Inferenza statistica parametrica: stima puntuale

13 maggio 2019

Inferenza parametrica

Sia $X_1, \ldots, X_n$ è un campione di dimensione n estratto da una popolazione con distribuzione F_θ nota a meno di un parametro incognito (oppure un vettore di parametri incogniti) θ.

Esempi

- Sappiamo che le X_i sono v.a. con densità di Poisson:

$$f_\theta(k) = P(X_i = k) = e^{-\theta}\frac{\theta^k}{k!} \qquad k = 0, 1, \ldots$$

  ma non conosciamo il valore di θ.

- Sappiamo che le X_i sono v.a. gaussiane, quindi F_θ è assolutamente continua con densità

$$f_\theta(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

  ma il vettore di parametri $\theta = (\mu, \sigma^2)$ è incognito.

Definizione: caratteristica della popolazione

Sia $X_1, \ldots, X_n$ un campione aleatorio estratto da F_θ. Una **caratteristica della popolazione** è una funzione (non costante) di θ.

D'ora in poi indicheremo con $\mathbb{E}_\theta$, $\mathbb{V}\mathrm{ar}_\theta$, $\mathbb{P}_\theta$
media, varianza e probabilità calcolate per un θ ammissibile, (quindi in funzione di θ).

Esempio

- Le X_i sono v.a. gaussiane con vettore dei parametri $\theta = (\mu, \sigma^2)$ (media e varianza) incognito e
$$k(\theta) = \mathbb{E}_\theta(X_i^2) = \sigma^2 + \mu^2.$$

Problema

Dire qualcosa (FARE INFERENZA) sui parametri incogniti o su una caratteristica della popolazione, usando i dati (OSSERVAZIONI CAMPIONARIE).

Stima puntuale

Gli STIMATORI PUNTUALI sono chiamati così perché forniscono, per ogni realizzazione campionaria, un solo valore come "stima" del parametro incognito θ o di una sua funzione $k(\theta)$.

Nel seguito forniremo:

- la definizione di stimatore puntuale;
- la definizione di stima corrispondente, sulla base di una realizzazione campionaria;
- criteri per valutare la "bontà" di uno stimatore puntuale;
- due metodi per la costruzione di stimatori puntuali.

Sia $X_1, \ldots, X_n$ un campione aleatorio estratto da F_θ e sia $k(\theta)$ una **caratteristica della popolazione**.

Ricordiamo che $\mathbb{E}_\theta$, $\mathbb{V}\mathrm{ar}_\theta$, $\mathbb{P}_\theta$ sono media, varianza e probabilità calcolate in funzione di ogni θ ammissibile.

Definizione: stimatore

Chiamiamo stimatore di $k(\theta)$ una statistica $\hat{K}_n = d_n(X_1, \ldots, X_n)$ usata per fare inferenza su $k(\theta)$.

Definizione: stima

Data la realizzazione campionaria $x_1, \ldots, x_n$ e lo stimatore $\hat{K}_n = d_n(X_1, \ldots, X_n)$ di $k(\theta)$, chiamiamo stima di $k(\theta)$ il valore $\hat{k}_n = d_n(x_1, \ldots, x_n)$ della statistica $\hat{K}_n$ in corrispondenza delle osservazioni $x_1, \ldots, x_n$.

N.B. Lo stimatore è una statistica e quindi una v.a. (o vettore aleatorio), mentre la stima è un numero (o vettore di numeri).

Proprietà degli stimatori

Sia $X_1, \ldots, X_n$ un campione aleatorio estratto da una popolazione di densità f_θ nota a meno di un parametro θ e sia $k(\theta)$ una caratteristica reale della popolazione. Sia $D = d(X_1, \ldots, X_n)$ uno stimatore di $k(\theta)$.

Per definizione, uno stimatore è una qualunque statistica usata per fare inferenza su $k(\theta)$. Quindi è fondamentale stabilire strumenti e criteri per valutare l'efficacia di $D = d(X_1, \ldots, X_n)$ come stimatore di $k(\theta)$.

Criterio: valutare una distanza tra $D = d(X_1, \ldots, X_n)$ e $k(\theta)$, per esempio

$$r_\theta(d, k(\theta)) := \mathbb{E}_\theta\left[\left(d(X_1, \ldots, X_n) - k(\theta)\right)^2\right]$$

dove $\mathbb{E}_\theta$ sta ad indicare che la media è calcolata rispetto a f_θ, dipendente dal parametro θ.

