5. La Statistica Inferenziale

5.1. Introduzione ai problemi statistici.

Ogni rilevamento statistico produce un campione di dati (relativamente piccolo). La Statistica Descrittiva si occupa di organizzare e riassumere in modo significativo questi dati, e qui termina il suo compito. La Statistica Inferenziale, invece, utilizzando metodi e nozioni del Calcolo delle Probabilità, cerca di fare previsioni sul futuro, o di ottenere risultati estendibili all'intera popolazione (a partire solo dal piccolo campione effettivamente osservato).

Nella pratica, tipicamente si deve effettuare un esperimento che produce una variabile aleatoria X di cui non si conosce la legge, e si vogliono ricavare informazioni su di essa. Vediamo qualche esempio.

(5.1.1) Esempio. Si deve decidere se una data moneta è truccata oppure no. Cioè, se X è la v. a.

$$X = \begin{cases} 1 & \text{se esce } T \\ 0 & \text{se esce } C, \end{cases}$$

allora $X \sim B(1,p)$, ma p non è nota.

(5.1.2) Esempio. (di tipo biologico). Si vuole sapere se in una certa coltura batterica i batteri si distribuiscono in modo uniforme oppure tendono a formare aggregazioni. Supponiamo che il numero di batteri nella coltura sia N.

Suddividiamo la coltura in n parti (per es. n mm^3), e sia X il numero di batteri presenti in una parte fissata. La coltura sarà omogenea se ogni batterio sceglie a caso e indipendentemente la parte in cui impiantarsi; allora le v.a.

$$Y_k = \begin{cases} 1 & \text{se il } k\text{-esimo batterio sceglie la parte fissata} \\ 0 & \text{se no,} \end{cases} \qquad (k = 1, \ldots, N)$$

hanno densità $B(1, 1/n)$ e sono tra loro indipendenti. Pertanto si ha

$$X = Y_1 + \cdots + Y_N \sim B(N, 1/n).$$

Se N e n sono grandi, si può approssimare la legge di X con una Poisson di parametro $\lambda = N/n$. Concludiamo che la distribuzione dei batteri sarà

Nella Stat. inferenz. non conosciamo il parametro della legge di distribuzione relativa alla v.a. X

omogenea se scopriremo che X ha densità di Poisson. In altri termini la domanda iniziale è diventata: X ha legge di Poisson oppure no?

Si osservi che in questo caso non è neppure noto il tipo di legge seguita da X (e non solo un parametro di essa).

5.2. Il concetto di stimatore.

(5.2.1) GENERALITÀ. Uno dei problemi che si presentano più frequentemente allo statistico che voglia ottenere informazioni su una data v.a. X è quello di doverne decidere la legge: supponiamo che egli sappia che tale legge appartiene ad una data famiglia, dipendente da un parametro θ non noto: ad esempio, in (5.1.1) si sa che la legge considerata è di tipo bernoulliano, ma non se ne conosce il parametro p.

CONVENZIONE. Nel caso generale il parametro da studiare è indicato con θ. In casi specifici il simbolo usato potrà essere diverso (per esempio, in (5.1.1) era p, in (5.1.2) era λ.

Chiediamoci cosa può fare lo sperimentatore in questa situazione: la cosa più naturale è quella di "procurarsi delle osservazioni" del fenomeno di cui X è l'espressione, effettuando qualche tipo di esperimento, e decidere (in statistica si dice anche "fare inferenza") in base ai risultati ottenuti.

Tipicamente come risultato del suo esperimento egli otterrà dei numeri $x_1, \ldots, x_n$. Essi vanno pensati come valori assunti (nel corso dell'esperimento) da certe v.a. $X_1, \ldots, X_n$ aventi una legge congiunta dipendente da θ. Le chiameremo *osservazioni* di X. In corrispondenza i numeri $x_1, \ldots, x_n$ (che sono i valori assunti dalle osservazioni dopo che l'esperimento è stato effettuato) si chiameranno *valori osservati.*

Un caso molto frequente (ma non l'unico!) è quello in cui $X_1, \ldots, X_n$ sono tra loro indipendenti ed hanno tutte la stessa legge di X: è la formalizzazione matematica del caso in cui lo sperimentatore decide di ripetere n volte, in condizioni di indipendenza, proprio l'esperimento che produce X. Si dice allora che $X_1, \ldots, X_n$ costituiscono un *campione di taglia n estratto* dalla legge di X.

Comunque sia, lo sperimentatore userà i numeri trovati per calcolare, a partire da essi, una stima del parametro incognito; in altri termini sceglierà una opportuna (secondo lui) funzione t di n variabili reali e stimerà il parametro θ con il numero $t(x_1, \ldots, x_n)$. Ovviamente la funzione t andrà scelta non dipendente dal parametro incognito (dato che essa va usata appunto per stimarlo!). Queste considerazioni giustificano la seguente:

STIMATORE

(5.2.2) DEFINIZIONE. Sia $t : \mathbb{R}^n \to \mathbb{R}$ una funzione non dipendente da θ.
*(i) Si chiama *stimatore* di θ la v. a. $T = t(X_1, \ldots, X_n)$, dove $X_1, \ldots, X_n$ sono n osservazioni di X. ↳ possibili valori

→ valori esperimento eseguito

(ii)Si chiama *stima* di θ il numero $t = t(x_1, \ldots, x_n)$, dove $x_1, \ldots, x_n$ sono gli n valori osservati (corrispondenti alle osservazioni del punto (i)).

