# Capitolo 8

# Stima dei Parametri: Metodo di Massima Verosimiglianza

Lo scopo dello studio sperimentale dei fenomeni fisici è quello di scoprire quali sono le leggi che legano le grandezze fisiche e di ottenere il valore delle costanti che compaiono nella formulazione matematica di queste leggi. Per questo studio è necessario disporre della relazione matematica con la quale descrivere gli andamenti attesi delle misure delle grandezze fisiche studiate. Questa relazione è il cosiddetto "modello matematico[1]" (o più semplicemente "modello") del fenomeno studiato e consiste in un'espressione matematica nella quale compaiono oltre ai valori delle grandezze fisiche, un certo numero di parametri incogniti $\lambda_1, \lambda_2, \ldots, \lambda_k$ (il cui insieme sarà indicato con $\Lambda$ per brevità). Ottenere, o come si suole dire, stimare il valore di questi parametri dai dati sperimentali è uno dei risultati più importanti dello studio dei fenomeni fisici.

I metodi con cui si ottene la stima del valore dei parametri che caratterizzano una popolazione statistica sono un'importante capitolo della statistica detto *teoria degli stimatori*. La teoria degli stimatori si divide in *puntuale*, quando valuta il valore di un parametro, e in stima di *intervallo* o intervallare, quando assegna un intervallo di valori che includa il parametro cercato con un prefissato livello di confidenza[2]

Per esemplificare quanto esposto si consideri un esperimento in cui si vuole stimare l'accelerazione di gravità locale studiando il moto di un grave in caduta libera; le grandezze fisiche osservate sono le distanze percorse dal grave in corrispondenza ad un certo numero di tempi fissati: $(t_i, s_i)$, $i = 1, \ldots, N$. La teoria, in particolare il secondo principio della dinamica, stabilisce che tra lo spazio percorso e il tempo esiste la relazione[3] $s = gt^2/2$ dove $g$ è l'accelerazione di gravità; il modello matematico da adattare ai dati sarà quindi una forma quadratica: $s = \lambda_1 + \lambda_2 t + \lambda_3 t^2$ dove il parametro $\lambda_1$ tiene conto di una eventuale valore non nullo della reale coordinata di inizio del moto, $\lambda_2$ tiene conto di un'eventuale velocità iniziale non nulla e infine $\lambda_3$ è il parametro che stima l'accelerazione di gravità diviso 2. Se i dati sperimentali sono in numero sufficiente (almeno quanti sono i parametri $\Lambda$, ma

[1]Il modello matematico del fenomeno studiato può essere derivato da considerazioni teoriche sul fenomeno oppure può essere un andamento fenomenologico escogitato per la descrizione quantitativa del fenomeno oppure poò essere una distribuzione di probabilità con cui si descrivono le misure di una grandezza fisica.

[2]Si parla di *confidenza* o *fiducia* e non di *probabilità* perché nell'interpretazione frequentista della probabilità i parametri delle teorie hanno valori fissi per quanto non noti e quindi *non ammettono una distribuzione di probabilità*. L'interpretazione soggettivista della probabilità al contrario, ammette che i parametri incogniti in quanto tali abbiano una distribuzione di probabilità.

[3]Supponendo velocità iniziale e coordinata di inizio del moto nulle

possibilmente in numero molto maggiore) la teoria degli estimatori fornisce gli strumenti per stimare tutti questi parametri dai dati sperimentali.

Nel seguito saranno esposti i due metodi più usati per la stima puntuale dei parametri: il metodo detto di *massima verosimiglianza* e il metodo detto dei *minimi quadrati*.

