5 ESTIMATORI E METODI DI INFERENZA STATISTICA

Una delle principali preoccupazioni nelle analisi dei dati consiste nell'inferire dai dati misurati i parametri che descrivono un modello teorico. Una qualsiasi funzione del mio set di n misure sperimentali $\vec{x} = \{x_i\}$ di una data grandezza X che non includa parametri sconosciuti è detta *statistico*. Quando lo statistico $\hat{\theta} = f(\vec{x})$ si riferisce a parametri θ quali il valore di aspettazione μ, la varianza σ^2 o altri parametri che definiscono la pdf associata alla grandezza X, si parla di *estimatori* di tali parametri. Il modulo di teoria è entrato nel dettaglio della definizione di estimatori. Qui conviene richiamarene soltanto alcune proprietà: un estimatore è detto

- *consistente* se all'aumentare delle misure ($n \to \infty$) $\hat{\theta}$ converge, nel senso della probabilità, al parametro che stima θ;
- *non distorto (unbiased)* se è nulla la differenza tra il valore di aspettazione dell'estimatore e del parametro vero, ovvero è nullo il *bias* $b = E[\hat{\theta}] - \theta$)

5.1 ESTIMATORI DEL VALORE DI ASPETTAZIONE E DELLA VARIANZA

Si dimostra facilmente che il valore medio del mio set di misure, ottenuto secondo la (1.9) è un estimatore consistente e non distorto, assumendo ovviamente che ciascuna misura x_i della stessa grandezza X sia assimilabile ad una variabile aleatoria con la stessa pdf $f(x)$.

Verifichiamo qui che il seguente estimatore

$$\begin{aligned} s^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{x})^2 \\ &= \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 + n\hat{x}^2 - 2\hat{x}\sum_{i=1}^{n} x_i\right) \\ &= \frac{1}{n-1}\left(\frac{n}{n}\sum_{i=1}^{n} x_i^2 + n\hat{x}^2 - 2\hat{x}\frac{n}{n}\sum_{i=1}^{n} x_i\right) \\ &= \frac{n}{n-1}\left(\hat{x^2} - \hat{x}^2\right) \end{aligned} \tag{5.1}$$

sia un estimatore non distorto della varianza σ^2 della variabile aleatoria x associata alla grandezza X.

$$
\begin{aligned}
E[s^2] &= E\left[\frac{n}{n-1}\left(\widehat{x^2}-\hat{x}^2\right)\right]\\
&= E\left[\frac{n}{n-1}\left(\frac{\sum_i x_i^2}{n}-\left(\frac{\sum_i x_i}{n}\right)^2\right)\right]\\
&= \frac{n}{n-1}\left(\frac{1}{n}\sum_i E\left[x_i^2\right]-\frac{1}{n^2}E\left[\sum_i x_i^2+2\sum_{i>j}x_i x_j\right]\right)\\
&= \frac{n}{n-1}\left(\frac{1}{n}\sum_i E\left[x_i^2\right]-\frac{1}{n^2}\sum_i E\left[x_i^2\right]-\frac{2}{n^2}\mu^2\frac{n(n-1)}{2}\right)\\
&= \frac{n}{n-1}\left(\frac{1}{n^2}(n-1)\sum_i E\left[x_i^2\right]-\mu^2\frac{(n-1)}{n}\right)\\
&= \frac{1}{n}\sum_i E\left[x_i^2\right]-\mu^2\frac{n}{n}\\
&= \frac{1}{n}\left(\sum_i E\left[x_i^2\right]-\sum_i \mu^2\right)\\
&= \frac{1}{n}\sum_i\left(E\left[x_i^2\right]-\mu^2\right)\\
&\quad \text{assumendo che ciascuna } x_i \text{ sia riferita alla stessa pdf e quindi alla stessa varianza } \sigma^2\\
&= \frac{1}{n}\sum_i \sigma^2\\
&= \sigma^2
\end{aligned}
$$

La stima della varianza s^2, ottenuta da un set di n misure $\vec{x}=\{x_i\}$, porta subito a ottenere la stima della deviazione standard associabile a ciascuna misura, ovvero $\hat{\sigma}=\sqrt{s^2}$. Tale stima è assunta come l'errore statistico di ciascuna delle misure effettuate. Inoltre si può utilizzare $\hat{\sigma}$ nella (1.12) per stimare l'incertezza del valore medio delle misure (che come abbiamo detto stima il valore di aspettazione).

5.2 Calcolo della miglior stima di conteggi attesi col Metodo della Massima Verosimiglianza

Supponiamo di ripetere per N volte un esperimento di conteggio di uno stesso osservabile, sempre per un certo *live time* ΔT. Ciascun conteggio r_k relativo al k-esimo esperimento segue la distribuzione di Poisson (2.8), dove con μ si indica il valore di aspettazione dei conteggi relativo al live time ΔT.

Utilizzando il metodo della massima verosimiglianza (ML, dall'inglese *Maximum Likelihood*), si può calcolare la miglior stima di μ dai dati misurati.