(5.2.3) OSSERVAZIONE. Per noi la situazione più comune sarà quella in cui $\theta \in \mathbb{R}$, ma in generale θ va pensato come un vettore (cioè la legge può dipendere da più di un parametro reale, come accade per esempio per la $\mathcal{N}(\mu, \sigma^2)$, quando sia la media μ che la varianza σ^2 non sono note).

(5.2.4) STIMATORI CORRETTI. Siano $X_1, \cdots, X_n$ n osservazioni aventi tutte la stessa legge, dipendente da un parametro θ.
Sia $T = t(X_1, \ldots, X_n)$ uno stimatore di una funzione $\psi(\theta)$ del parametro. La situazione "ideale" sarebbe che valesse l'uguaglianza

$$T(\omega) = \psi(\theta) \qquad \forall \omega \in \Omega, \tag{5.2.5}$$

cioè che lo stimatore fornisse sempre e esattamente la quantità da stimare. Ciò non è ovviamente possibile; più ragionevole è chiedersi se l'uguaglianza (5.2.5) possa valere almeno *in media*; in effetti una buona proprietà di uno stimatore è la seguente:

(5.2.6) DEFINIZIONE. Lo stimatore $T = t(X_1, \cdots, X_n)$ di $\psi(\theta)$ si dice *corretto* (o *non distorto*, *unbiased* in inglese) se vale la relazione

$$\mathbf{E}^\theta[T] = \psi(\theta) \qquad \forall \theta \in \Theta.$$

NOTE. (i) In tutti gli esempi che seguono, indicheremo con μ e σ^2 rispettivamente la media e la varianza della comune legge delle X_i. Osservare che μ e σ^2 sono, naturalmente, due funzioni del parametro θ, ma questo fatto non viene messo in risalto nella notazione usata, per motivi di brevità.

(ii) Nelle scritture del tipo $\mathbf{E}^\theta$ o $\mathbf{Var}^\theta$ o simili sottintenderemo il parametro θ, (cioè scriveremo semplicemente $\mathbf{E}$ o $\mathbf{Var}$, naturalmente sempre che ciò non dia luogo ad equivoci).

(iii) Queste convenzioni di scrittura saranno tacitamente usate anche nei paragrafi successivi.

(5.2.7) ESEMPI. (a) La media campionaria $\overline{X}$ è uno stimatore corretto di μ. Infatti

$$\mathbf{E}[\overline{X}] = \frac{1}{n} \sum_{i=1}^{n} \mathbf{E}[X_i] = \frac{n\mu}{n} = \mu.$$

$$\overline{X} = \frac{X_1 + X_2 + \ldots X_n}{n}$$

(b) Se μ è nota, lo stimatore

$$T = \frac{\sum_{i=1}^{n}(X_i - \mu)^2}{n} \tag{5.2.8}$$

è uno stimatore corretto di σ^2. Infatti

$$\mathbf{E}\Big[\frac{\sum_{i=1}^{n}(X_i - \mu)^2}{n}\Big] = \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}[(X_i - \mu)^2] = \frac{1}{n}\sum_{i=1}^{n}\sigma^2 = \sigma^2.$$

(c) Supponiamo ora in più che le osservazioni $X_1, \cdots, X_n$ (oltre ad essere equidistribuite) siano tra loro indipendenti. Vogliamo trovare uno stimatore corretto di σ^2 nel caso, molto frequente, che μ non sia nota (in tale situazione la v. a. indicata in (5.2.8) non è uno stimatore, perché dipende da μ). L'idea è quella di sostituire μ con il suo stimatore $\overline{X}$ nella formula (5.2.8), cioè usare lo stimatore

$$Z = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}. \tag{5.2.9}$$

Calcoliamo dunque $\mathbf{E}[Z]$. Cominciamo calcolando la media del numeratore della frazione in (5.2.9):

$$\begin{aligned}
\mathbf{E}\big[\sum_{i=1}^{n}(X_i - \overline{X})^2\big] &= \mathbf{E}\Big[\sum_{i=1}^{n}X_i^2 - 2\overline{X}\frac{\sum_{i=1}^{n}X_i}{n}n + \sum_{i=1}^{n}\overline{X}^2\Big] \\
&= \mathbf{E}\big[\sum_{i=1}^{n}X_i^2 - 2n\overline{X}^2 + n\overline{X}^2\big] = \mathbf{E}\big[\sum_{i=1}^{n}X_i^2 - n\overline{X}^2\big] \\
&= \sum_{i=1}^{n}\mathbf{E}[X_i^2] - n\mathbf{E}[\overline{X}^2].
\end{aligned} \tag{5.2.10}$$