Esempio. Se le X_i sono assolutamente continue con densità f_θ

$$r_\theta(d, k(\theta)) = \mathbb{E}_\theta\left[\Big(D - k(\theta)\Big)^2\right] = \mathbb{E}_\theta\left[\Big(d(X_1, \dots, X_n) - k(\theta)\Big)^2\right]$$
$$= \int_{\mathbb{R}^n} (d(x_1, \dots, x_n) - k(\theta))^2 f_\theta(x_1) \cdots f_\theta(x_n) dx_1 \cdots dx_n.$$

Scrivete l'espressione analoga per un campione discreto.

Definizione: Errore Quadratico Medio (o M.S.E.)

Sia $D = d(X_1, \dots, X_n)$ uno stimatore di $k(\theta)$ che ammette momento secondo finito. Si definisce errore quadratico medio (o mean square error) di $D = d(X_1, \dots, X_n)$ come stimatore di $k(\theta)$ la funzione (positiva) di θ

$$r_\theta(d, k(\theta)) := \mathbb{E}_\theta\left[\Big(d(X_1, \dots, X_n) - k(\theta)\Big)^2\right]$$

E' comunemente usata anche la notazione:

$$MSE_\theta(D) := \mathbb{E}_\theta\left[\Big(D - k(\theta)\Big)^2\right] = \mathbb{E}_\theta\left[\Big(d(X_1, \dots, X_n) - k(\theta)\Big)^2\right]$$

Trovare uno stimatore che minimizzi l'errore quadratico medio per ogni θ non è possibile.

Esempio. Se $k(\theta) = \theta$ e $d(X_1, \ldots, X_n) = 5$, allora $r_\theta(d, k(\theta)) = \mathbb{E}_\theta\left[\left(d(X_1, \ldots, X_n) - \theta\right)^2\right] = 0$ se $\theta = 5$ e quindi nessun altro stimatore fornisce un errore quadratico medio inferiore se $\theta = 5$.

Ma a volte si può trovare uno stimatore che minimizzi lo scarto quadratico medio in una classe ristretta di stimatori. Per esempio...

...una "buona" proprietà per uno stimatore è la non distorsione.

Definizione: Distorsione (o bias)

Sia $D = d(X_1, \ldots, X_n)$ uno stimatore di $k(\theta)$ che ammette media. Si definisce distorsione (o bias) di $D = d(X_1, \ldots, X_n)$ come stimatore di $k(\theta)$ la funzione di θ

$$b_\theta(d) := \mathbb{E}_\theta\Big[d(X_1, \ldots, X_n)\Big] - k(\theta)$$

Definizione: Stimatore non distorto o corretto

Lo stimatore $D = d(X_1, \ldots, X_n)$ si dice non distorto o corretto per $k(\theta)$ se, per ogni θ,

$$b_\theta(d) = \mathbb{E}_\theta\Big[d(X_1, \ldots, X_n)\Big] - k(\theta) = 0.$$

E' comunemente usata anche la notazione:

$$Bias_\theta(D) := \mathbb{E}_\theta\Big[d(\overrightarrow{X})\Big] - k(\theta)$$

Proposizione.

Se $D = d(X_1, \ldots, X_n)$ è uno stimatore di $k(\theta)$ basato sul campione $X_1, \ldots, X_n$ che ammette momento secondo finito, allora

$$r_\theta(d, k(\theta)) = \mathbb{V}\mathrm{ar}_\theta(d(X_1, \ldots, X_n)) + (b_\theta(d))^2.$$

... conseguenza: se $D = d(X_1, \ldots, X_n)$ è uno stimatore non distorto (i.e. $b_\theta(d) = 0$) allora l'errore quadratico medio coincide con la varianza. In simboli:

$$b_\theta(d) = 0 \quad \Longrightarrow \quad r_\theta(d, k(\theta)) = \mathbb{V}\mathrm{ar}_\theta(d(X_1, \ldots, X_n)).$$

Dimostrazione.

