9 Stimatori e loro proprietà

Definition 78 *Dediniamo campione aleatorio (random sample) un insieme di variabili aleatorie indipendenti ed identicamente distribuite* $X_1, ..., X_n$ *definite su uno spazio di probabilità* $\{\Omega, \Im, P\}$.

Remark 79 *In realtà né l'ipotesi di indipendenza né quella di identica distribuzione sono necessarie e saranno entrambe abbandonate più avanti; partiamo da queste condizioni super-semplificate per necessità.*

Supponiamo ora che la misura di probabilità P sia nota solo a meno del valore di un certo parametro θ. In genere, si definisce *Statistica Parametrica* la disciplina che studia le tecniche per risalire al valore di θ nel caso in cui esso appartenga ad uno spazio di dimensione finita, $\theta \in \mathbb{R}^p$, $p \in \mathbb{N}$; si parla invece di *Statistica NonParametrica* la disciplina che studia il caso in cui θ ha dimensione infinita. Ad esempio, se sappiamo che P è una legge Gaussiana ma ne ignoriamo valor medio e varianza siamo nell'ambito della Statistica Parametrica; se sappiamo che P ammette una densità e cerchiamo di ricavarla senza sapere che tipo di legge sia, siamo nell'ambito della Statistica Nonparametrica. Per distinguere un valore generico del parametro dal valore "vero" che caratterizza la legge da cui è stato estratto un certo campione aleatorio, puòà essere conveniente scrivere θ_0 in questo ultimo caso.

Definition 80 *(Stimatore) Uno stimatore di un certo parametro* $\theta \in \mathbb{R}^p$ *è una funzione (o meglio una successione di funzioni) misurabile* $T_n : \mathbb{R}^n \to \mathbb{R}^p$ *di un campione aleatorio; scriveremo equivalentemente*

$$T_n = \widetilde{\theta}_n = T_n(X_1, ...X_n) \ .$$

Remark 81 *La definizione è volutamente molto generica: qualsiasi funzione misurabile può essere considerata uno stimatore, quindi il punto cruciale sarà determinare quali siano le proprietà che consentono di valutare la bontà della scelta fatta.*

Example 82 *L'esempio più ovvio di stimatore (per il valore medio* $\mu = E[X_i]$*) è la media aritmetica , cioè*

$$\widetilde{X}_n = \widetilde{\mu}_n := \frac{1}{n} \sum_{i=1}^{n} X_i \ .$$

Altro stimatore naturale è la cosiddetta varianza empirica (che stima $\sigma^2 = Var[X_i]$*), cioè*

$$S_n^2 = \widetilde{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \widetilde{\mu}_n)^2.$$

9.1 Criteri per Valutare gli stimatori

- Non-distorsione/asintotica non distorsione
- Consistenza (in probabilità, in media r-esima, quasi certa, completa)
- Efficienza (assoluta e relativa)
- Asintotica Gaussianità

Example 83 *Media aritmetica - banale mostrare non-distorsione, consistenza ed asintotica Gaussianità; si può mostrare con Cramer-Rao (vedi dopo) che efficiente in senso assoluto nel caso Gaussiano. nel caso non-Gaussiano, si consideri un altro stimatore lineare*

$$\sum_{i=1}^{n} c_i X_i \ , \ \sum_{i=1}^{n} c_i = 1$$

(dove la condizione che la somma dei pesi sia 1 garantisce la non-distorsione). La varianza è evidentemente $\sum_{i=1}^{n} c_i^2 Var[X_i]$*, e pertanto nel caso di variabili incorrelate ed identicamente distribuite possiamo porcil problema di massimo vincolato*

$$\min \sum_{i=1}^{n} c_i^2 \ s.t. \ \sum_{i=1}^{n} c_i = 1 \ .$$

Con i moltiplicatori di Lagrange troviamo

$$\psi(c_1, ..., c_n; \lambda) = \sum_{i=1}^{n} c_i^2 + \lambda \left\{ \sum_{i=1}^{n} c_i - 1 \right\}$$

con derivate prime

$$\frac{\partial \psi(c_1, ..., c_n; \lambda)}{\partial c_i} = 2c_i + \lambda \ .$$

Ne segue che il problema di minimo sarà risolto imponendo che i coefficienti c_i *siano tutti uguali e pertanto pari a* n^{-1}*; questo rende la media aritmetica uno stimatore BLUE (Best Linear Unbiased Estimator).*

Example 84 *(Varianza Campionaria) In questo caso possiamo scrivere*

$$\begin{aligned} S_n^2 &= \frac{1}{n} \sum_{i=1}^{n} (X_i - \widetilde{\mu}_n)^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \widetilde{\mu}_n^2 \ , \\ E[S_n^2] &= \frac{1}{n} \sum_{i=1}^{n} E[X_i^2] - (E\widetilde{\mu}_n)^2 - Var[\widetilde{\mu}_n] \\ &= \frac{1}{n} \left(\sum_{i=1}^{n} \left\{ E[X_i^2] - \mu^2 \right\} - \sigma^2 \right) = \frac{n-1}{n} \sigma^2. \end{aligned}$$

La varianza campionaria risulta quindi distorta, anche se asintoticamente non-distorta; in un certo senso è come se perdessimo un grado di liberta centrando le osservazioni sulla media empirica invece di quella teorica (per $n = 1$ la varianza campionaria è identicamente nulla). Per quello che riguarda la consistenza debole, è una semplice conseguenza del Lemma di Slutzky; se le X_i hanno momenti quarti finiti, si può dimostrare la convergenza in media quadratica usando la disuguaglianza di Chebyshev.

Exercise 85 *Determinare opportune condizioni per dimostrare la convergenza quasi certa della varianza campionaria al valore della varianza σ^2; sotto quali condizioni vale il teorema del limte centrale?*