D'altra parte

$\mathrm{Var}(x) \doteq E(x^2) - (E(x))^2$

$$\mathbf{E}[X_i^2] = \mathbf{Var}X_i + \mathbf{E}^2[X_i] = \sigma^2 + \mu^2; \tag{5.2.11}$$

$$\begin{aligned}
\mathbf{E}[\overline{X}^2] &= \mathbf{Var}\overline{X} + \mathbf{E}^2[\overline{X}] = \mathbf{Var}\big(\frac{\sum_{i=1}^{n}X_i}{n}\big) + \mu^2 \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\mathbf{Var}X_i + \mu^2 = \frac{\sigma^2}{n} + \mu^2
\end{aligned} \tag{5.2.12}$$

$\frac{n\sigma^2}{n^2}$

(nella (5.2.12) è stata usata l'indipendenza delle X_i; dove?).
Usando le relazioni (5.2.11) e (5.2.12) nella (5.2.10) si ottiene:

$$\mathbf{E}\Big[\sum_{i=1}^{n}(X_i - \overline{X})^2\Big] = \sum_{i=1}^{n}(\sigma^2 + \mu^2) - n\Big(\frac{\sigma^2}{n} + \mu^2\Big) = (n-1)\sigma^2.$$

Dunque

$$\mathbf{E}\big[Z\big] = \frac{n-1}{n}\sigma^2, \tag{5.2.13}$$

cioè Z è uno stimatore distorto! Tuttavia il calcolo appena fatto ci dice che è corretto lo stimatore (non molto diverso da Z per n grande)

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}.$$

(5.2.14) CONFRONTO DI STIMATORI. Sostituendo il vero valore di $\psi(\theta)$ con il suo stimatore $T = t(X_1, \dots, X_n)$ si commette un errore, che è opportuno saper misurare. Si dà dunque la seguente

(5.2.15) DEFINIZIONE. Si chiama *rischio quadratico* dello stimatore T la funzione $\theta \mapsto R_T(\theta)$ definita su Θ da

$$R_T(\theta) = \mathbf{E}^\theta\big[(T - \psi(\theta)^2)\big].$$

(5.2.16) OSSERVAZIONE. La definizione data sopra si motiva come abbiamo fatto a suo tempo per la varianza (vedere le considerazioni fatte prima della definizione (3.1.15)). pag 56

(5.2.17) OSSERVAZIONE. Se lo stimatore T è corretto, si ha evidentemente

$$R_T(\theta) = \mathbf{Var}^\theta T.$$

In presenza di due stimatori S e T della quantità $\psi(\theta)$, preferiremo ovviamente lo stimatore con rischio più piccolo, per ogni valore di θ (sempre che uno dei due realizzi questa richiesta). Cioè

(5.2.18) DEFINIZIONE. (i) Si dice che S è *preferibile* a T se $R_S(\theta) \le R_T(\theta)$ $\forall \theta \in \Theta$; se in più $\exists \theta_0 \in \Theta$ tale che $R_S(\theta_o) < R_T(\theta_o)$, allora si dice che S è *strettamente preferibile* a T. Def strettamente preferibile

(ii) Uno stimatore è detto $\boxed{\textit{ammissibile}}$ se non esistono stimatori ad esso strettamente preferibili.

(5.2.19) Stimatori consistenti. Molte proprietà degli stimatori (ad esempio la correttezza) hanno bisogno solo di un numero finito di osservazioni $X_1, \ldots, X_n$ (cioè in questo momento n va pensato come un numero intero fissato).

Talvolta, però, è utile conoscere il comportamento *asintotico* di uno stimatore (cioè per $n \to \infty$, e in tal caso bisogna immaginare di avere a disposizione una successione di osservazioni $X_1, X_2, X_3, \ldots$).

Ad esempio, uno stimatore assegnato può non essere corretto per nessun valore finito di n, ma "tendere a diventare corretto" quando $n \to \infty$. In tal caso, se possibile e in pratica non troppo costoso, questo potrebbe indurci ad aumentare il numero delle osservazioni, in modo da avvicinarci in modo abbastanza soddisfacente alla correttezza. (Questo è quello che accade ad esempio per lo stimatore Z della varianza σ^2 definito in (5.2.9): abbiamo visto che Z non è corretto, ma si ha (vedi (5.2.13))

$$\lim_{n\to\infty} \mathbf{E}[Z] = \lim_{n\to\infty} \frac{n-1}{n}\sigma^2 = \sigma^2,$$

D = 1

esempio di uno stimatore Z

e dunque Z è asintoticamente corretto.

Un'altra proprietà asintotica che può essere importante per uno stimatore $T_n = t_n(X_1, \ldots, X_n)$ è la "consistenza". Essa riguarda la funzione di ripartizione di T_n e la sua variazione al crescere di n (per sottolineare il fatto che in questo momento stiamo parlando di una successione di stimatori, li indichiamo con T_n invece che semplicemente con T). Al solito indichiamo con $\psi(\theta)$ una funzione del parametro θ.

