Stima Puntuale e Test d'Ipotesi

Corso di **Modelli Statistici e Statistical Learnig**

a.a. 2023/2024 - Primo Semestre

Corso di Laurea Magistrale in

Ingegneria Informatica (percorso AI-ML)

DIIMES

Prof. Filippo DOMMA

(filippo.domma@unical.it)

Dipartimento di Economia, Statistica e Finanza, Università della Calabria

Proprietà Asintotiche

Gli studi di tipo asintotico riguardano il comportamento degli stimatori al divergere della dimensione campionaria. In tale contesto, ci si aspetta che all'aumentare della dimensione campionaria la *performance* dello stimatore migliori in quanto aumenta l'informazione circa l'obiettivo di stima.

A tal fine, dato lo stimatore T=T(**X**) della funzione g(θ), consideriamo la successione di stimatori costruita nel seguente modo

$$
\begin{aligned}
X_1 &\longrightarrow T_1 = T(X_1) \\
X_1, X_2 &\longrightarrow T_2 = T(X_1, X_2) \\
X_1, X_2, X_3 &\longrightarrow T_3 = T(X_1, X_2, X_3) \\
\ldots. & \quad \ldots. \\
X_1, X_2, X_3, \ldots, X_n &\longrightarrow T_n = T(X_1, X_2, X_3, \ldots, X_n) \\
\ldots. & \quad \ldots.
\end{aligned}
$$

Ad esempio, consideriamo lo stimatore media campionaria

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

della media di una popolazione g(θ)=E(X), la successione di stimatori media campionaria è data da:

$$X_1 \longrightarrow \bar{X}_1 = X_1$$

$$X_1, X_2 \longrightarrow \bar{X}_2 = \frac{X_1 + X_2}{2}$$

$$X_1, X_2, X_3 \longrightarrow \bar{X}_3 = \frac{X_1 + X_2 + X_3}{3}$$

$\ldots.$ $\ldots.$

$$X_1, X_2, X_3, \ldots, X_n \longrightarrow \bar{X}_n = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n}$$

$\ldots.$ $\ldots.$

Stimatori asintoticamente non-distorti

Se uno stimatore è distorto per n fissato, possiamo chiederci se la sua distorsione si riduce all'aumentare della dimensione campionaria e, al limite, se si annulla.

Definizione. **Stimatore asintoticamente non-distorto**

Lo stimatore $T_n = T(X_1, X_2, X_3, \dots, X_n)$ di g(θ) è asintoticamente non-distorto se e solo se

$$\lim_{n \to \infty} E(T_n) = g(\theta) \quad \forall\, \theta \in \Theta$$

Esempio. Abbiamo visto che lo stimatore naturale della varianza di una popolazione, $S^{2\prime}$, è uno stimatore distorto per V(X) perché $\mathrm{E}\left(S^{2\prime}\right) = \left(1 - \frac{1}{n}\right) \times V(X)$.

D'altra parte, il $\lim_{n\to\infty} \mathrm{E}\left(S^{2\prime}\right) = \lim_{n\to\infty} \left(1 - \frac{1}{n}\right) \times V(X) = V(X)$ possiamo, quindi, concludere che $S^{2\prime}$ è uno stimatore asintoticamente non-distorto.

La Consistenza

Una proprietà asintotica legata alla dispersione dello stimatore intorno alla funzione del parametro da stimare è la cosiddetta consistenza. Si richiede che all'aumentare della dimensione campionaria e, quindi, all'aumentare delle informazioni sulla quantità da stimare, lo stimatore debba fornire stime sempre più "*vicine*" alla funzione g(θ). In altri termini, all'aumentare di *n* ci si aspetta che la dispersione di T_n intorno a g(θ) diminuisca. Tale proprietà è formalizzata nella seguente

Definizione. **Consistenza forte**

Lo stimatore $T_n = T(X_1, X_2, X_3, \dots, X_n)$ di g(θ) è fortemente consistente se converge quasi certamente a $g(\theta)$, cioè se

$$P\left\{\lim_{n \to \infty} T_n = g(\theta)\right\} = 1 \qquad \forall\, \theta \in \Theta$$

Una condizione più debole della consistenza forte è la consistenza semplice espressa nella definizione riportata di seguito

Definizione. **Consistenza semplice (o debole)**

Lo stimatore $T_n = T(X_1, X_2, X_3, \dots, X_n)$ di g(θ) è debolmente consistente per $g(\theta)$ se, scelto un $\varepsilon > 0$ qualsiasi, si ha

$$\lim_{n\to\infty} P\{|T_n - g(\theta)| \leq \varepsilon\} = 1 \qquad \forall\, \theta \in \Theta$$