Ricorda $D = d(X_1, \ldots, X_n)$.

$$
\begin{aligned}
r_\theta(d, k(\theta)) =& \mathbb{E}_\theta\left[\left(D - k(\theta)\right)^2\right] \\
=& \mathbb{E}_\theta\left[\left(D - \mathbb{E}_\theta(D) + \mathbb{E}_\theta(D) - k(\theta)\right)^2\right] \\
=& \mathbb{E}_\theta\left[\left(D - \mathbb{E}_\theta(D) + b_\theta(d)\right)^2\right] \\
=& \mathbb{E}_\theta\left[\left(D - \mathbb{E}_\theta(D)\right)^2\right] + \left(b_\theta(d)\right)^2 \\
&+ 2b_\theta(d)\mathbb{E}_\theta\left[d(\overrightarrow{X}) - \mathbb{E}_\theta(d(\overrightarrow{X}))\right] \\
=& \mathbb{V}\mathrm{ar}_\theta(d(\overrightarrow{X})) + (b_\theta(d))^2
\end{aligned}
$$

in quanto l'ultimo addendo è uguale a zero. □

Proprietà asintotiche degli stimatori

Sia $X_1, X_2, \ldots$ una successione di v.a. i.i.d. con densità f_θ dipendente da un parametro incognito (o vettore di parametri incogniti) θ e sia $k(\theta)$ una caratteristica della popolazione. Sia inoltre $(D_n)_n = (d_n(X_1, \ldots, X_n))_n$ una successione di stimatori di $k(\theta)$ che ammettono momento secondo finito.

Definizione: Non distorsione asintotica

La successione di stimatori $(D_n)_n = (d_n(X_1, \ldots, X_n))_n$ è detta asintoticamente non distorta o corretta per $k(\theta)$ se

$$b_\theta(d_n) = \mathbb{E}_\theta\Big[d_n(X_1, \ldots, X_n)\Big] - k(\theta) = \mathbb{E}_\theta(D_n) - k(\theta) \to 0$$

per $n \to +\infty$, per ogni valore ammissibile del parametro θ.

Definizione: Consistenza debole.

Una successione di stimatori $(D_n)_n = (d_n(X_1, \dots, X_n))_n$ è detta debolmente consistente per $k(\theta)$ se per ogni $\epsilon > 0$

$$\mathbb{P}_\theta(\mid D_n - k(\theta) \mid > \epsilon) \to 0$$

per $n \to +\infty$, per ogni valore ammissibile del parametro θ.

Definizione: Consistenza in media quadratica.

La successione di stimatori $(D_n)_n = (d_n(X_1, \dots, X_n))_n$ è detta consistente in media quadratica per $k(\theta)$ se

$$r_\theta(d_n, k(\theta)) = \mathbb{E}_\theta\left[\left(D_n - k(\theta)\right)^2\right] \to 0$$

per $n \to +\infty$, per ogni valore ammissibile del parametro θ.

Conseguenze

Data una successione di campioni $(X_1, X_2, \dots)$ e $(D_n)_n = (d_n(X_1, \dots, X_n))_n$ una successione di stimatori di $k(\theta)$, allora:

1. Se $b_\theta(d_n) \to 0$ e $\mathbb{V}\mathrm{ar}_\theta(D_n) \to 0$ allora $r_\theta(d_n, k(\theta)) = \mathbb{V}\mathrm{ar}_\theta(D_n) + \Big(b_\theta(d_n)\Big)^2 \to 0$. Ovviamente vale anche il viceversa perché i due addendi sono positivi.
2. Se la successione di stimatori è consistente in media quadratica allora è consistente. Infatti per la disuguaglianza di Markov:

$$\mathbb{P}_\theta(\mid D_n - k(\theta) \mid > \epsilon) \leq \frac{1}{\epsilon^2}\mathbb{E}_\theta\Big[\Big(D_n - k(\theta)\Big)^2\Big] = \frac{r_\theta(d_n, k(\theta))}{\epsilon^2} \to 0$$

Esercizio. a) Mostrare che la successione delle medie campionarie campionarie $(\overline{X}_n)_n$ è sempre una successione di stimatori della media della popolazione non distorti e consistenti in media quadratica.
b) Mostrare che la successione delle varianze campionarie

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

è sempre una successione di stimatori della varianza della popolazione non distorti e, assumendo l'esistenza del momento quarto, consistenti in media quadratica.

Metodi per ottenere uno stimatore puntuale

In alcuni casi c'è un naturale candidato per uno stimatore puntuale.

Esempio. Sia X il numero di telefonate ad un numero verde in un giorno feriale. Abbiamo visto che X può essere modellizzato mediante una variabile di Poisson; supponiamo che il numero di telefonate in giorni differenti siano v.a. i.i.d..
La distribuzione di Poisson dipende da un parametro $\lambda > 0$ che rappresenta la media. Se $\lambda > 0$ è incognito e abbiamo a disposizione il numero di chiamate $x_1, \ldots, x_{30}$ al numero verde in 30 giorni (dati a disposizione), è naturale stimare λ con il valore

$$\hat{\lambda}_{30} = \frac{x_1 + \cdots + x_{30}}{30}$$

media aritmetica del numero delle chiamate nei 30 giorni.

