ANALISI 2 | ma 12/5/20

Stimatori di massima verosimiglianza

Dato un modello statistico parametrico $(\Omega, \mathcal{A}, \{P^\theta : \theta \in D\})$, sia $(X_1, \dots, X_N)$ un campione statistico di taglia N, e siano $x_1 = X_1(\omega), \dots, x_N = X_N(\omega)$ i valori osservati. Supponiamo per comodità che le X_j abbiano densità (discreta o continua) f_θ rispetto alla probabilità P^θ; se F_θ è la densità congiunta di $X_1, \dots, X_N$, grazie all'indipendenza si ha

$$F_\theta(x_1, \dots, x_N) = f_\theta(x_1) \cdot f_\theta(x_2) \cdot \dots \cdot f_\theta(x_N).$$

Si chiama funzione di verosimiglianza la funzione

$$\theta \mapsto \mathcal{M}(\theta; x_1, \dots, x_N) = F_\theta(x_1, \dots, x_N)$$

definita su D, avente $x_1, \dots, x_N$ come parametri.
Il nome deriva dal fatto che, nel caso di densità discrete, risulta

$$\mathcal{M}(\theta; x_1, \dots, x_N) = P^\theta(X_1 = x_1, \dots, X_N = x_n),$$

e quindi tale funzione misura, secondo P^θ, la probabilità che il campione assuma i valori $x_1, \dots, x_N$: cioè, ci dice quanto tali valori siano plausibili, ovvero verosimili.

Volendo stimare il parametro θ della legge del campione, sarà allora naturale adottare come stima di θ il valore $\hat{\theta}$ che rende massima, rispetto a θ, la funzione di verosimiglianza quando i valori assunti dai parametri sono $x_1, \dots, x_N$.

<u>Definizione</u> Supponiamo che la funzione M, per ogni scelta dei parametri $x_1, \dots, x_N$, abbia un unico punto di massimo assoluto $\hat{\theta}(x_1, \dots, x_N)$. Allora la statistica

$$T = \hat{\theta}(X_1, \dots, X_N)$$

è detta <u>stimatore di massima verosimiglianza</u> di θ.

Si noti che $\theta \in D$, e D è un sottoinsieme di $\mathbb{R}^k$. Dunque $\theta = (\theta_1, \dots, \theta_k)$ e $T = (T_1, \dots, T_k)$ è un vettore aleatorio. Le componenti T_j saranno stimatori del parametro θ_j, $j = 1, \dots, k$.

In generale lo stimatore di massima verosimiglianza può non esistere, se $M(\theta; x_1, \dots, x_N)$ non ha massimo per qualche scelta dei parametri, o non essere unico, se $\mathcal{M}$ ha più di un punto di massimo assoluto. Però, usualmente, se N è sufficientemente grande questi problemi non si presentano. In tal caso, se D è un aperto di $\mathbb{R}^k$ e se $\mathcal{M}$ è di classe C^1 rispetto a ϑ, la ricerca dello stimatore di massima verosimiglianza si può fare con i metodi dell'analisi.

Osserviamo che, essendo $\mathcal{M}$ spesso il prodotto di densità marginali, conviene massimizzare $\ln \mathcal{M}$ in luogo di $\mathcal{M}$, tenendo conto del fatto che

$$\frac{\partial}{\partial \vartheta_j} \ln \mathcal{M} = \frac{1}{\mathcal{M}} \frac{\partial}{\partial \vartheta_j} \mathcal{M}.$$

<u>Esempio</u> Stimiamo il parametro della legge di Poisson: dato un campione $(X_1, \dots, X_N)$ con legge $P(\lambda)$, ove λ è sconosciuto, la funzione di verosimiglianza è

$$\mathcal{M}(\lambda; x_1, \dots, x_N) = P^\lambda(X_1 = x_1, \dots, X_N = x_N) =$$
$$= e^{-\lambda} \frac{\lambda^{x_1}}{x_1!} \cdot \ldots \cdot e^{-\lambda} \frac{\lambda^{x_N}}{x_N!} = e^{-N\lambda} \prod_{j=1}^N \frac{\lambda^{x_j}}{x_j!}.$$