Equivalentemente, si può scrivere

$$\lim_{n\to\infty} P\{g(\theta) - \varepsilon \leq T_n \leq g(\theta) + \varepsilon\} = 1 \qquad \forall\, \theta \in \Theta\; e\; \varepsilon > 0$$

In letteratura, alcuni autori fanno riferimento alla c**onsistenza in media quadratica**, affermando che lo stimatore $T_n(\mathbf{X})$ di $g(\theta)$ converge in media quadratica se

$$\lim_{n\to\infty} EQM(T_n) = 0 \quad \forall\, \theta \in \Theta$$

Metodi di Stima

Dato il modello parametrico

$$\boldsymbol{M} = \{\boldsymbol{P}, \boldsymbol{X}\}$$

Supponiamo che la famiglia di distribuzione P sia parametrizzata da un vettore di parametri di dimensione r, cioè

$$\boldsymbol{P} = \{f(.;\boldsymbol{\theta}): \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \Re^{\boldsymbol{r}}\}$$

L'obiettivo è quello di descrivere dei metodi che consentono di costruire stimatori per gli elementi del vettore dei parametri $\boldsymbol{\theta}$.

In letteratura esistono diverse tecniche, in questa parte del vedremo solo il Metodo dei Momenti e il Metodo della Massima Verosimiglianza. Il Metodo dei Minimi Quadrati verrà esposto nelle prossime lezioni.

Metodo dei Momenti

Abbiamo visto che il momento dall'origine di ordine k, indicato con $\mu_k = E[X^k] = \mu_k(\boldsymbol{\theta})$, in generale, è una funzione del vettore sconosciuto dei parametri della popolazione. Dato un campione casuale di dimensione n, indichiamo con M_j il momento j-esimo campionario, cioè $M_j = \frac{1}{n}\sum_{i=1}^{n} X_i^j$.

Uguagliando, ordinatamente, i primi r momenti campionari ai primi r momenti della popolazione, otteniamo un sistema di r-equazioni in r-incognite, cioè

$$\begin{cases} M_1 = \mu_1(\boldsymbol{\theta}) \\ M_2 = \mu_2(\boldsymbol{\theta}) \\ \cdots\cdots\cdots \\ M_r = \mu_r(\boldsymbol{\theta}) \end{cases} \Rightarrow \begin{cases} \frac{1}{n}\sum_{i=1}^{n} X_i = \mu_1(\theta_1, \ldots, \theta_r) \\ \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \mu_2(\theta_1, \ldots, \theta_r) \\ \cdots\cdots\cdots \\ \frac{1}{n}\sum_{i=1}^{n} X_i^r = \mu_r(\theta_1, \ldots, \theta_r) \end{cases}$$

Se $\widehat{\boldsymbol{\theta}} = \left(\hat{\theta}_1, \ldots, \hat{\theta}_r\right)^t$ è l'unica soluzione del sistema allora $\widehat{\boldsymbol{\theta}} = \left(\hat{\theta}_1, \ldots, \hat{\theta}_r\right)^t$ è lo stimatore di $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_r)$ ottenuto con il metodo dei momenti.

Esempio. Dato il c.c. $\mathbf{X} = (X_1, X_2, X_3, \dots, X_n)$ di dimensione n, estratto da una v.c. Normale di parametri μ e σ^2, determinare lo stimatore dei momenti del vettore di parametri $\theta = (\mu, \sigma^2)$. In questo caso, abbiamo due parametri da stimare e, quindi, dobbiamo costruire un sistema di 2 equazioni in due incognite, imponendo che i primi due momenti della popolazione siano uguali ai primi due momenti campionari, cioè

$$\begin{cases} M_1 = \mu_1(\boldsymbol{\theta}) \\ M_2 = \mu_2(\boldsymbol{\theta}) \end{cases} \Rightarrow \begin{cases} \frac{1}{n}\sum_{i=1}^{n} X_i = \mu \\ \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \mu^2 + \sigma^2 \end{cases} \Rightarrow \begin{cases} \hat{\mu} = \bar{X} \\ \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 \end{cases}$$

Esempio. Sia $\boldsymbol{X} = (X_1, \ldots, X_n)$ un campione casuale estratto da una popolazione con funzione di densità data da:

$$f(x;\theta) = \frac{3x^2}{\theta^3} \qquad 0 \leq x \leq \theta$$

a) determinare lo stimatore di θ con il metodo dei momenti;
b) stabilire se lo stimatore ottenuto è non-distorto e consistente in media quadratica

Metodo della Massima Verosimiglianza

Tale metodo si basa sull'idea che fd (o fp) diverse diano luogo, in generale, a campioni diversi e, quindi, sia più plausibile che il campione osservato provenga da una fd (o fp) con determinati valori di θ piuttosto che da altre fd (o fp) per le quali il suo realizzarsi sarebbe meno plausibile.