Se $X_1, \dots, X_{30}$ sono le v.a. cosí definite: X_i rappresenta il numero di telefonate nell' i-esimo giorno ($i = 1, \dots, 30$), allora le X_i sono i.i.d. e $X_i \sim \mathit{Poiss}(\lambda)$. La stima sopra proposta per il parametro incognito λ non è altro che il valore assunto dallo stimatore media campionaria

$$\overline{X}_{30} = \frac{X_1 + \cdots + X_{30}}{30}$$

per la realizzazione campionaria $X_1 = x_1, \dots, X_{30} = x_{30}$, cioè

$$\hat{\lambda}_{30} = \overline{x}_{30} = \frac{x_1 + \cdots + x_{30}}{30}$$

Il fatto che la stima $\hat{\lambda}_{30}$ sia il valore assunto dalla statistica $\overline{X}_{30}$ ci permette di dire, per esempio, che tale stima è "mediamente corretta", nel senso che

$$\mathbb{E}_\lambda\left(\frac{X_1 + \cdots + X_{30}}{30}\right) = \lambda,$$

cioè lo stimatore è non distorto per il parametro λ.

Metodo dei momenti

Sia $X_1, \ldots, X_n$ un campione aleatorio estratto da una densità $f_\theta(x)$ discreta o assolutamente continua, $\theta = (\theta_1, \ldots, \theta_k)$ vettore di parametri incogniti.

Supponiamo esistano finiti i primi k momenti della densità $f_\theta(x)$. Indicato con μ_j il momento di ordine j di $f_\theta(x)$, tale momento risulta funzione del parametro incognito $\theta = (\theta_1, \ldots, \theta_k)$. Infatti:

$$\mu_j = \mathbb{E}_\theta(X_1^j) = \begin{cases} \int_{-\infty}^{+\infty} x^j f_\theta(x)dx & \text{se } X_1 \text{ è ass. continua} \\ \sum_h x_h^j f_\theta(x_h) & \text{se } X_1 \text{ è discreta} \end{cases}$$

N.B. $\mu_j = \mathbb{E}_\theta(X_1^j)$ significa che calcolo la media quando il valore del parametro è θ, e quindi è una caratteristica della popolazione. Scriviamo, per $j = 1, \ldots, k$,

$$\mu_j = \mu_j(\theta_1, \ldots, \theta_k).$$

Eguagliamo i primi k momenti campionari ai corrispondenti k momenti della popolazione:

$$\begin{cases} \frac{1}{n}\sum_{i=1}^{n} X_i = \mu_1(\theta_1, \ldots, \theta_k) \\ \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \mu_2(\theta_1, \ldots, \theta_k) \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} X_i^k = \mu_k(\theta_1, \ldots, \theta_k) \end{cases}$$

Ne risulta un sistema di k equazioni nelle k incognite $\theta_1, \ldots, \theta_k$. Supponiamo che tale sistema abbia una soluzione $\hat{\Theta}_1, \ldots, \hat{\Theta}_k$. Ovviamente ciascun $\hat{\Theta}_i$ è funzione di $X_1, \ldots, X_n$. In simboli:

$$\begin{cases} \hat{\Theta}_1 = d_1(X_1, \ldots, X_n) \\ \hat{\Theta}_2 = d_2(X_1, \ldots, X_n) \\ \vdots \\ \hat{\Theta}_k = d_k(X_1, \ldots, X_n) \end{cases}$$

per delle opportune funzioni $d_1, \ldots, d_k$.

Lo stimatore $\hat{\Theta} = (\hat{\Theta}_1, \ldots, \hat{\Theta}_k)$ cosí ottenuto è detto **stimatore del metodo dei momenti**.

Se osserviamo $X_1 = x_1, \ldots, X_n = x_n$, allora il valore:

$$(\hat{\theta}_1 = d_1(x_1, \ldots, x_n), \ldots, \hat{\theta}_k = d_k(x_1, \ldots, x_n))$$

è la **stima dei momenti** di $\theta = (\theta_1, \ldots, \theta_k)$ corrispondente alla realizzazione campionaria $(x_1, \ldots, x_n)$.