Dunque

$$\ln \mathcal{M}(\lambda, x_1, \dots, x_N) = -N\lambda + \sum_{j=1}^N x_j \ln \lambda - \sum_{j=1}^N \ln x_j!,$$

da cui

$$\frac{d}{d\lambda} \ln \mathcal{M}(\lambda; x_1, \dots, x_N) = -N + \frac{1}{\lambda} \sum_{j=1}^N x_j = 0 \iff \lambda = \frac{1}{N} \sum_{j=1}^N x_j.$$

Dunque lo stimatore di massima verosimiglianza è

$$\hat{\lambda}(X_1,\dots,X_N) = \frac{1}{N}\sum_{j=1}^{N} X_j = \overline{X},$$

che coincide con la media empirica: quindi è uno stimatore corretto.

Esempio La tabella sottostante mostra il numero di incidenti stradali in 24 giornate serene del 2019 nella località di Paperopoli.

0 3 0 0 6 4 5 1 0 1 1 3 2 6 0 7 4 3 1 1 0 1 2 1.

Vogliamo stimare la probabilità di giornate serene con meno di tre incidenti a Paperopoli nel 2019.

Poichè vi è un numero elevato di automobilisti, con una bassa probabilità di incidenti, è ragionevole assumere che il numero di incidenti in una giornata serena sia una v.a. di Poisson.

Lo stimatore di massima verosimiglianza è la media empirica: dunque il parametro λ può essere stimato come

$$\overline{\lambda} = \frac{1}{24}\sum_{i=1}^{24} x_i = \frac{52}{24} = 2.1\overline{6} \simeq 2.17.$$

Dunque, il numero X di incidenti in una giornata serena avrà legge di Poisson, approssimativamente $P(\overline{\lambda})$: pertanto

$$P(X<3) = P(X\le 2) = e^{-\overline{\lambda}}\sum_{j=0}^{2}\frac{\overline{\lambda}^j}{j!} = \left[1+2{,}17+\frac{(2.17)^2}{2}\right]e^{-2.17} \simeq 0.63.$$

Quindi, secondo la nostra stima, la probabilità che vi siano meno di 3 incidenti in una giornata serena è poco più del 60%.

Esempio Stimiamo il parametro della legge geometrica: sia $X_1 \cdots X_N$ un campione dotato di legge geometrica $G(\theta)$; la funzione di verosimiglianza è

$$\mathcal{M}(\theta; x_1, \ldots, x_N) = P^\theta(X_1 = x_1, \ldots, X_N = x_N) =$$
$$= \theta(1-\theta)^{x_1 - 1} \cdot \ldots \cdot \theta(1-\theta)^{x_N - 1} = \theta^N (1-\theta)^{\sum_{j=1}^N x_j - N} .$$

Dunque

$$\ln \mathcal{M}(\theta; x_1, \ldots, x_N) = N \ln\theta + \left(\sum_{j=1}^N x_j - N\right) \ln(1-\theta) ,$$

e pertanto

$$\frac{d}{d\theta} \ln \mathcal{M}(\theta; x_1, \ldots, x_N) = \frac{N}{\theta} - \frac{\sum_{j=1}^N x_j - N}{1-\theta} = \frac{N - \theta \sum_{j=1}^N x_j}{\theta(1-\theta)} = 0$$

se e solo se

$$\theta = \frac{N}{\sum_{j=1}^N x_j} .$$

Dunque lo stimatore di massima verosimiglianza è

$$T = \hat{\theta}(X_1 \ldots X_N) = \frac{N}{\sum_{j=1}^N X_j} = \frac{1}{\overline{X}} ,$$

cioè il reciproco della media empirica.