## 8.1 Il metodo di Massima Verosimiglianza

**Esempio introduttivo.** Per introdurre il concetto della *verosimiglianza* (ing. *Likelihood*) in statistica e del suo uso nella valutazione dei parametri si consideri un semplice esperimento consistente in $N$ lanci di un dado e l'evento favorevole sia l'uscita del numero "$6$". Come è noto, il numero degli eventi favorevoli segue la distribuzione binomiale[4]

$$\mathcal{B}(k|p) = \binom{N}{k} p^k (1-p)^{N-k} \tag{8.1}$$

dove $p$ è la probabilità del singolo evento e, fissato il numero $N$ degli eventi, $p$ è l'unico parametro della distribuzione. Se il dado è ben equilibrato è lecito supporre $p = 1/6$ e la forma della distribuzione di probabilità (8.1) è completamente determinata e appare come quella disegnata nella figura 8.1.

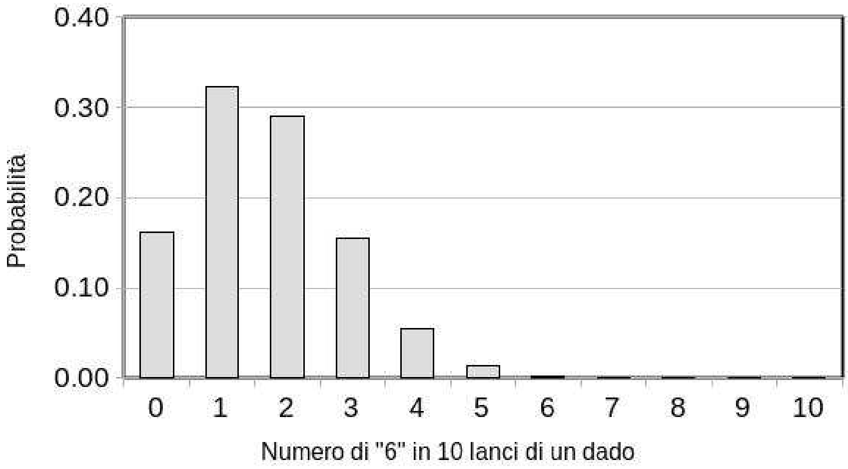


Figura 8.1: Distribuzione binomiale di probabilità per $N = 10$ e $p = 1/6$

In generale se si conosce sia il tipo di distribuzione di probabilità sia il valore dei parametri che la caratterizzano si conosce la probabilità con cui si realizzano tutti i possibili eventi.

Tuttavia il problema che si deve affrontare nell'attività sperimentale è quello inverso, e cioè *noti i dati si vuole ottenere la stima dei parametri*. Continuando con l'esempio del dado, supponiamo che si abbia il sospetto che il dado sia "truccato" (ad esempio appesantendo una faccia) e che quindi la probabilità del "$6$" abbia un valore $p$, differente da $1/6$. Per stimare il valore di $p$ eseguiamo un esperimento lanciando il dado $N$ volte e sia $k = k_o$ il numero di volte che è uscito il "$6$". Un modo di stimare il valore di $p$, consiste nel calcolare la probabilità assegnata all'evento osservato come una funzione del *parametro* $p$ avendo fissato

[4] Sulla distribuzione binomiale vedi il paragrafo 5.8.2

il valore della variabile aleatoria a quello osservato ($k = k_o$):

$$L(p|k_0) = \binom{N}{k_0} p^{k_0}(1-p)^{N-k_0} \tag{8.2}$$

la funzione $L(p|k_0)$, detta *funzione di verosimiglianza*, descrive l'informazione guadagnata eseguendo l'esperimento. La funzione di verosimiglianza ha la stessa forma algebrica della distribuzione di probabilità (8.1) ma ha un significato totalmente differente, infatti la funzione $L$ dipende unicamente dai parametri (in questo esempio $p$) mentre le variabili aleatorie prendono un valore fisso ($k = k_0$). Nella figura 8.2 è mostrato l'andamento della funzione di verosimiglianza (8.2) per due valori differenti di $k_o$.