Esempio 1. Sia $X_1, \dots, X_n$ un campione aleatorio estratto da una popolazione gaussiana di media μ e varianza σ^2 incognite. Determinare lo stimatore dei momenti di $\theta = (\mu, \sigma^2)$.

Con la precedente notazione:

$$\mathbb{E}_\theta(X_1) = \mu_1(\mu, \sigma^2) = \mu \qquad \mathbb{E}_\theta(X_1^2) = \mu_2(\mu, \sigma^2) = \sigma^2 + \mu^2.$$

Il sistema è:

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

che ha come unica soluzione

$$\begin{cases} \hat{\Theta}_1 = \overline{X}_n \\ \hat{\Theta}_2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - (\overline{X}_n)^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X}_n)^2. \end{cases}$$

Esempio 2. Sia $X_1, \ldots, X_n$ un campione aleatorio estratto da una popolazione gamma di parametro di forma α e di scala λ incogniti. Determinare lo stimatore dei momenti di $\theta = (\alpha, \lambda)$.

Con la precedente notazione:

$$\mathbb{E}_\theta(X_1) = \mu_1(\alpha, \lambda) = \frac{\alpha}{\lambda} \qquad \mathbb{E}_\theta(X_1^2) = \mu_2(\alpha, \lambda) = \frac{\alpha(\alpha+1)}{\lambda^2}.$$

Il sistema è:

$$\begin{cases} \frac{1}{n}\sum_{i=1}^n X_i = \frac{\alpha}{\lambda} \\ \frac{1}{n}\sum_{i=1}^n X_i^2 = \frac{\alpha(\alpha+1)}{\lambda^2} \end{cases}$$

che ha come unica soluzione

$$\begin{cases} \hat{\Theta}_1 = \frac{\overline{X}_n^2}{\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2} \\ \hat{\Theta}_2 = \frac{\overline{X}_n}{\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2}. \end{cases}$$

Stimatori di massima verosimiglianza o M.L.E.

Uno stimatore di massima verosimiglianza si ottiene col seguente ragionamento.

Sia f_θ una densità discreta o assolutamente continua dipendente da un parametro incognito θ (o da un vettore di parametri incogniti $\theta = (\theta_1, \dots, \theta_k)$). Se $X_1, \dots, X_n$ è un campione aleatorio estratto da f_θ, la densità del campione è

$$L(\theta, \mathbf{x}) = \prod_{i=1}^{n} f_\theta(x_i)$$

per ogni $\mathbf{x} = (x_1, \dots, x_n)$, essendo le X_i v.a. i.i.d..
Supponiamo di osservare i dati $X_1 = \tilde{x}_1, \dots, X_n = \tilde{x}_n$ e consideriamo la funzione di θ, detta verosimiglianza del campione o likelihood,

$$L(\theta, \tilde{\mathbf{x}}) = \prod_{i=1}^{n} f_\theta(\tilde{x}_i)$$

dove $\tilde{\mathbf{x}} = (\tilde{x}_1, \dots, \tilde{x}_n)$.

Se $X_1, \ldots, X_n$ sono v.a. discrete allora

$$L(\theta, \tilde{\mathbf{x}}) = \prod_{i=1}^{n} f_\theta(\tilde{x}_i) = \prod_{i=1}^{n} \mathbb{P}_\theta(X_i = \tilde{x}_i) = \mathbb{P}_\theta(X_1 = \tilde{x}_1, \ldots, X_n = \tilde{x}_n),$$

dove $\mathbb{P}_\theta$ è la probabilità calcolata in corrispondenza del valore θ del parametro. Quindi la funzione $L(\theta, \tilde{\mathbf{x}})$ rappresenta, al variare di θ, la probabilità di osservare i valori $\tilde{x}_1, \ldots, \tilde{x}_n$ per il campione quando θ è il valore del parametro.

Quando il campione è assolutamente continuo

$$L(\theta, \tilde{\mathbf{x}}) = \prod_{i=1}^{n} f_\theta(\tilde{x}_i)$$

non è più una probabilità, ma può essere interpretata come la “verosimiglianza” (o credibilità) di osservare la n-upla di dati $\tilde{x}_1, \ldots, \tilde{x}_n$ per ogni valore θ fissato del parametro.

È quindi ragionevole prendere come stima di θ quel valore che massimizza la “verosimiglianza” per i dati osservati.

La stima di massima verosimiglianza è definita come quel valore $\hat{\theta}_n$ che massimizza la verosimiglianza per ogni realizzazione del campione $\mathbf{x} = (x_1, \ldots, x_n)$ e quindi è una funzione di $(x_1, \ldots, x_n)$.