Esempio Stimiamo il parametro della legge esponenziale: se $X_1 \ldots X_N$ è un campione statistico con legge $\mathcal{E}(\lambda)$, la funzione di verosimiglianza è

$$\mathcal{M}(\lambda; x_1 \ldots x_N) = f_\lambda(x_1) \cdot \ldots \cdot f_\lambda(x_N) = \lambda^N e^{-\lambda \sum_{j=1}^N x_j} ,$$

da cui

$$\ln M(\lambda; x_1 \cdots x_N) = N \ln \lambda - \lambda \sum_{j=1}^{N} x_j \,,$$

e dunque

$$\frac{d}{d\lambda} \ln M(\lambda; x_1 \cdots x_N) = \frac{N}{\lambda} - \sum_{j=1}^{N} x_j = 0 \iff \lambda = \frac{N}{\sum_{j=1}^{N} x_j} \,.$$

Si trova nuovamente, come stimatore di massima verosimiglianza, la statistica

$$T = \bar{\lambda}(X_1 \cdots X_N) = \frac{N}{\sum_{j=1}^{N} X_j} = \frac{1}{\bar{X}}$$

reciproca della media empirica.

<u>Esempio</u> (Stima per la media e la varianza della legge normale)

Sia $X_1, \dots, X_N$ un campione con legge $N(\mu, \sigma^2)$. La funzione di verosimiglianza è

$$M(\mu, \sigma; x_1 \cdots x_N) = \frac{1}{[\sqrt{2\pi}\,\sigma]^n} \prod_{j=1}^{N} e^{-\frac{(x_j-\mu)^2}{2\sigma^2}} \,.$$

Ne segue

$$\ln M(\mu, \sigma; x_1 \cdots x_N) = -n \ln \sqrt{2\pi} - n \ln \sigma - \sum_{j=1}^{N} \frac{1}{2\sigma^2}(x_j - \mu)^2 \,,$$

e dunque:

$$\frac{\partial}{\partial\mu} \ln M(\mu, \sigma; x_1 \cdots x_N) = +\frac{1}{\sigma^2} \sum_{j=1}^{N} (x_j - \mu) = \frac{1}{\sigma^2}\left[\sum_{j=1}^{N} x_j - N\mu\right] ,$$

$$\frac{\partial}{\partial\sigma} \ln M(\mu, \sigma; x_1 \cdots x_N) = -\frac{N}{\sigma} + \frac{1}{\sigma^3} \sum_{j=1}^{N} (x_j - \mu)^2 \,.$$

Le due derivate si annullano per $\hat{\mu} = \frac{1}{N} \sum_{j=1}^{N} x_j$ e $\hat{\sigma} = \sqrt{\frac{1}{N} \sum_{j=1}^{N} (x_j - \hat{\mu})^2}$.

Lo stimatore di massima verosimiglianza per la media è

$$T_1 = \hat{\mu}(X_1, \dots, X_N) = \bar{X},$$

mentre quello per la varianza è

$$T_2 = \hat{\sigma}^2(X_1, \dots, X_N) = \frac{1}{N} \sum_{j=1}^{N} (X_j - \bar{X})^2 = \frac{N-1}{N} S^2 =: \Sigma^2 .$$

La statistica Σ^2 è uno stimatore distorto, essendo

$$\mathbb{E}^{\mu,\sigma}[\Sigma^2] = \frac{N-1}{N} \mathbb{E}^{\mu,\sigma}[S^2] = \frac{N-1}{N} \sigma^2 .$$

Come si vede, non è detto che gli stimatori di massima verosimiglianza siano corretti. Tuttavia essi godono di buone proprietà asintotiche: se $N \to \infty$, essi approssimano il parametro meglio di tutti gli altri. Non è importante, insomma, se lo stimatore che minimizza il rischio è corretto o no.