(5.2.20) Definizione. Una successione (T_n) di stimatori di $\psi(\theta)$ si dice *fortemente consistente* se T_n converge verso $\psi(\theta)$ quasi certamente.

(5.2.21) Esempio. Per la legge forte dei grandi numeri, la media campionaria

$$\overline{X} = \frac{X_1 + \ldots X_n}{n}$$

è uno stimatore fortemente consistente della media. (Per essere precisi, dovremmo mettere un indice n a $\overline{X}$, poichè n varia, e meglio sarebbe dire "la successione delle medie campionarie" ....).

(5.2.22) Definizione. Una successione (T_n) di stimatori di $\psi(\theta)$ si dice *debolmente consistente* se T_n converge verso $\psi(\theta)$ in probabilità.

Nella pratica, per trovare stimatori consistenti, è utile il seguente risultato, di cui omettiamo la dimostrazione.

(5.2.23) TEOREMA. *Sia (T_n) una successione di stimatori fortemente consistenti del parametro θ, e supponiamo che $\theta \mapsto \psi(\theta)$ sia una funzione continua. Allora $U_n = \psi(T_n)$ è una successione di stimatori fortemente consistenti di $\psi(\theta)$.*

(5.2.24) ESEMPIO (tipico). Sia X una v. a. avente densità

$$f_\alpha(x) = \begin{cases} \alpha x^{\alpha-1} & 0 < x < 1; \\ 0 & \text{altrove}, \end{cases}$$

dove $\alpha > 0$ non è noto. Trovare una successione di stimatori consistenti per α.

SOLUZIONE. In questo, come in altri casi simili, il trucco è quello di considerare la legge dipendente <u>non</u> dal parametro α, ma dal parametro $\mu = \mathbf{E}[X]$, nel modo seguente. Calcoliamo prima di tutto μ (ovviamente in funzione di α!). Si ha

$$\mu = \mathbf{E}[X] = \int_0^1 x\alpha x^{\alpha-1} = \frac{\alpha}{\alpha+1}. \tag{5.2.25}$$

Si osserva intanto che $0 < \mu < 1$. Inoltre, invertendo la relazione (5.2.25), si ottiene

$$\alpha = \frac{\mu}{1-\mu} \doteq \psi(\mu).$$

Nell'intervallo aperto $(0,1)$, $\mu \mapsto \psi(\mu)$ è una funzione continua. Sappiamo (esempio (5.2.21)) che $T_n = \overline{X}$ è uno stimatore fortemente consistente di μ. Posto allora

$$U_n = \psi(\overline{X}) = \frac{\overline{X}}{1-\overline{X}}, \tag{5.2.26}$$

si ottiene dal Teorema (5.2.23) che (U_n) è uno stimatore fortemente consistente di α.

(5.2.27) OSSERVAZIONE. L'espressione al secondo membro della relazione (5.2.26) non ha senso sull'evento $\{\overline{X} = 1\}$, ma si può dimostrare che questo evento ha probabilità nulla.

(5.2.28) STIMATORI DEI MOMENTI.

(5.2.29) MOTIVAZIONE. Come si sarà già intuito dopo aver letto il paragrafo sugli stimatori consistenti, generalmente si adotta qualche particolare criterio per individuare uno stimatore adatto (ad esempio, appunto, la correttezza). In questa sezione e nella successiva vedremo altri due criteri importanti.

DEF Sia X una v. a. la cui legge dipende da un certo numero di parametri $\theta_1, \ldots, \theta_r$ (il "parametro" θ è in generale un vettore, $\theta = (\theta_1, \ldots, \theta_r)$ appunto). Supponiamo che $\theta_1, \ldots, \theta_r$ non siano noti, e come al solito il nostro scopo è darne una stima dipendente dalle osservazioni $(X_1, \ldots, X_n)$.

(5.2.30) DEFINIZIONE. Sia k un numero intero ≥ 1 fissato. Si chiama *momento teorico di ordine* k di X il numero

$$m_k \doteq \mathbf{E}[X^k],$$

(a patto che esso sia finito, e cioè che $\mathbf{E}[|X|^k] < \infty$).

(5.2.31) OSSERVAZIONE. Dato che la legge di X dipende da $\theta_1, \ldots, \theta_r$, lo stesso accadrà per il momento teorico m_k (che, a sua volta, dipende solo dalla legge di X); in altre parole esisterà una funzione f_k di r variabili tale che

$$m_k = f_k(\theta_1, \ldots, \theta_r).$$

Supponiamo che la v. a. X ammetta i primi q momenti (q è un numero intero ≥ 1). Ciò significa che possiamo scrivere la relazione precedente $\forall k = 1, \ldots, q$, ottenendo così il sistema

$$\begin{cases} m_1 = f_1(\theta_1, \ldots, \theta_r) \\ m_2 = f_2(\theta_1, \ldots, \theta_r) \\ \cdot \\ \cdot \\ \cdot \\ m_q = f_q(\theta_1, \ldots, \theta_r). \end{cases}$$