Definizione: M.L.E (Maximum Likelihood Estimator)

La **stima di massima verosimiglianza o stima M.L.** in corrispondenza dell'osservazione $\mathbf{x} = (x_1, \ldots, x_n)$ è

$$\hat{\theta}_n = \underset{\theta}{argmax}\ L(\theta, \mathbf{x}) = \underset{\theta}{argmax} \prod_{i=1}^{n} f_\theta(x_i)$$

Ovviamente $\hat{\theta}_n$ è una funzione di $(x_1, \ldots, x_n)$, diciamo $\hat{\theta}_n = t_n(x_1, \ldots, x_n)$ per una opportuna funzione t_n. Il corrispondente stimatore

$$\hat{\Theta}_n = t_n(X_1, \ldots, X_n)$$

è detto **stimatore di massima verosimiglianza o M.L.E.**.

Operativamente, in molti casi, si usa il fatto che $L(\theta, \mathbf{x})$ e $\ln L(\theta, \mathbf{x})$ assumono il massimo in corrispondenza dello stesso valore. Quindi per ottenere la stima si massimizza $\ln L(\theta)$ detta funzione di **log-verosimiglianza** (log-likelihood).

Supponiamo ora di voler stimare una caratteristica $k(\theta)$.

Principio d'invarianza degli M.L.E.

Se $\hat{\Theta}_n = t_n(X_1, \dots, X_n)$ è l'M.L.E. di θ basato sul campione $X_1, \dots, X_n$ estratto da f_θ, allora per ogni funzione k, l' **M.L.E.** di $k(\theta)$ è

$$\hat{K}_n := k(\hat{\Theta}_n) = k(t_n(X_1, \dots, X_n)).$$

ESEMPI DI STIMATORI E STIME M.L.

Es 1. Stima M.L.E. del parametro di una popolazione bernoulliana. Sia $X_1, \dots, X_n$ un campione aleatorio estratto da un popolazione bernoulliana di parametro p incognito. Con le notazioni prima introdotte:

$$f_p(x) = p^x(1-p)^{1-x}\mathbb{I}_{\{0,1\}}(x)$$

Data un'osservazione campionaria $(x_1, \dots, x_n)$, posso supporre che $\mathbb{I}_{\{0,1\}}(x_i) = 1$ per ogni i (in quanto osservazioni da variabili di Bernoulli) e quindi

$$L(p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}$$

$$\ln L(p) = (\sum_{i=1}^{n} x_i)\ ln(p) + (n - \sum_{i=1}^{n} x_i)\ ln(1-p).$$

Supponiamo $\sum_{i=1}^{n} x_i \neq 0, n$. Cerchiamo il punto di massimo della log-verosimiglianza $\ln L(p)$ (e quindi anche della verosimiglianza $L(p)$) risolvendo l'equazione:

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

che ammette l'unica soluzione

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

che è anche un punto di massimo. Possiamo quindi concludere che $\hat{p}_n$ è la stima M.L. di p. Si vede facilmente che $\hat{p}_n$ è la stima M.L. di p anche se $\sum_{i=1}^{n} x_i = 0$ oppure n.

...in conclusione: la stima M.L. di p è

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

che corrisponde allo stimatore M.L. di p

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

Per il Principio d'invarianza l'M.L.E. di $k(p) = p(1-p)$ è quindi

$$\hat{K}_n = k(\hat{P}_n) = \overline{X}_n(1 - \overline{X}_n) = \frac{n-1}{n} S_n^2$$

N.B. $k(p) = p(1-p) = \mathbb{V}\mathrm{ar}_p(X_1)$, dove con l'ultima notazione intendo la varianza calcolata al variare dei valori possibili del parametro p.

Es. 2. Stima M.L. della media e della deviazione standard di una popolazione gaussiana. Sia $X_1, \ldots, X_n$ un campione aleatorio estratto da una popolazione gaussiana di media μ e varianza σ^2 incognite. Sia $\theta = (\mu, \sigma)$.

$$L(\theta) = L(\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}$$
$$= \left(\frac{1}{2\pi}\right)^{n/2} \frac{1}{\sigma^n} e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2}.$$

e

$$\ln L(\mu, \sigma) = -\frac{n}{2}\,\ln 2\pi - n\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