Stimatore dei momenti

Descriviamo un metodo ulteriore per ottenere uno stimatore del parametro $\psi(\theta)$, molto utile nella pratica anche se non troppo preciso in certi casi. Sia $X_1 \dots X_N$ un campione statistico di taglia N e legge con densità f_θ (continua o discreta). I momenti di ordine $r \in \mathbb{N}^+$ sono le quantità (uguali per ciascuna delle X_j)

$$m_r(\theta) = \mathbb{E}^\theta [X_j^r];$$

in particolare $m_1(\theta)$ è la media $\mathbb{E}^\theta [X_1]$, mentre

$$m_2(\theta) - m_1(\theta)^2 = \mathrm{Var}^\theta [X_1] .$$

Supponiamo che la funzione $\psi(\theta)$ si possa scrivere nella forma

$$\psi(\theta) = g(m_1(\theta), \dots, m_r(\theta));$$

come abbiamo visto, ciò accade per la media e per la varianza.
Definiamo il <u>momento empirico di ordine r</u>: essa è la statistica

$$M_r = \frac{1}{N} \sum_{j=1}^{N} X_j^r .$$

Si tratta di uno stimatore corretto del momento $m_r(\theta)$, perché

$$\mathbb{E}^\theta[M_r] = \frac{1}{N} \sum_{j=1}^{N} \mathbb{E}^\theta[X_j^r] = \frac{1}{N} N\, m_r(\theta) = m_r(\theta).$$

Il <u>metodo dei momenti</u> consiste nello stimare il parametro $\psi(\theta) = g(m_1(\theta), \dots, m_r(\theta))$ mediante la statistica

$$T = g(M_1, \dots, M_r),$$

che è chiamata <u>stimatore dei momenti</u>.
Come sappiamo, prendendo $\psi(\theta) = m_1(\theta) = \mu$ si ottiene $T = M_1 = \overline{X}$, mentre, prendendo $\psi(\theta) = m_2(\theta) - m_1(\theta)^2 = \sigma^2$, si ottiene $T = M_1 - M_2^2 = \overline{X} - \frac{N-1}{N} S^2$; si ritrovano così i due stimatori di massima verosimiglianza per la media e la varianza di campioni con leggi normali.

<u>Esempio</u> Sia $X_1, \dots, X_N$ un campione di taglia N con legge esponenziale $\mathcal{E}(\lambda)$: allora $m_1(\lambda) = \frac{1}{\lambda}$, e dunque $\lambda = \frac{1}{m_1(\lambda)}$.
Perciò lo stimatore dei momenti del parametro λ è

$$\Lambda = \frac{1}{M_1} = \frac{1}{\overline{X}},$$

cioè è lo stesso stimatore ottenuto col metodo della massima verosimiglianza.

Esempio (stima dei parametri della legge gamma).

Sia $X_1 \dots X_N$ un campione di taglia N con legge $\gamma(\alpha, \lambda)$. Si sa allora che

$$m_1(\alpha, \lambda) = \frac{\alpha}{\lambda}, \qquad m_2(\alpha, \lambda) = \frac{\alpha^2 + \alpha}{\lambda^2},$$

da cui, risolvendo rispetto a α e λ,

$$\alpha = \frac{m_1^2}{m_2 - m_1^2}, \qquad \lambda = \frac{m_1}{m_2 - m_1^2}.$$

Quindi gli stimatori dei momenti di α e λ sono

$$A = \frac{M_1^2}{M_2 - M_1^2}, \qquad \Lambda = \frac{M_1}{M_2 - M_1^2}.$$

Ricordando che $\bar{X} = M_1$ e $S^2 = \frac{N}{N-1}(M_2 - M_1^2)$, possiamo scrivere

$$A = \frac{N \bar{X}^2}{(N-1) S^2}, \qquad \Lambda = \frac{N \bar{X}}{(N-1) S^2}.$$