La Likelihood L è ottenuta dalla *joint* p.d.f. $L=\prod_k^N P(r_k,\mu)$, con $\vec{r}=\{r_k, k\in(1,N)\}$ il set di misure disponibili.

In forma esplicita:

$$L(N, \vec{r}, \mu) = \prod_{k=1}^{N} \frac{e^{-\mu} \mu_k^r}{r_k!}$$

Si noti che il numero ed i risultati degli esperimenti effettuati fissano la L a meno del parametro μ che non conosciamo a priori. Il metodo della ML asserisce che la miglior stima di μ ottenibile con le misure fatte è quel valore che massimizza la L, quindi occorre derivare rispetto a μ. Prima però conviene prendere il logaritmo naturale della Likelihood, $\ln L$, sfruttando la monotonicità della funzione logaritmo:

$$\begin{aligned}
\ln L &= \sum_{k=1}^{N} \ln \frac{e^{-\mu} \mu_k^r}{r_k!} \\
&= \sum_{k=1}^{N} \left[\ln e^{-mu} + r_k \ln \mu + \text{cost} \right] \\
&= \sum_{k=1}^{N} \left[-\mu + r_k \ln \mu + \text{cost} \right] \\
&= -N\mu + \ln \mu \sum_{k=1}^{N} r_k + \text{cost}
\end{aligned}$$

Di conseguenza, derivando rispetto a μ:

$$\begin{aligned}
\frac{d \ln L}{d\mu} &= -N + \frac{1}{\mu} \sum_{k=1}^{N} r_k = 0 \\
\Rightarrow \mu_{best} &= \frac{\sum_{k=1}^{N} r_k}{N} \qquad (5.2)
\end{aligned}$$

Il risultato ottenuto ci indica che la miglior stima del valore atteso di eventi altro non è che il valor medio tra i conteggi misurati negli N esperimenti, da confrontare con quanto detto all'inizio della sezione 5.1. Facilmente si trova la varianza di μ_{best} :

$$\begin{aligned}
V[\mu_{best}] &= V\left[\frac{\sum_{k=1}^{N} r_k}{N} \right] \\
&= \frac{1}{N^2} \sum_{k=1}^{N} V[r_k] \\
&= \frac{\mu}{N} \qquad (5.3)
\end{aligned}$$

Si noti che gli ultimi due passaggi sono fatti assumendo legittimo $V[r_k] = \mu$, considerando r_k come una variabile aleatoria, e non una misura definita con un valore certo ed unico.

Posso però dare una stima della varianza (5.3), e quindi poter attribuire alla stima di μ_{best} un'incertezza statistica, proprio calcolando la varianza in base al μ_{best}: $V_{best} = \frac{\mu_{bset}}{N}$

Naturalmente, se facessi un solo esperimento ($N = 1$) e misurassi r conteggi, allora avrei:

$$\mu_{best} = r$$

$$V_{best} = r$$

Si confronti quest'ultimo risultato con quanto scritto nella sezione 2.2 al momento della discussione degli errori da attribuire al numero di eventi presenti in un bin di un istogramma.

5.3 Disuguaglianza di Rao-Cramer-Frechet

La disuguaglianza RCF è fondamentale perché fornisce il metodo per calcolare la varianza degli estimatori in modo generalizzato. La sua formulazione nel caso di un singolo parametro θ è la seguente:

$$V(\theta_{best}) \geq \frac{\left(1 + \frac{\partial b}{\partial \theta}\right)}{E\left[-\frac{\partial^2}{\partial \theta^2} \ln L\right]} \tag{5.4}$$

dove b è il *bias*.

Se l'estimatore è efficiente, ovvero ha la minima varianza, allora la diseguaglianza (5.4) coincide con l'uguaglianza.

Mettiamoci nel caso in cui anche il bias sia nullo $b = 0$, allora nel caso di più parametri $\vec{\theta}$ possiamo scrivere l'elemento generico della matrice di covarianza inversa $(V^{-1})_{ij}$ come:

$$(V^{-1})_{ij} \quad = \quad E\left[-\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln L\right] \tag{5.5}$$

Per scrivere esplicitamente l'operazione di media mostrata nella (5.5) comporta chidersi rispetto a cosa si sta facendo tale operazione di media: ovviamente la p.d.f. da considerare è la joint p.d.f. delle variabili aleatorie che vado a misurare. Se il nostro problema riguarda la misura di N variabili aleatorie $\{x_l, l \in (1, N)\}$, allora posso riscrivere la (5.5) come

$$(V^{-1})_{ij} \quad = \quad \int_{\Omega_1} \dots \int_{\Omega_N} \left(-\frac{\partial^2}{\partial \theta_i \partial \theta_j} \sum_{k=1}^{N} \ln f(x_k, \vec{\theta})\right) \prod_{l=1}^{N} f(x_l, \vec{\theta}) dx_l \tag{5.6}$$

