

STIME PUNTUALI

Definizione 1

Se la stima di un parametro della distribuzione delle X_i è data da un singolo numero ottenuto dalle misurazioni relative a un campione, tale valore è detto stima puntuale del parametro.

Esempio La media campionaria osservata \bar{x}_n è una stima del parametro μ della distribuzione delle variabili aleatorie X_i . Il numero (stima) \bar{x}_n è una realizzazione dello stimatore X_n .

La statistica \bar{X}_n ha valore atteso μ , quindi, in questo caso, il valore atteso dello stimatore è uguale al parametro che si vuole stimare.

Definizione 2

Uno stimatore con valore atteso uguale al parametro che si vuole stimare si dice corretto (o non distorto).

La media campionaria $\bar{X}_n = (X_1 + \dots + X_n)/n$ è quindi uno stimatore corretto di μ .

Definizione 3

Se due statistiche sono entrambe stimatori corretti di un parametro, la statistica con minore varianza è detta stimatore più efficiente.

La stima ottenuta con la media campionaria è tanto più accurata tanto più il campione è numeroso. Infatti \bar{x}_n è un valore osservato di \bar{X}_n la cui varianza è σ/n . Pertanto **se si considerano due campioni di taglia m e n con $m > n$ si avrà che \bar{X}_m è più efficiente di \bar{X}_n nello stimare il parametro μ .**

In seguito potremo confrontare con l'efficienza di altri stimatori corretti diversi dalla media campionaria.

STIMA DELLA VARIANZA CAMPIONARIA

Abbiamo visto che una stima puntuale corretta per il valore atteso μ delle variabili aleatorie X_i è $\bar{x}_n = (x_1 + \dots + x_n)/n$.

Una stima puntuale della varianza σ^2 delle variabili aleatorie X_i è la varianza campionaria s_n^2

$$s_n^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n - 1}$$

Mostriamo che anche questa stima puntuale è **corretta** ossia che il valore medio dello stimatore

$$S_n^2 = \frac{(X_1 - \bar{X}_n)^2 + (X_2 - \bar{X}_n)^2 + \dots + (X_n - \bar{X}_n)^2}{n - 1}$$

è uguale alla varianza σ^2 delle variabili aleatorie X_i .

Si consideri un **campione** di numerosità n da una popolazione: \mathbf{n} variabili aleatorie $\mathbf{X}_1, \dots, \mathbf{X}_n$, con stessa distribuzione di cui **non si conosce la varianza** σ^2 . Otteniamo \mathbf{n} numeri $\mathbf{x}_1, \dots, \mathbf{x}_n$ (x_i è il valore osservato della variabile aleatoria X_i).

Sembra naturale prendere come stima di σ^2 il numero

$$a_n = \frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \dots + (x_n - \mu)^2}{n}$$

In effetti a_n è una stima corretta se il valore atteso μ delle variabili X_i è noto. Sia a_n una realizzazione della variabile aleatoria A_n (lo stimatore: stessa definizione di a_n ma con le realizzazioni x_i sostituite dalle corrispondenti variabili aleatorie X_i). Verificare che $\mathbb{E}(A_n) = \sigma^2$ è semplice, infatti dalla equazione qui sopra, tenendo presente che le variabili X_i sono identicamente distribuite, si ha che

$$\mathbb{E}(\mathbf{A}_n) = \mathbb{E}[(\mathbf{X}_1 - \mu)^2] = \sigma^2$$

Tuttavia molto spesso **anche il valore atteso è ignoto** e quindi sembra ragionevole sostituire μ con la sua stima che è la media campionaria \bar{x}_n .

$$b_n = \frac{(x_1 - \bar{x}_n)^2 + (x_2 - \bar{x}_n)^2 + \dots (x_n - \bar{x}_n)^2}{n}$$