Cerchiamo il punto di massimo come soluzione del sistema

$$\begin{cases} \frac{\partial}{\partial \mu} \ln L(\mu, \sigma) = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) = 0 \\ \frac{\partial}{\partial \sigma} \ln L(\mu, \sigma) = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{n} (x_i - \mu)^2 = 0. \end{cases}$$

L'unica soluzione che è anche un punto di massimo (verificarlo!) è

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i \qquad \hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\mu})^2}.$$

Quindi gli M.L.E. di μ e σ sono, rispettivamente,

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \qquad \sqrt{\frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2} = \sqrt{\frac{n-1}{n} S_n^2}.$$

Es. 3. Stima M.L. della media di una popolazione uniforme su un intervallo. Sia $X_1, \ldots, X_n$ un campione aleatorio estratto da una popolazione uniforme sull'intervallo $[0, \theta]$, con $\theta > 0$ parametro incognito.

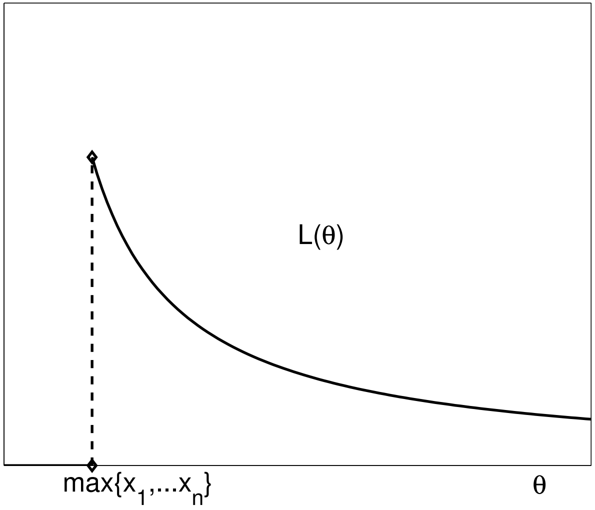
$$
\begin{aligned}
f_\theta(x) &= \frac{1}{\theta}\, \mathbb{I}_{[0,\theta]}(x) \\
L(\theta) &= \prod_{i=1}^{n} \frac{1}{\theta}\, \mathbb{I}_{[0,\theta]}(x_i) = \frac{1}{\theta^n} \prod_{i=1}^{n} \mathbb{I}_{[0,\theta]}(x_i) \\
&= \frac{1}{\theta^n} \mathbb{I}_{[\max\{x_1,\ldots,x_n\}),\infty)}(\theta).
\end{aligned}
$$

Infatti

$$
\prod_{i=1}^{n} \mathbb{I}_{[0,\theta]}(x_i) = 1 \;\Leftrightarrow\; x_i \leq \theta \quad \forall i = 1, \ldots, n \;\Leftrightarrow \theta \geq \max\{x_1, \ldots, x_n\}
$$

altrimenti è uguale a zero.

Grafico di $L(\theta) = \frac{1}{\theta^n} \mathbb{I}_{[\max\{x_1,\dots,x_n\}),\infty)}(\theta)$



Quindi la stima M.L.E. è $\hat{\theta}_n = \max\{x_1, \ldots, x_n\}$ e il corrispondente stimatore è

$$\hat{\Theta}_n = \max\{X_1, \ldots, X_n\}.$$

Poiché la media della popolazione è $k(\theta) = \frac{\theta}{2}$ segue che lo stimatore M.L.E per la media è

$$k(\hat{\Theta}_n) = \frac{\max\{X_1, \ldots, X_n\}}{2}.$$

Esercizio. Calcolare la densità dello stimatore M.L.E. $\hat{\Theta}_n$.

Esercizio. Determinare le proprietà dello stimatore M.L.E. del parametro di una densità di Poisson.

Soluzione. Sia $X_1, \ldots, X_n$ un campione di dimensione n estratto da una popolazione di Poisson di parametro λ incognito, cioè con densità:

$$f_\lambda(x) = e^{-\lambda} \frac{\lambda^x}{x!} \mathbb{I}_{\{0,1,\ldots\}}(x).$$

Lo stimatore M.L.E. è la media campionaria $\overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ (dimostratelo per esercizio). Al variare di n, la successione degli stimatori cosí ottenuti ha le seguenti proprietà.