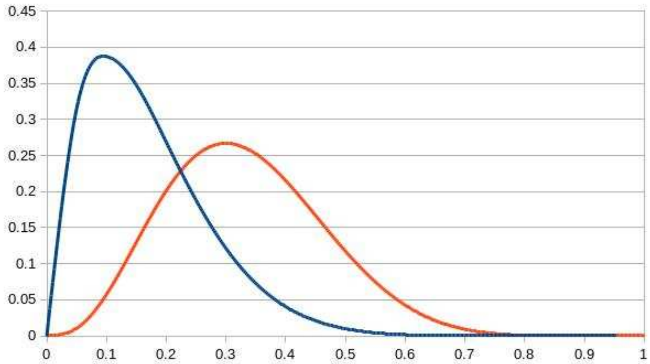


Figura 8.2: Funzione di verosimiglianza (8.2) in funzione del parametro $p$. Le due curve di verosimiglianza mostrate corrispondono a due serie di 10 lanci di un dado in cui si sono avuti $k_0 = 1$ eventi favorevoli (curva con il massimo a $p = 0.1$) e $k_0 = 3$ eventi favorevoli (curva con il massimo a $p = 0.3$).

Il criterio per stimare $p$, noto come il *criterio di massima verosimiglianza*, afferma che l'evento che si è realizzato (in questo caso esemplificato da $k = k_0$) ha la massima probabilità di accadere (infatti è accaduto) e la stima del parametro $p$ (con il metodo di massima verosimiglianza) si ottiene *massimizzando la funzione di verosimiglianza* (8.2) in funzione di $p$. Tornando all'esempio, il valore di $p$ per cui la funzione di verosimiglianza (8.2) ha il suo massimo è *la stima di massima verosimiglianza* ($\hat{p}$) del parametro $p$. Per semplificare i calcoli è consuetudine utilizzare al posto della funzione $L$ il suo logaritmo[5]: $\mathcal{L} = \ln L$. In formule:

$$\mathcal{L} = \ln\left[\binom{N}{k_0} p^{k_0}(1-p)^{N-k_0}\right] = k_o \ln p + (N-k_0)\ln(1-p) + \ln\binom{N}{k_0}$$

Per trovare il massimo di $\mathcal{L}$ annulliamo la sua derivata prima:

$$\frac{d\mathcal{L}}{dp} = \frac{k_0}{p} - \frac{N-k_0}{1-p} = 0, \quad \text{da cui} \quad \hat{p} = \frac{k_0}{N}$$

Il valore $\hat{p}$ così ottenuto è la *stima di massima verosimiglianza del parametro* $p$. Si noti che la stima di massima verosimiglianza $\hat{p}$ è uguale a quella che si ottiene dalla legge dei grandi numeri.

[5]Si può facilmente dimostrare che $L$ e $\mathcal{L} = \ln L$ hanno massimi (e minimi) per gli stessi valori delle variabili da cui dipendono

**Caso generale.** Generalizzando l'esempio precedente si consideri un evento definito da un insieme di $N$ grandezze aleatorie $X_1, X_2, \ldots, X_N$ indipendenti[6] tra loro e siano $f_i(X_i|\Lambda)$ le distribuzioni di probabilità che danno la probabilità di osservare $X_i$ dato $\Lambda$; $\Lambda = \{\lambda_1, \ldots, \lambda_m\}$ rappresenta l'insieme degli $m$ parametri da cui dipendo le $f_i$.

Data l'indipendenza delle realizzazioni $X_i$ la probabilità di osservare una particolare realizzazione delle $X_i$, che indichiamo con $\{x_1, x_2, \ldots, x_N\}$ è proporzionale a:

$$f_1(x_1|\Lambda) \cdot f_2(x_2|\Lambda) \cdot \ldots f_n(x_N|\Lambda) = \prod_{i=1}^{N} f_i(x_i|\Lambda) \tag{8.3}$$