In effetti b_n NON è una stima corretta. Verificare che $\mathbb{E}(B_n) \neq \sigma^2$ è semplice, infatti dalla equazione qui sopra, tenendo presente che le variabili X_i sono identicamente distribuite, si ha

$$\mathbb{E}(B_n) = \mathbb{E}[(X_1 - \bar{X}_n)^2] = \mathbb{E} \left[\left(X_1 - \frac{X_1 + X_2 + \dots X_n}{n} \right)^2 \right]$$

Tenendo presente che X_i e X_j sono indipendenti se $i \neq j$ si ha $\mathbb{E}(X_i X_j) = \mathbb{E}(X_i)\mathbb{E}(X_j) = \mu^2$ per $i \neq j$ (si noti che la prima uguaglianza, che non dimostriamo, è una diretta conseguenza dell'indipendenza). Si ha inoltre $\mathbb{E}[X_i^2] = \mu^2 + \sigma^2$ per ogni i , usando questi risultati si ottiene

$$\mathbb{E}(\mathbf{B}_n) = \frac{n-1}{n}\sigma^2$$

e quindi B_n **non è uno stimatore corretto**. Se invece di dividere per n si divide per $n-1$ si ottiene uno stimatore è corretto. In definitiva uno stimatore puntuale corretto della varianza σ^2 è

$$S_n^2 = \frac{(\mathbf{X}_1 - \bar{\mathbf{X}})^2 + (\mathbf{X}_2 - \bar{\mathbf{X}})^2 + \dots (\mathbf{X}_n - \bar{\mathbf{X}})^2}{n-1}$$

ossia è la **varianza campionaria** perchè $\mathbb{E}(S_n^2) = \frac{n}{n-1}\mathbb{E}(\mathbf{B}_n) = \sigma^2$.

Abbiamo visto che la media campionaria è uno stimatore corretto del valore atteso μ e la varianza campionaria è uno stimatore corretto della varianza σ^2 . Lo stimatore di un parametro non è unico: per decidere quale tra 2 stimatori corretti sia il migliore si usa il criterio di scegliere quello più efficiente ossia quello per cui la varianza è minore.

Esercizio 1. Si consideri un campione x_1, \dots, x_n di variabili aleatorie discrete X_i che assumono i tre valori 0, -1, 1 con probabilità

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

(1) Per quali a la funzione p è una distribuzione di probabilità ?

Risposta: deve essere $0 \leq p(i) \leq 1$ per ogni $i = 0, -1, 1$ quindi a deve essere $a \in [0, 1]$. Inoltre deve essere $p(-1) + p(0) + p(1) = 1$, cosa vera per ogni a .

(2) Calcolare valore atteso e varianza.

Risposta:

$$\mathbb{E}(X_1) = (-1) \cdot \frac{a}{2} + 0 \cdot (1 - a) + 1 \cdot \frac{a}{2} = 0.$$

$$\text{var}(X_1) = \mathbb{E}(X_1^2) = (-1)^2 \cdot \frac{a}{2} + 0^2 \cdot (1 - a) + 1^2 \cdot \frac{a}{2} = a$$

(3). Si considerino i seguenti 2 stimatori di a :

$$\Gamma = \frac{1}{n} \sum_{i=1}^n |X_i|, \quad \Lambda = |X_n|$$

I due stimatori sono corretti?

Risposta: $\mathbb{E}(\Lambda) = \mathbb{E}(|X_n|) = |-1| \cdot p(-1) + 0 \cdot p(0) + |1|p(1) = a$

Poichè le X_i . hanno stessa distribuzione:

$$\mathbb{E}(\Gamma) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(|X_i|) = \frac{1}{n} \sum_{i=1}^n a = a$$

Quindi sia Γ che Λ sono stimatori corretti.

(4) Si calcoli la varianza dei due stimatori.

Risposta: la varianza di Λ è

$$\text{var}(\Lambda) = \mathbb{E}\left((|X_n| - a)^2\right) = \mathbb{E}(|X_n|^2) - a^2$$

$$\mathbb{E}(|X_n|^2) = (-1)^2 p(-1) + 0^2 p(0) + 1^2 p(1) = \frac{a}{2} + \frac{a}{2} = a$$

quindi $\text{var}(\Lambda) = a - a^2 = a(1 - a)$.