In altri termini, una volta osservato il campione $(x_1, \ldots, x_n)$, si cerca di determinare quale distribuzione ha generato con maggiore plausibilità il campione stesso. Così, fissate le osservazioni campionarie $(x_1, \ldots, x_n)$ si osserva come variano i valori di f(x;θ) al variare di θ nello spazio parametrico Θ. In tal modo, otteniamo una funzione definita sullo spazio parametrico Θ, cioè $L: \Theta \rightarrow \Re^+$.

ESEMPIO: GRAFICO

Definizione. Dato il modello parametrico $\mathcal{M} = \{P, \mathcal{X}\}$, sia $(X_1, \ldots, X_n)$ un campione casuale indipendente ed identicamente distribuito estratto da f(x;θ), e sia $\boldsymbol{x} = (x_1, \ldots, x_n)$ il campione osservato. La funzione di densità (o di probabilità) congiunta f(**x**;θ) definita sullo spazio parametrico Θ, viene detta funzione di verosimiglianza del campione osservato **x** ed indicata con:

$$L(\theta; \boldsymbol{x}) = f(\boldsymbol{x}; \theta)$$

Quindi, la funzione di verosimiglianza è una applicazione definita nello spazio parametrico che associa valori nell'insieme dei numeri reali positivi, cioè

$$L: \boldsymbol{\Theta} \rightarrow \Re^+$$

Possiamo dire che massimizzando L(θ;**x**) individuiamo quel valore di θ, diciamo $\hat{\theta} = \hat{\theta}(\boldsymbol{x})$ e, quindi, quella fd (o fp) che con maggior verosimiglianza ha generato le osservazioni campionarie $(x_1, \ldots, x_n)$. Da qui la seguente:

Definizione. Dato il modello parametrico $\boldsymbol{M} = \{\chi, \boldsymbol{P}\}$, sia $L(\theta; \boldsymbol{x})$ la f.v. del campione osservato $\boldsymbol{x} = (x_1, \ldots, x_n)$. Il valore $\hat{\theta} = \hat{\theta}(\boldsymbol{x})$ tale per cui

$$\boldsymbol{L}\left[\widehat{\boldsymbol{\theta}}(\boldsymbol{x}); \boldsymbol{x}\right] = \boldsymbol{\sup_{\theta \in \Theta} L}[\boldsymbol{\theta}; \boldsymbol{x}]$$

se esiste è detta stima di massima verosimiglianza. Tale valore valutato nel campione casuale **X**=(X_1,…,X_n) fornisce lo stimatore di massima verosimiglianza $\hat{\theta} = \hat{\theta}(\boldsymbol{X})$.

Sotto l'ipotesi di indipendenza, la fd (o fp) congiunta campionaria fattorizza nel prodotto delle marginali e, quindi, la funzione di verosimiglianza può essere scritta nel seguente modo

$$L(\theta; \boldsymbol{x}) = f(\boldsymbol{x}; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

Spesso, per semplificare i calcoli, si preferisce far riferimento al logaritmo della funzione di verosimiglianza, denominata funzione di log-verosimiglianza

$$\ell(\theta; \boldsymbol{x}) = ln\, L\,(\theta; \boldsymbol{x}) = ln \prod_{i=1}^{n} f(x_i; \theta) = \sum_{i=1}^{n} ln\, f\,(x_i; \theta)$$

poiché la trasformazione è monotona il punto che massimizza L(θ;**x**) massimizzerà anche $\ell(\theta; \boldsymbol{x})$; infatti, si ha:

$$\frac{\partial \ell(\theta;\boldsymbol{x})}{\partial \theta} = \frac{1}{L(\theta;\boldsymbol{x})} \frac{\partial L(\theta;\boldsymbol{x})}{\partial \theta} = 0 \quad \Leftrightarrow \quad \frac{\partial L(\theta;\boldsymbol{x})}{\partial \theta} = 0.$$