Si tratta evidentemente di un sistema di q equazioni nelle r incognite $\theta_1, \ldots, \theta_r$, che si può cercare di risolvere. Se questo è possibile, otterremo

r espressioni del tipo seguente

$$(5.2.32)\qquad \begin{cases} \theta_1 = g_1(m_1, m_2, \ldots, m_q) \\ \theta_2 = g_2(m_1, m_2, \ldots, m_q) \\ . \\ . \\ . \\ \theta_r = g_r(m_1, m_2, \ldots, m_q). \end{cases}$$

(5.2.33) DEFINIZIONE. Supponiamo che le osservazioni $(X_1. \ldots, X_n)$ siano tra loro indipendenti. Si definisce *momento empirico* di X la quantità (aleatoria!)

$$\hat{m}_k \doteq \frac{\sum_{i=1}^{n} X_i^k}{n}.$$

Per la legge dei grandi numeri, si ha

$$(5.2.34)\qquad \hat{m}_k = \frac{\sum_{i=1}^{n} X_i^k}{n} \to_{n\to\infty} \mathbf{E}[X^k] = m_k;$$

(5.2.35) OSSERVAZIONE. Questa relazione spiega i nomi di momento *teorico* e momento *empirico* dati alle due quantità sopra definite; ci dice anche che i momenti empirici sono stimatori consistenti di quelli teorici; rivedere la definizione di consistenza.

La relazione (5.2.34) suggerisce il procedimento seguente: dato che per n grande $\hat{m}_k \simeq m_k$, nel sistema (5.2.32) sostituiamo $\hat{m}_k$ al posto di m_k per ogni $k = 1, \ldots, q$, ottenendo le relazioni

$$\begin{cases} \theta_1 \simeq g_1(\hat{m}_1, \hat{m}_2, \ldots, \hat{m}_q) \\ \theta_2 \simeq g_2(\hat{m}_1, \hat{m}_2, \ldots, \hat{m}_q) \\ . \\ . \\ . \\ \theta_r \simeq g_r(\hat{m}_1, \hat{m}_2, \ldots, \hat{m}_q), \end{cases}$$

ovvero avremo “espresso” (approssimativamente!) ciascuno dei parametri in termini delle v. a. $\hat{m}_1, \hat{m}_2, \ldots, \hat{m}_q$ che sono note, perchè dipendenti solo dalle osservazioni.

Dunque, a sua volta, ciascuna delle funzioni $g_i(\hat{m}_1, \hat{m}_2, \ldots, \hat{m}_q)$, per ogni $i = 1, \ldots, r$, dipende solo dalle osservazioni, ed è dunque uno stimatore $\hat{\theta}_i = \hat{\theta}_i(X_1, \ldots, X_n)$ di θ_i.

(5.2.36) DEFINIZIONE. Il vettore di v. a. $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_r)$ si chiama *stimatore dei momenti* del parametro $\theta = (\theta_1, \ldots, \theta_r)$.

(5.2.37) OSSERVAZIONE. Se si deve stimare con il metodo dei momenti una funzione del parametro θ, sia essa $\psi(\theta) = \psi(\theta_1, \ldots, \theta_r)$, si prende lo stimatore

$$\psi(\hat{\theta}_1, \ldots, \hat{\theta}_r).$$

Per esercizio, dare una giustificazione teorica di questa "regola".

(5.2.38) OSSERVAZIONE. Seguendo la definizione (5.2.2), chiameremo *stime dei momenti* i numeri che si ottengono mettendo i valori osservati, cioè $x_1, \ldots, x_n$ al posto delle osservazioni $X_1, \ldots, X_n$ nelle espressioni, sopra definite, $\hat{\theta}_i(X_1, \ldots, X_n)$.

(5.2.39) OSSERVAZIONE. Spesso il numero q dei momenti che servono per fare i conti è uguale al numero r dei parametri da stimare, cioè si cerca di scrivere (e risolvere) un sistema di r equazioni in r incognite.

(5.2.40) ESEMPIO. Stimatore dei momenti del parametro della legge esponenziale basato sulle osservazioni $(X_1, \ldots, X_n)$.

È noto che, se $X \sim \mathcal{E}(\lambda)$, allora

$$m_1 = \mathbf{E}[X] = \frac{1}{\lambda}t = 0$$

e, risolvendo, si trova

$$\lambda = \frac{1}{m_1}.$$

Si pone allora

$$\hat{\lambda} = \frac{1}{\hat{m}_1} = \frac{n}{X_1 + \cdots X_n}.$$

Naturalmente questo stimatore è definito solo sull'evento $\{X_1 + \cdots X_n t = 0\}$ (che però si può dimostrare avere probabilità nulla).

Gli stimatori dei momenti sono spesso distorti, ma consistenti (come si può intuire dalla consistenza dei momenti empirici, di cui essi sono funzione). Tuttavia in generale essi non sono dei buoni stimatori. Stimatori migliori si ottengono con il *metodo della massima verosimiglianza*, che vedremo nella prossima sezione.

(5.2.40) STIMATORI DI MASSIMA VEROSIMIGLIANZA.