se in questa relazione le $x_i$ sono gli esiti di un esperimento, ovvero uno specifico campione statistico, le $x_i$ hanno un valore numerico definito e la relazione (8.3) diviene una funzione dei soli parametri $\Lambda$. Questa funzione, in analogia con l'esempio precedente, è detta funzione di *verosimiglianza*:

$$L = \prod_{i=1}^{N} f(x_i|\Lambda) \tag{8.4}$$

Come già anticipato nell'esempio introduttivo, in luogo della (8.4) si usa definire come funzione di verosimiglianza il logaritmo della (8.4):

$$\mathcal{L} = \ln L = \ln \prod_{i=1}^{N} f(x_i|\Lambda) = \sum_{i=1}^{N} \ln f(x_i|\Lambda) \tag{8.5}$$

**Il criterio di massima verosimiglianza.** Questo criterio, originariamente sviluppato da R. A. Fisher nel 1922, stabilisce che

> *i valori preferiti dei parametri di una funzione di verosimiglianza sono quelli che rendono massima la probabilità di ottenere i dati osservati.*

Ne segue che le stime dei parametri con il metodo di massima verosimiglianza sono le soluzioni del seguente sistema:

$$\frac{\partial \mathcal{L}}{\partial \lambda_1} = \frac{\partial \mathcal{L}}{\partial \lambda_2} = 0 \ldots \frac{\partial \mathcal{L}}{\partial \lambda_m} = 0$$

La stima dei parametri che si ottengono massimizzando la funzione (8.5) è detta *stima di massima verosimiglianza* dei parametri.

**Esempi.** Sono riportati nel seguito alcuni esempi di stima di massima verosimiglianza di alcuni parametri. Si noti che per l'applicazione del metodo di massima verosimiglianza è necessario conoscere il tipo di distribuzione di probablità delle variabili aleatorie coinvolte.

1. *Stima del valore medio della gaussiana.* Consideriamo $n$ misurazioni ripetute di una grandezza $\{x_1, x_2, \ldots, x_n\}$ che sappiamo avere una distribuzione normale di varianza nota $\sigma$ e valore medio da determinare. Supponiamo inoltre che le misurazioni ripetute della grandezza siano indipendenti. Determiniamo la stima di massima verosimiglianza del valore medio $\mu$. Da quanto detto la funzione di verosimiglianza in questo caso è

$$\mathcal{L}(\mu) = \ln \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-(x_i-\mu)^2/2\sigma^2} = \sum_{i}^{n} -\frac{(x_i-\mu)^2}{2\sigma^2} + \frac{n}{2} \ln 2\pi\sigma^2 \tag{8.6}$$

[6] La condizione di indipendenza non è essenziale per nella definizione del metodo di massima verosimiglianza ma ne semplifica la formulazione matematica

Derivando la funzione di verosimiglianza rispetto a $\mu$ e annullando la derivata si ha:

$$\frac{d\mathcal{L}}{d\mu} = \sum_i^n \frac{(x_i - \mu)}{\sigma^2} = 0, \quad \text{da cui} \quad \hat{\mu} = \frac{1}{n}\sum_i^n x_i$$

La stima di massima verosimiglianza del valore medio di una gaussiana $\hat{\mu}$ è data dalla media campionaria del campione di dati acquisiti.

2. *Stima della deviazione standard della gaussiana.* Riprendiamo l'esempio precedente supponendo che sia la varianza $\sigma$ il parametro da determinare. Derivando rispetto a $\sigma$ l'espressione della funzione di verosimiglianza (8.6) e annullando la derivata si ha:

$$\frac{d\mathcal{L}}{d\sigma} = \sum_i^n \frac{(x_i - \mu)^2}{\sigma^3} + \frac{n}{\sigma} = 0, \quad \text{da cui} \quad \hat{\sigma} = \sqrt{\frac{1}{n}\sum_i^n (x_i - \mu)^2}$$

Si noti che in questo caso la stima di massima verosimiglianza della deviazione standard è *distorta* se, come succede nella maggior parte dei casi, il valore medio $\mu$ è ignoto ed è sostituito con la sua stima $\hat{\mu}$.