Analogamente a quanto detto in precedenza, se la fd (o fp) è funzione di un vettore di parametri $\boldsymbol{\theta}$, con $\theta \in \Theta \subseteq \Re^k$, allora la stima di massima verosimiglianza $\hat{\theta}(\boldsymbol{x}) = \left(\hat{\theta}_1(\boldsymbol{x}), \ldots, \hat{\theta}_k(\boldsymbol{x})\right)^t$ di $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_k)^t$ è quel vettore, se esiste, tale per cui

$$\boldsymbol{L}\left[\widehat{\boldsymbol{\theta}}(\boldsymbol{x}); \boldsymbol{x}\right] = \boldsymbol{\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} L}[\boldsymbol{\theta}; \boldsymbol{x}].$$

Sostituendo il campione osservato, **x**, con il campione casuale, **X**, si ottiene lo stimatore di m.v. $\hat{\theta}(\boldsymbol{X}) = \left(\hat{\theta}_1(\boldsymbol{X}), \ldots, \hat{\theta}_k(\boldsymbol{X})\right)^t$ del vettore dei parametri $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_k)^t$.

Nei casi in cui ℓ ($\boldsymbol{\theta}$;$\mathbf{x}$) è differenziabile rispetto a $\boldsymbol{\theta}$ allora, utilizzando i metodi dell'analisi matematica, costruiamo il seguente sistema

$$\begin{cases} \dfrac{\partial \ell(\theta;\boldsymbol{x})}{\partial \theta_1} = 0 \\ \cdots\cdots \\ \dfrac{\partial \ell(\theta;\boldsymbol{x})}{\partial \theta_k} = 0 \end{cases}$$

formato da k equazioni di verosimiglianza (le derivate parziali della log-verosimiglianza rispetto alle k componenti del vettore $\boldsymbol{\theta}$) in k incognite (gli elementi del vettore incognito $\boldsymbol{\theta}$). La soluzione di detto sistema fornisce la stima di massima verosimiglianza $\hat{\theta} = \left(\hat{\theta}_1, \ldots, \hat{\theta}_k\right)$ del vettore di parametri incogniti $\theta = (\theta_1, \ldots, \theta_k)$, una volta verificato che la matrice delle derivate seconde

Esempio. Bernoulli

Esempio. Normale.

Proprietà stimatori di massima verosimiglianza

Teorema (*Principle of invarance*).
Sia g(.) una funzione definita nello spazio parametrico con immagine in $\Omega \subset \Re$. Se $\hat{\theta} = \hat{\theta}(\boldsymbol{x})$ è la stima di massima verosimiglianza di $\theta \in \Theta \subset \Re$, allora $g(\hat{\theta})$ è la corrispondente stima di massima verosimiglianza di g(θ).

In estrema sintesi, gli stimatori di massima verosimiglianza:
- non necessariamente sono non-distorti;
- sono consistenti
- sono asintoticamente normali e pienamente efficienti

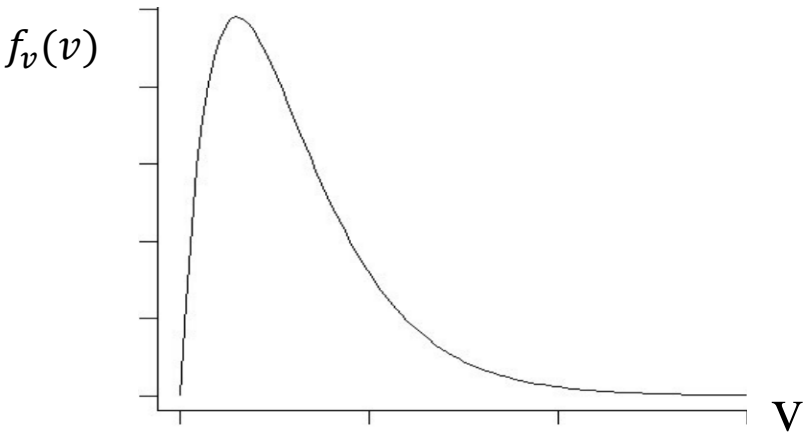
Alcune utili variabili casuali ottenute da trasformazioni della variabile casuale Normale.