(5.2.41) ESEMPIO INTRODUTTIVO. Un'urna contiene 1000 palline, alcune rosse e altre bianche. La percentuale di palline di ciascun colore non è nota: si sa soltanto che l'urna contiene o 1 pallina rossa e 999 palline bianche oppure, viceversa, 1 pallina bianca e 999 palline rosse. Un tizio deve stabilire quale delle due composizioni è quella vera, basandosi su qualche tipo di osservazione, a sua scelta. Egli allora decide di effettuare 100 estrazioni con rimpiazzo, e ottiene in ciascuna estrazione una pallina bianca. A questo punto, come è facile capire, egli è propenso a credere che l'urna contenga 999 palline bianche e 1 rossa. Ritiene infatti che il risultato ottenuto (100 palline bianche in 100 estrazioni) potrebbe sì verificarsi anche se l'urna avesse l'altra composizione, ma con una probabilità inferiore.

Cerchiamo di formalizzare quello che è accaduto. Indichiamo con X il risultato di una generica estrazione:

$$X = \begin{cases} 1 & \text{se esce una pallina bianca} \\ 0 & \text{se esce una pallina rossa.} \end{cases}$$

Allora le 100 estrazioni effettuate costituiscono un campione di 100 osservazioni del fenomeno, che indicheremo, come sempre, con $(X_1, \ldots, X_{100})$. Sulla legge di X l'informazione è la seguente:

$$X \sim B(1,p), \text{ con } p \in \Pi = \left\{ \frac{1}{1000}, \frac{999}{1000} \right\}.$$

Allora la probabilità di ottenere 100 palline bianche in 100 estrazioni è

$$P^p(X_1 = 1, X_2 = 1, \ldots, X_{100} = 1) = p^{100} = \begin{cases} (\frac{1}{1000})^{100} & \text{se } p = \frac{1}{1000}, \\ (\frac{999}{1000})^{100} & \text{se } p = \frac{999}{1000}. \end{cases}$$

Il tizio ha dunque deciso di considerare vero il valore di p per il quale il risultato effettivamente osservato è il più probabile. In altre parole, egli ha calcolato

$$\max_{p \in \Pi} P^p(X_1 = 1, X_2 = 1, \ldots, X_{100} = 1) = \max_{p \in \Pi} p^{100},$$

ed ha deciso per il valore del parametro in cui tale massimo è raggiunto, cioè per il punto di massimo della funzione

$$p \mapsto P^p(X_1 = 1, X_2 = 1, \ldots, X_{100} = 1) = p^{100}.$$

Questo esempio suggerisce il procedimento che ora descriveremo del (metodo per il calcolo dello *stimatore di massima verosimiglianza*).

Def

Sia X una v.a. la cui legge dipende da un parametro θ. Disponiamo di n osservazioni di X (non necessariamente indipendenti, come capitava nell'esempio), che indichiamo con $(X_1, \ldots, X_n)$. Per semplicità, in questo momento supporremo che la legge congiunta delle osservazioni sia discreta, e indicheremo con P^θ la loro densità congiunta. (In seguito toglieremo questa restrizione). Dopo avere effettuato l'esperimento, il campione di osservazioni avrà prodotto un campione di n valori osservati (ricordare la distinzione che abbiamo fatto tra i termini *osservazione* e *valore osservato*), $(x_1, \ldots, x_n)$. La probabilità (in funzione di $\theta \in \Theta$) che il risultato sia quello effettivamente osservato è

$$\theta \mapsto P^\theta(X_1 = x_1, \ldots, X_n = x_n), \qquad \theta \in \Theta.$$

Essa è evidentemente una funzione di θ dipendente dai parametri reali $x_1, \ldots, x_n$. Nel caso generale (cioè se il vettore di osservazioni non è necessariamente discreto) essa si chiama *funzione di verosimiglianza* (*likelihood* in inglese) e il simbolo usato sarà piuttosto

$$\theta \mapsto L(\theta | x_1, \ldots, x_n), \qquad \theta \in \Theta.$$

Vedremo fra poco come calcolarla in alcuni casi importanti.

Supponiamo di essere riusciti a calcolare (in qualche modo) il massimo di questa funzione (al variare di $\theta \in \Theta$); il corrispondente punto di massimo dipenderà anch'esso dai parametri $x_1, \ldots, x_n$; indichiamolo con $\hat{\theta} = \hat{\theta}(x_1, \ldots, x_n)$.

(5.2.43) DEFINIZIONE. Il numero $\hat{\theta}(x_1, \ldots, x_n)$ si chiama *stima di massima verosimiglianza* di θ (in corrispondenza dei valori osservati $x_1, \ldots, x_n$).

È evidente dalla definizione che la stima di massima verosimiglianza è funzione delle variabili reali $x_1, \ldots, x_n$. Dunque

(5.2.44) Definizione. La v.a. $\hat{\theta}(X_1, \dots, X_n)$ (in cui semplicemente abbiamo messo le v.a. $X_1, \dots, X_n$ al posto dei numeri $x_1, \dots, x_n$) si chiama *stimatore di massima verosimiglianza* di θ.