Poichè le variabili aleatorie sono indipendenti e con stessa distribuzione:

$$\begin{aligned}\text{var}(\Gamma) &= \text{var}\left(\frac{1}{n} \sum_{i=1}^n |X_i|\right) = \frac{1}{n^2} \sum_{i=1}^n \text{var}(|X_i|) \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{var}(|X_n|) = \frac{1}{n^2} \sum_{i=1}^n a(1-a) = \frac{1}{n} a(1-a)\end{aligned}$$

(5) Quale è lo stimatore più efficiente?

Risposta: si ha

$$\frac{\text{var}(\Gamma)}{\text{var}(\Lambda)} = \frac{1}{n} < 1, \quad \text{per ogni } n > 1$$

Quindi Γ è più efficiente di Λ .

Esercizio 2. Si consideri un campione x_1, \dots, x_n di variabili aleatorie discrete X_i che assumono i due valori 0 e 1 con probabilità $1 - p$ e p rispettivamente.

(1) Si calcolino $\mu = \mathbb{E}(X_i)$ e $\sigma = \text{var}(X_i)$ in funzione di p .

Risposta: Si deve avere $\mu = 0 \cdot (1 - p) + 1 \cdot p = p$ per cui $\mu = p$, che implica $\mu \in [0, 1]$.

Inoltre $\sigma^2 = \mathbb{E}(X_i^2) - \mu^2 = 0^2 \cdot (1 - p) + 1^2 \cdot p - p^2 = p(1 - p)$.

(2) Lo stimatore di p

$$\Gamma = \frac{1}{12}X_1 + 3(X_2 + X_3)$$

è corretto? Risposta: Poichè le X_i hanno stessa distribuzione:

$$\mathbb{E}(\Gamma) = \frac{1}{12}\mathbb{E}(X_1) + 3\mathbb{E}(X_2) + 3\mathbb{E}(X_3) = \mathbb{E}(X_1)\left[\frac{1}{12} + 6\right] = p \frac{73}{12} = p \frac{73}{12}$$

quindi Γ è distorto (il suo valore atteso non coincide con p).

(3) Si determini una costante c tale che posto $\Lambda = c\Gamma$, Λ è uno stimatore corretto di p .

Risposta: siccome $\mathbb{E}(\Gamma) = 73p/12$, per $c = \frac{12}{73}$ si ha che

$$\mathbb{E}(\Lambda) = \frac{12}{73}\mathbb{E}(\Gamma) = p$$

quindi

$$\Lambda = \frac{1}{73}X_1 + \frac{36}{73}(X_2 + X_3)$$

è uno stimatore corretto di p .

(4) Si compari lo stimatore corretto di μ (e quindi di $p = \mu$) $\bar{X}_3 = (X_1 + X_2 + X_3)/3$ con Λ e si dica quale è quello più efficiente

Risposta: per le proprietà della varianza

$$\text{var}(\Lambda) = \left(\frac{1}{73}\right)^2 \text{var}(X_1) + \left(\frac{36}{73}\right)^2 \cdot \text{var}(X_2) + \left(\frac{36}{73}\right)^2 \cdot \text{var}(X_3)$$

inoltre tenendo presente che le variabili X_i sono identicamente distribuite con varianza σ^2

$$\text{var}(\Lambda) = \sigma^2 \left[\left(\frac{1}{73}\right)^2 + 2 \cdot \left(\frac{36}{73}\right)^2 \right] = 0.49 \cdot \sigma^2$$

Si ha invece $\text{var}(\bar{X}_3) = \sigma^2/3 = 0,33 \cdot \sigma^2$ e quindi

$$\frac{\text{var}(\bar{X}_3)}{\text{var}(\Lambda)} = \frac{0,33}{0,49} < 1$$

Quindi \bar{X}_3 è più efficiente di Λ .