Variabile Casuale chi-quadro χ^2

Definizione. Date k variabili casuali Normali e ***indipendenti,*** $X_i \sim N(\mu_i, \sigma_i^2)$ per i=1,.., k, la variabile casuale così definita

$$V = \sum_{i=1}^{k} \left(\frac{X_i - \mu_i}{\sigma_i} \right)^2$$

si distribuisce secondo una chi-quadrato di parametro k. Viene indicata con $\chi^2(k)$.



$$\boldsymbol{E(V) = k}$$

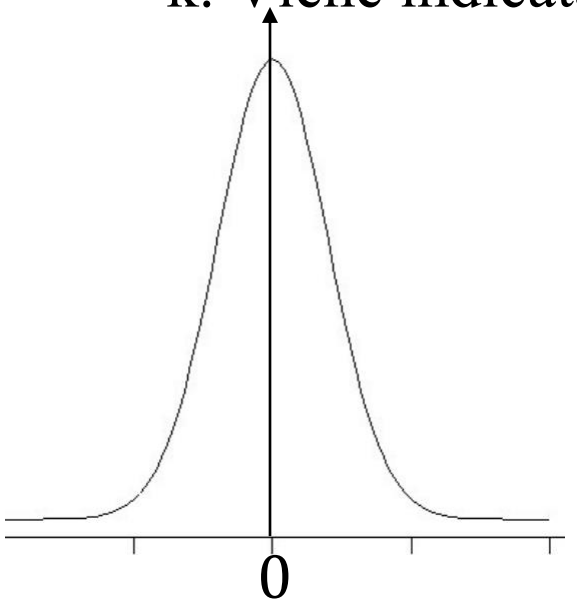
$$\boldsymbol{\mathrm{Var}(V) = 2k}$$

Variabile Casuale *t* – *Student*

Definizione. Se $X \sim N(\mu, \sigma^2)$ e $V \sim \chi^2(k)$ sono ***indipendenti***, allora il rapporto

$$T = \frac{\frac{X-\mu}{\sigma}}{\sqrt{\frac{V}{k}}} = \frac{Z}{\sqrt{\frac{V}{k}}}$$

si distribuisce secondo una t-Student di parametro k. Viene indicata con $t(k)$.



$$E(T) = 0$$

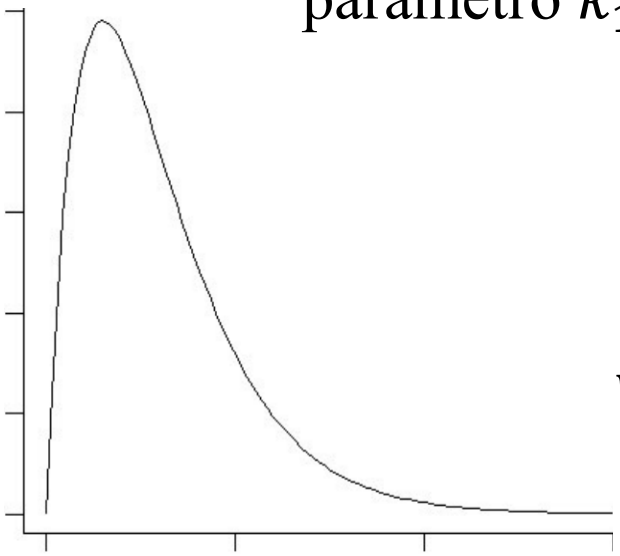
$$\mathrm{Var}(T) = \frac{k}{k-2}$$

Variabile Casuale *F di Fisher*

Definizione. Se $V_1 \sim \chi^2(k_1)$ e $V_2 \sim \chi^2(k_2)$ sono ***indipendenti***, allora il rapporto

$$F = \frac{\frac{V_1}{k_1}}{\frac{V_2}{k_2}}$$

si distribuisce secondo una F di Fisher di parametri parametro k_1 e k_2. Viene indicata con $F(k_1, k_2)$.



$$E(F) = \frac{n}{n-2} \qquad \text{per n>2}$$

$$V(F) = \frac{n^2(2m + 2n - 4)}{m(n-2)^2(n-4)} \qquad \text{per n>4}$$

Gradi di libertà

I gradi di libertà sono il numero di componenti del campione casuale che possono essere scelti liberamente dato dalla dimensione campionaria meno il numero di vincoli sullo spazio campionario costituite dalle stime preliminari che bisogna effettuare per calcolare lo stimatore '*corrente*'.

Dato un c.c. **X** di dimensione n, lo stimatore '*corrente*' $T_c = h(\boldsymbol{X}, T_1, \dots, T_r)$ ha *n-r* gradi di libertà.

Esempio. Supponiamo che la media campionaria su un campione di dimensione 5 risulta essere pari a 20. Se vogliamo calcolare la devianza campionaria (numeratore della varianza) $\sum_{i=1}^{5}(x_i - \bar{x})^2$ che rappresenta lo stimatore corrente, allora la media campionaria rappresenta un vincolo sullo spazio campionario e possiamo scegliere liberamente solo 4 componenti del campione.