(5.2.44) Osservazione. Normalmente lo stimatore di massima verosimiglianza è ancora indicato con $\hat{\theta}$, anche se la notazione può generare confusione tra stimatore e stima.

Occupiamoci ora un po' più in dettaglio di alcuni metodi per il calcolo della funzione di verosimiglianza. Una situazione semplice (del resto già vista sopra) è quella in cui il vettore delle osservazioni ha densità (congiunta) discreta. Un caso particolare di questa situazione si ha quando le osservazioni costituiscono un campione, cioè sono tra loro indipendenti. Supponiamo cioè che la v.a. X abbia densità discreta $p_\theta(x) = P^\theta(X = x)$. Allora

$$P^\theta(X_1 = x_1, \dots, X_n = x_n) = p_\theta(x_1) \cdots p_\theta(x_n) \doteq L(\theta|x_1, \dots, x_n).$$

Per analogia, se X è assolutamente continua, con densità $f_\theta(x)$ (e naturalmente le osservazioni sono ancora tra loro indipendenti), la funzione di verosimiglianza è definita dalla formula

$$L(\theta|x_1, \dots, x_n) \doteq f_\theta(x_1) \cdots f_\theta(x_n).$$

Per concludere, diamo un rapido elenco di alcuni metodi usati per trovare il punto di massimo di L.

(i) Come nell'esempio dell'urna fatto all'inizio, supponiamo che l'insieme Θ sia costituito da un numero finito di elementi $\theta_1, \dots, \theta_M$. Se M non è troppo grande, un modo del tutto elementare consiste nel calcolare $L(\theta_i)$ $\forall i = 1, \dots, M$ e scegliere il valore di θ_i per cui $L(\theta_i)$ risulta massimo. Se M è grande, bisognerà ricorrere a qualche trucco, da vedere caso per caso.

(ii) Se Θ è un intervallo della retta (eventualmente non limitato), si possono applicare i metodi studiati in "Analisi I" per trovare i punti di massimo e minimo di una funzione: in particolare, la stima di massima verosimiglianza, se interna all'intervallo Θ, è soluzione dell'equazione

$$\frac{d}{d\theta} L(\theta|x_1, \dots, x_n) = 0.$$

Attenzione agli (eventuali) estremi dell'intervallo, per i quali non si può applicare il criterio dei punti stazionari.

(iii) Nel caso che la stima di massima verosimiglianza sia stata ottenuta con il solo criterio dei punti stazionari, è necessario ricordare che il fatto che un punto sia stazionario non garantisce automaticamente che esso sia di massimo. Di regola, bisognerebbe proseguire nell'indagine (come è noto dall'analisi) studiando la derivata seconda di L, oppure in altro modo. Noi ci accontenteremo quasi sempre della sola stazionarietà, per non appesantire i calcoli (in alcuni casi esistono criteri specifici, che garantiscono la correttezza del procedimento).

(iv) In molti casi, prima di effettuare la derivazione di L, conviene passare al logaritmo, cioè derivare non la funzione $\theta \mapsto L(\theta|x_1, \dots, x_n)$, ma $\theta \mapsto \log L(\theta|x_1, \dots, x_n)$. Il passaggio al logaritmo non cambia infatti i punti stazionari, né la loro natura (perché?). In questi casi, dunque, la stima di massima verosimiglianza viene trovata come soluzione dell'equazione

$$\frac{d}{d\theta} \log L(\theta|x_1, \dots, x_n) = 0,$$

che è nota come *equazione di verosimiglianza*. Di essa esiste una versione multidimensionale (che si utilizza quando il parametro θ è un vettore), ma la trattazione va oltre gli scopi di questi appunti (in particolare, il caso multidimensionale ha bisogno dei metodi dell'"Analisi II").

(5.2.46) OSSERVAZIONE. Per stimare con il metodo della massima verosimiglianza una funzione $\psi(\theta)$ del parametro θ, si utilizza la quantità $\psi(\hat{\theta})$, dove $\hat{\theta}$ è la stima (o lo stimatore) di massima verosimiglianza di θ. Non daremo una giustificazione teorica di questa regola.

(5.2.47) ESEMPIO. Calcolare la stima e lo stimatore di massima verosimiglianza del parametro p della legge bernoulliana $B(1, p)$, basato sul campione $(X_1, \dots, X_n)$. Si suppone che $p \in (0, 1)$.