3. *Media pesata.* La formula della media pesata, utilizzata per combinare due o più misure della stessa grandezza con incertezze differenti, può essere ottenuta attraverso il metodo di massima verosimiglianza. Per semplicità si considerino solo due misure indipendenti di una stessa grandezza: $x_1 \pm u_1, x_2 \pm u_2$. Si supponga inoltre che $x_1$ e $x_2$ siano distribuite in modo normale con deviazioni standard rispettivamente $u_1$ e $u_2$.

Se con $\mu$ si indica il valore della grandezza (parametro da stimare) le *pdf* delle due variabili aleatorie $x_1$ e $x_2$ sono date da:

$$\frac{1}{\sqrt{2\pi}u_1}\mathrm{e}^{-(x_1-\mu)^2/2u_1^2}; \qquad \frac{1}{\sqrt{2\pi}u_2}\mathrm{e}^{-(x_2-\mu)^2/2u_2^2}$$

La funzione di verosimiglianza $\mathcal{L}$ in questo caso si scrive come:

$$\mathcal{L}(\mu) = \ln\left[\frac{1}{2\pi}\frac{1}{u_1 u_2}\mathrm{e}^{-\frac{(x_1-\mu)^2}{2u_1^2}}\,\mathrm{e}^{-\frac{(x_2-\mu)^2}{2u_2^2}}\right] = -\left[\frac{(x_1-\mu)^2}{2u_1^2} + \frac{(x_2-\mu)^2}{2u_2^2}\right] + \text{cost.}$$

Per trovare il massimo di $\mathcal{L}$, uguagliamo a zero la derivata rispetto a $\mu$

$$\frac{d}{d\mu}\mathcal{L}(\mu) = -\frac{d}{d\mu}\left(\frac{(x_1-\mu)^2}{2u_1^2} + \frac{(x_2-\mu)^2}{2u_2^2}\right) = 0$$

Introducendo i pesi, definiti da $w_i = 1/u_i^2$, $(i = 1, 2)$, si ottiene $w_1(x_1 - \mu) + w_2(x_2 - \mu) = 0$, e infine risolvendo per $\mu$, si ottiene la stima di massima verosimiglianza di $\mu$:

$$\hat{\mu} = \frac{w_1 x_1 + w_2 x_2}{w1 + w2}$$

che per misure distribuite in modo normale, coincide esattamente con la formula della media pesata.

ESERCIZIO. Calcolare la stima di massima verosimiglianza di due misure $x_1 \pm u_1, x_2 \pm u_2$ entrambe distribuite in modo normale di una stessa grandezza supponendo che siano correlate con coefficiente di correlazione $\rho$. Per eseguire questo calcolo utilizzare la relazione (5.57).

4. *Processo di Poisson.* Consideriamo un processo governato dalla statistica di Poisson come ad esempio il numero di conteggi di un contatore di raggi cosmici in un fissato intervallo temporale. Siano $k_1, k_2, \ldots, k_n$ $n$ realizzazioni del processo di Poisson considerato; indicando con $\mu$ il parametro della distribuzione di Poisson, la funzione di verosimiglianza si scrive:

$$\mathcal{L}(\mu) = \ln\left(\prod_i^n \frac{\mu^{k_i}}{k_i!}\mathrm{e}^{-\mu}\right) = \left(\sum_i k_i\right)\ln\mu - n\mu - \sum_i \ln k_i!$$

Derivando rispetto a $\mu$ e annullando la derivata si ottiene:

$$\frac{d\mathcal{L}}{d\mu} = \frac{\sum_i^n k_i}{\mu} - n = 0, \qquad \text{da cui} \quad \hat{\mu} = \frac{1}{n}\sum_i^n k_i \tag{8.7}$$

La stima di massima verosimiglianza del parametro $\mu$ della distribuzione di Poisson coincide con la media campionaria.