La prima sommatoria comprende N addendi distinti; la seconda è estesa a tutte le possibili *combinazioni* dei valori distinti di i e j presi a due a due: è costituita quindi da

$$C_2^N = \frac{N\,(N-1)}{2}$$

addendi distinti.

Il fattore 2 che compare davanti ad essa è dovuto al fatto che una coppia di valori degli indici si presentava nella sommatoria su $i \neq j$ una volta come ${x_i}^2{x_j}^2$ e un'altra come ${x_j}^2{x_i}^2$, termini diversi per l'ordine ma con lo stesso valore. In definitiva, passando ai valori medi e tenendo conto dell'indipendenza statistica di x_i e x_j quando è $i \neq j$, risulta

$$E\left\{\left(\sum_i {x_i}^2\right)^2\right\} = N\,E(x^4) + N\,(N-1)\,\left[E(x^2)\right]^2 \quad .$$

Con simili passaggi, si ricava per il secondo termine

$$\begin{aligned}\left(\sum_i {x_i}^2\right)\left(\sum_j x_j\right)^2 &= \left(\sum_i {x_i}^2\right)\left(\sum_j {x_j}^2 + \sum_{\substack{j,k\\ j\neq k}} x_j x_k\right)\\ &= \sum_i {x_i}^4 + \sum_{\substack{i,j\\ i\neq j}} {x_i}^2{x_j}^2 + \sum_{\substack{i,j\\ i\neq j}} {x_i}^3 x_j + \sum_{\substack{i,j,k\\ i\neq j\neq k}} {x_i}^2 x_j\, x_k\end{aligned}$$

dove gli indici aventi simboli diversi si intendono avere anche valori sempre diversi tra loro nelle sommatorie.

Il valore medio del terzo e del quarto termine si annulla essendo $E(x) = 0$; inoltre gli addendi nella prima sommatoria sono in numero di N e quelli nella seconda in numero di $N\,(N-1)/2$ e vanno moltiplicati per un fattore 2. Pertanto anche

$$E\left\{\left(\sum_i x_i^2\right)\left(\sum_i x_i\right)^2\right\} = N\,E(x^4) + N\,(N-1)\,\left[E(x^2)\right]^2 \quad .$$

Infine avremo, con la medesima convenzione sugli indici,

$$\begin{aligned}\left(\sum_i x_i\right)^4 &= \left(\sum_i x_i\right)\left(\sum_j x_j\right)\left(\sum_k x_k\right)\left(\sum_l x_l\right)\\ &= \sum_i {x_i}^4 + \sum_{\substack{i,j\\ i\neq j}} {x_i}^3 x_j + \sum_{\substack{i,j\\ i\neq j}} {x_i}^2{x_j}^2 + \sum_{\substack{i,j,k\\ i\neq j\neq k}} {x_i}^2 x_j\, x_k + \sum_{\substack{i,j,k,l\\ i\neq j\neq k\neq l}} x_i\, x_j\, x_k\, x_l \quad .\end{aligned}$$

I valori medi del secondo, quarto e quinto termine (che contengono potenze dispari delle x) sono nulli. Gli addendi nella prima sommatoria sono

in numero di N; nella terza vi sono $N\,(N-1)/2$ termini distinti: ma ciascuno appare in 6 modi diversi solo per l'ordine, corrispondenti al numero C_2^4 di combinazioni dei quattro indici originari i, j, k ed l presi a due a due. Allora

$$E\left(\sum_i x_i\right)^4 = N\,E(x^4) + 3\,N\,(N-1)\,\left[E(x^2)\right]^2 \quad ;$$

e, riprendendo la formule di partenza,

$$E(s^4) = \frac{(N-1)^2}{N^3}\,E(x^4) + \frac{(N-1)(N^2-2N+3)}{N^3}\,\left[E(x^2)\right]^2 \quad .$$

Per il valore medio di s^2, già sappiamo come risulti per la varianza del campione

$$E(s^2) = \sigma^2 - {\sigma_{\bar{x}}}^2$$

inoltre

$$\sigma^2 \;=\; E\left\{(x-x^*)^2\right\} \;=\; E(x^2)$$

(essendo $x^* = 0$) e

$${\sigma_{\bar{x}}}^2 \;=\; E\left\{(\bar{x}-x^*)^2\right\} \;=\; \frac{\sigma^2}{N}$$

da cui abbiamo ottenuto a suo tempo la

$$E(s^2) \;=\; \frac{N-1}{N}\,\sigma^2 \;=\; \frac{N-1}{N}\,E(x^2) \quad .$$

Per la varianza di s^2, che vogliamo determinare:

$$\begin{aligned}
\mathrm{Var}(s^2) &= E(s^4) - \left[E(s^2)\right]^2 \\
&= \frac{(N-1)^2}{N^3}\,E(x^4) + \\
&\quad + \left[\frac{(N-1)(N^2-2N+3)}{N^3} - \frac{(N-1)^2}{N^2}\right]\left[E(x^2)\right]^2 \\
&= \frac{(N-1)^2}{N^3}\,E(x^4) \;-\; \frac{(N-1)(N-3)}{N^3}\,\left[E(x^2)\right]^2 \\
&= \frac{N-1}{N^3}\,\left\{(N-1)\,E(x^4) \;-\; (N-3)\,\left[E(x^2)\right]^2\right\} \quad .
\end{aligned}$$