Sia X una v. a. di legge $B(1, p)$; la sua densità, come è noto, è

$$P^p(X = x) = p^x (1 - p)^{1-x}, \qquad x = 0, 1.$$

Dunque la funzione di verosimiglianza è

$$\begin{aligned} p \mapsto L(p|x_1, \dots, x_n) &= p^{x_1}(1-p)^{1-x_1} \cdots p^{x_n}(1-p)^{1-x_n} \\ &= p^{x_1+\dots+x_n}(1-p)^{n-(x_1+\dots+x_n)}, \end{aligned}$$

per $(x_1, \ldots, x_n) \in \{0,1\}^n$. Per semplicità poniamo

$\frac{d}{dp}\log(1-p) = \frac{1}{1-p}\cdot(0-1) = -\frac{1}{1-p}$

$$\alpha = x_1 + \ldots + x_n.$$

L'equazione di verosimiglianza è

$$0 = \frac{d}{dp}\log L(p|x_1, \ldots, x_n) = \frac{d}{dp}(\alpha \log p + (n-\alpha)\log(1-p)) = \frac{\alpha}{p} - \frac{n-\alpha}{1-p},$$

la cui unica soluzione è

$$\hat{p} = \frac{\alpha}{p} = \frac{x_1 + \ldots + x_n}{n} = \overline{x}.$$

(media del campione di valori osservati). È semplice vedere che il valore così trovato è effettivamente un punto di massimo (verifica per esercizio), e si conclude che esso è la stima di massima verosimiglianza di p cercata. Lo stimatore di massima verosimiglianza è dunque

$$\hat{p} = \frac{X_1 + \ldots + X_n}{n} = \overline{X},$$

ovvero la media campionaria delle osservazioni.

(5.2.48) Esempio. Calcolare la stima e lo stimatore di massima verosimiglianza del parametro λ della legge esponenziale $\mathcal{E}(\lambda)$, basato sul campione $(X_1, \ldots, X_n)$. Si suppone che $\lambda \in (0, +\infty)$.

Sia X una v. a. di legge $\mathcal{E}(\lambda)$; la sua densità (assolutamente continua), come è noto, è

$$f_\lambda(x) = \begin{cases} \lambda e^{-\lambda x} & \text{per } x > 0 \\ 0 & \text{altrove.} \end{cases}$$

Dunque la funzione di verosimiglianza è

$$p \mapsto L(\lambda|x_1, \ldots, x_n) = \begin{cases} \lambda e^{-\lambda x_1} \ldots \lambda e^{-\lambda x_n} & x_i > 0 \ \forall i, \\ 0 & \text{altrove} \end{cases}$$
$$= \begin{cases} \lambda^n e^{-\lambda(x_1 + \ldots + x_n)} & x_i > 0 \ \forall i, \\ 0 & \text{altrove .} \end{cases}$$

È chiaro che il punto di massimo di questa funzione si troverà nel quadrante $\{x_1 > 0, \ldots, x_n > 0\}$ (dove la funzione è strettamente positiva). Ponendo di nuovo, come nell'esempio precedente,

$$\alpha = x_1 + \ldots + x_n.$$

l'equazione di verosimiglianza è

$$0 = \frac{d}{d\lambda}\big(\log(\lambda^n e^{-\lambda\alpha})\big) = \frac{d}{d\lambda}(n\log\lambda - \lambda\alpha) = \frac{n}{\lambda} - \alpha,$$

da cui si ricava

$$\hat{\lambda} = \frac{n}{\alpha} = \frac{n}{x_1 + \ldots + x_n} = \frac{1}{\overline{x}}.$$

Tralasciamo anche qui la verifica che questo è effettivamente il punto di massimo della funzione nell'intervallo considerato $(0, +\infty)$, e dunque la stima di massima verosimiglianza. Corrispondentemente, lo stimatore sarà

$$\hat{\lambda} = \frac{1}{\overline{X}}.$$

Supponiamo ora di dover stimare la media dell'esponenziale. Dato che essa è uguale a $\psi(\lambda) = 1/\lambda$, per il criterio dell'osservazione (5.2.46) si ottiene lo stimatore

$$\psi(\hat{\lambda}) = \overline{X}.$$

5.3. Intervalli di fiducia.

(5.3.1) MOTIVAZIONE. Nei problemi di stima di un parametro incognito θ, o di una sua funzione $\psi(\theta)$, spesso è utile poter dire che la probabilità con la quale θ può essere approssimato dallo stimatore scelto non è troppo bassa.

NOTA. In questo paragrafo, riguardante gli intervalli di fiducia, e nei successivi, in cui parleremo di test, ogni funzione delle osservazioni, non dipendente dal parametro incognito θ, sarà detta *statistica*. Dunque il concetto di statistica non è diverso da quello di *stimatore*, che abbiamo dato in precedenza (vedi def. (5.2.2)). Quello che cambia è solo il nome, per il differente punto di vista in cui ci mettiamo: ora non si tratta solo di dare un valore approssimato del parametro, ma, come vedremo, di utilizzare la nostra funzione delle osservazioni per confronti più specifici.

(5.3.2) DEFINIZIONE. Sia $\alpha \in (0, 1)$ un numero fissato. Date due statistiche $T_1 = t_1(X_1, \ldots, X_n)$ e $T_2 = t_2(X_1, \ldots, X_n)$, si dice che $I_X = [T_1, T_2]$ è un *intervallo di fiducia* per $\psi(\theta)$ *di livello* $1 - \alpha$ se, $\forall\theta \in \Theta$ si ha

$$P^\theta\big(\psi(\theta) \in I_X\big) \geq 1 - \alpha.$$