- $\mathbb{E}_\lambda(\overline{X}_n) = \lambda$ per ogni λ, i.e. è una successione di stimatori non distorti;
- Per la L.D.G.N. vale che $\lim_{n \to +\infty} \mathbb{P}_\lambda(|\overline{X}_n - \lambda| > \epsilon) = 0$, per ogni $\epsilon > 0$, i.e. la successione di stimatori $(\overline{X}_n)_n$ è consistente per λ;

- Essendo gli stimatori non distorti, l'errore quadratico medio coincide con la varianza. Inoltre, $\mathbb{V}\mathrm{ar}_\lambda(\overline{X}_n) = \frac{\lambda}{n} \to 0$ per $n \to +\infty$. Quindi la successione di stimatori è anche consistente in media quadratica per λ;
- Infine per il T.C.L. vale, per n "grande", $\overline{X}_n \approx \mathbb{N}(\lambda, \lambda/n)$. Più precisamente:

$$\lim_{n\to+\infty} \mathbb{P}_\lambda\left(\frac{\overline{X}_n - \lambda}{\sqrt{\lambda/n}} \leq t\right) = \Phi(t).$$

Queste proprietà non valgono solo per gli stimatori M.L.E. del parametro di una Poisson.

Perché gli M.L.E. sono molto usati?

Sia $X_1, X_2, \dots$ succ. di campioni estratti da f_θ con $\theta \in \mathbb{R}$ incognito.

"Teorema"

Se la densità f_θ soddisfa "opportune condizioni di regolarità", se $\hat{\Theta}_n = t_n(X_1, \dots, X_n)$, $n = 1, 2, \dots$, è la successione degli M.L.E. di θ e $k(\theta)$ è una funzione differenziabile di θ, allora la successione $(k(\hat{\Theta}_n))_n$ degli M.L.E. di $k(\theta)$ è:

1. asintoticamente non distorta, i.e. $\lim_{n\to+\infty} \mathbb{E}_\theta\left[k(\hat{\Theta}_n)\right] = k(\theta)$;
2. consistente in media quadratica, i.e.
   $$\lim_{n\to+\infty} \mathbb{E}_\theta\left[\left(k(\hat{\Theta}_n) - k(\theta)\right)^2\right] = 0$$
3. asintoticamente gaussiana di media $k(\theta)$ e varianza $\frac{\sigma^2(\theta)}{n} = \frac{(k'(\theta))^2}{n\mathbb{E}_\theta[(\frac{\partial}{\partial\theta}\ln(f_\theta(X_1)))^2]}$, i.e.

$$\lim_{n\to+\infty} \mathbb{P}_\theta\left(\frac{k(\hat{\Theta}_n) - k(\theta)}{\sigma(\theta)/\sqrt{n}} \leq t\right) = \Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-u^2/2} du.$$

Esercizio. Mostrare che nel caso di una popolazione con densità f_λ di Poisson con parametro incognito λ, la varianza asintotica degli stimatori M.L.E. di λ (cioè di $\overline{X}_n,\ n = 1, 2, \dots$), ottenuta con la formula al punto 3, è $\sigma^2(\lambda)/n = \lambda/n$.

Soluzione. In questo caso $k(\lambda) = \lambda$ e quindi $k'(\lambda) = 1$. Inoltre $\ln(f_\lambda(x)) = \ln(e^{-\lambda}\frac{\lambda^x}{x!}) = -\lambda + x\ln(\lambda) - \ln(x!)$ e quindi

$$\mathbb{E}_\lambda\Big[\Big(\frac{\partial}{\partial\lambda}\ln(f_\lambda(X_1))\Big)^2\Big] = \mathbb{E}_\lambda\Big[\Big(\frac{X_1}{\lambda} - 1\Big)^2\Big] = \frac{1}{\lambda^2}\mathbb{E}_\lambda\Big[\Big(X_1 - \lambda\Big)^2\Big] = \frac{\mathbb{V}\mathrm{ar}_\lambda(X_1)}{\lambda^2} = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}.$$

Quindi

$$\frac{\sigma^2(\lambda)}{n} = \frac{1}{n\mathbb{E}_\lambda\Big[\Big(\frac{\partial}{\partial\lambda}\ln(f_\lambda(X_1))\Big)^2\Big]} = \frac{1}{n/\lambda} = \frac{\lambda}{n}.$$

Osserviamo, che il Teorema precedente sulle proprietà degli M.L.E. non si applica, per esempio, al caso di una popolazione con densità uniforme su un intervallo $[0, \theta]$, cosí come a tutte le densità che hanno il supporto dipendente dal parametro (cioè il parametro compare nell'indicatore). Senza entrare nei dettagli, il motivo è che in questi casi non sono soddisfatte le "opportune condizioni di regolarità", ipotesi del Teorema.