Si noti che sia le x_k che le x_l sono da considerarsi variabili aleatorie, ovvero non stiamo fissando nulla al momento. Questo è consistente anche col modo di calcolare il valore di aspettazione e della varianza di un estimatore presentato nella sezione precedente. Adesso, sicuramente possiamo invertire l'ordine tra derivate parziali e sommatoria, poichè le derivate sono operatori lineari. Se le poi la log-likelihood e le p.d.f. sono funzioni regolari, possiamo portare fuori la sommatoria dagli integrali. A questo punto possiamo riscrivere la (5.6) :

$$(V^{-1})_{ij} \quad = \quad \sum_{k=1}^{N} \int_{\Omega_1} \dots \int_{\Omega_N} \left(-\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln f(x_k, \vec{\theta})\right) \prod_{l=1}^{N} f(x_l, \vec{\theta}) dx_l \tag{5.7}$$

A questo punto interpreto la (5.7) come una somma di integrali dove, volta a volta, k è fissato. Quindi in tutti i casi in cui $l \neq k$ possiamo separare gli integrali e porre:

$$(V^{-1})_{ij} = \sum_{k=1}^{N} \left[\prod_{l \neq k} \int_{\Omega_l} f(x_l, \vec{\theta}) dx_l \int_{\Omega_k} \left(-\frac{\partial^2}{\partial\theta_i \partial\theta_j} \ln f(x_k, \vec{\theta}) \right) f(x_k, \vec{\theta}) dx_k \right] \tag{5.8}$$

Nella (5.8) il termine relativo alla produttoria di integrali per $l \neq k$ è sempre uguale ad 1, per la definizione di p.d.f. . Infine, se le x_k variabili aleatorie si riferiscono sempre alla stessa variabile che misurerò N volte, e quindi associate ad una stessa p.d.f. $f(x, \vec{\theta})$, posso trasformare ancora la (5.8) interpretandola come uno stesso integrale sommato N volte:

$$(V^{-1})_{ij} = N \int_{\Omega} \left(-\frac{\partial^2}{\partial\theta_i \partial\theta_j} \ln f(x, \vec{\theta}) \right) f(x, \vec{\theta}) dx \tag{5.9}$$

Non sempre è possibile calcolare analiticamente la (5.5) o la (5.9), per questo, disponendo di un numero di dati sufficientemente grande, possiamo stimare la (V^{-1}) calcolando le derivate seconde della log-likelihood sostituendo alle variabili aleatorie i valori misurati e sostituendo ai parametri la loro miglior stima. In sostanza:

$$(V^{-1})_{ij} = -\frac{\partial^2}{\partial\theta_i \partial\theta_j} \ln L \bigg|_{\vec{x} \to \vec{x}_{meas}, \vec{\theta} \to \vec{\theta}'_{best}} \tag{5.10}$$

Ovviamente per ottenere le varianze e covarianze per i parametri occorre invertire la (V^{-1}), utilizzando i metodi di inversione delle matrici dell'algebra lineare.

5.4 Incertezza estimatori: il metodo grafico

Grazie ai risultati esposti nella sezione precedente possiamo trovare un metodo alternativo per trovare l'incertezza di un estimatore, utilizzando il grafico della Likelihood.

Supponiamo di sviluppare in serie di Taylor la nostra Likelihood attorno alla migliore stima di un parametro (per semplicità assumiamo che esista solo un parametro)

$$\begin{aligned} \ln L(\theta) \sim & \ln L(\theta_{best}) + \frac{d}{d\theta} \ln L(\theta) \bigg|_{\theta_{best}} (\theta - \theta_{best}) + \\ & + \frac{1}{2} \frac{d^2}{d\theta^2} \ln L(\theta) \bigg|_{\theta_{best}} (\theta - \theta_{best})^2 + O(\theta^2) \end{aligned} \tag{5.11}$$

Il secondo termine della somma al secondo membro della (5.11) si annulla, in quanto esso corrisponde alla marginalizzazione secondo la quale viene determinato θ_{best}. Pertanto:

$$\ln L(\theta) \sim \ln L(\theta_{best}) + \frac{1}{2} \frac{d^2}{d\theta^2} \ln L(\theta) \bigg|_{\theta_{best}} (\theta - \theta_{best})^2 + O(\theta^2) \tag{5.12}$$

Ricordando che $L(\theta_{best}) = L_{max}$, e che $\theta = \theta_{best} \pm \sigma_{best}$, possiamo scrivere:

$$\ln L(\theta_{best} \pm \sigma_{best}) \quad = \quad \ln L_{max} + \frac{1}{2} \frac{\sigma_{best}^2}{\frac{1}{\frac{d^2}{d\theta^2} \ln L\theta \Big|_{\theta_{best}}}} \tag{5.13}$$

Sostituendo nella (5.13) il risultato della (5.10), ovvero $\frac{1}{\frac{d^2}{d\theta^2} \ln L\theta \Big|_{\theta_{best}}} = -\sigma_{best}^2$ Si ottiene che

$$\ln L(\theta_{best} \pm \sigma_{best}) \quad \sim \quad \ln L_{max} - \frac{1}{2}. \tag{5.14}$$