metodo dei momenti

Per un carattere X con distribuzione di Binomiale di parametri (n, p) gli stimatori ottenuti con il metodo dei momenti si trovano risolvendo un sistema di due equazioni:

$$\begin{cases} E(X) & = \frac{1}{n} \Sigma_i X_i \\ E(X^2) & = \frac{1}{n} \Sigma_i X_i^2 \end{cases}$$

ovvero

$$\begin{cases} np & = \frac{1}{n} \Sigma_i X_i \\ np(1-p) + (np)^2 & = \frac{1}{n} \Sigma_i X_i^2 \end{cases}$$

Notiamo che, in questo caso, $k = 2$, dove k è il numero di parametri da stimare, quindi per ottenere lo stimatore dei momenti servono due equazioni (perchè ho due incognite, n e p)

Lascio a voi la soluzione del sistema rispetto ad n 4 p.