Questa relazione ha validità generale. *Nel caso poi che la popolazione ubbidisca alla legge normale*, potremo calcolare il valore medio di x^4 usando la forma analitica della funzione di Gauss: per distribuzioni normali qualsiasi, i momenti di ordine pari rispetto alla media sono dati dalla formula (8.5), che qui ricordiamo:

$$\mu_{2k} = E\left\{[x - E(x)]^{2k}\right\} = \frac{(2k)!}{2^k\, k!}\,{\mu_2}^k = \frac{(2k)!}{2^k\, k!}\,\sigma^{2k} \quad .$$

Per la varianza di s^2 se ne ricava

$$E(x^4) = 3\sigma^4$$

e, sostituendo,

$$\mathrm{Var}(s^2) = \frac{2(N-1)}{N^2}\left[E(x^2)\right]^2 = \frac{2(N-1)}{N^2}\,\sigma^4 \quad ;$$

insomma *l'errore quadratico medio della varianza s^2 del campione* vale

$$\sigma_{s^2} = \frac{\sqrt{2(N-1)}}{N}\,\sigma^2 \quad .$$

La varianza, invece, della stima della varianza della popolazione

$$\sigma^2 = \frac{N}{N-1}\,s^2$$

vale

$$\mathrm{Var}(\sigma^2) = \left(\frac{N}{N-1}\right)^2 \mathrm{Var}(s^2) = \frac{2}{N-1}\,\sigma^4 \quad ;$$

ed infine *l'errore quadratico medio della stima della varianza della popolazione* ricavata dal campione è

$$\boxed{\sigma_{\sigma^2} = \sqrt{\frac{2}{N-1}}\,\sigma^2}$$

Sottolineiamo ancora come queste formule che permettono di calcolare, per una popolazione *avente distribuzione normale*, gli errori quadratici medi sia della varianza di un campione di N misure che della stima della varianza della popolazione ricavata da un campione di N misure, siano *esatte*.

Se si vuole invece calcolare l'errore da attribuire agli *errori quadratici medi*, cioè alle quantità s e σ radici quadrate delle varianze di cui sopra, non è possibile dare delle formule esatte: la ragione ultima è che il valore medio di s non può essere espresso in forma semplice in termini di grandezze caratteristiche della popolazione.

Per questo motivo *è sempre meglio riferirsi ad errori di varianze* piuttosto che ad errori di scarti quadratici medi; comunque, in prima approssimazione, l'errore di σ si può ricavare da quello su σ^2 usando la formula di propagazione:

$$\mathrm{Var}(\sigma) \approx \left(\frac{1}{\frac{\mathrm{d}(\sigma^2)}{\mathrm{d}\sigma}}\right)^2 \mathrm{Var}(\sigma^2) = \frac{1}{4\,\sigma^2}\,\mathrm{Var}(\sigma^2) = \frac{\sigma^2}{2\,(N-1)} \;;$$

cioè

$$\boxed{\sigma_\sigma \approx \frac{\sigma}{\sqrt{2\,(N-1)}}} \tag{B.1}$$

(il fatto che questa formula sia approssimata risulta chiaramente se si considera che la relazione tra σ^2 e σ è non lineare).

Una conseguenza dell'equazione (B.1) è che l'errore relativo di σ dipende *solo dal numero di misure*; diminuisce poi all'aumentare di esso, ma questa diminuzione è inversamente proporzionale alla radice quadrata di N e risulta perciò lenta.

In altre parole, per diminuire l'errore relativo di σ di un ordine di grandezza occorre aumentare il numero delle misure di *due* ordini di grandezza; σ_σ/σ è (circa) il 25% per 10 misure, il 7% per 100 misure ed il 2% per 1000 misure effettuate: e questo è sostanzialmente il motivo per cui, di norma, si scrive l'errore quadratico medio *dando per esso una sola cifra significativa*.

Due cifre significative *reali* per σ corrisponderebbero infatti ad un suo errore relativo compreso tra il 5% (se la prima cifra significativa di σ è 1, ad esempio $\sigma = 10 \pm 0.5$) e lo 0.5% ($\sigma = 99 \pm 0.5$); e presupporrebbero quindi che siano state effettuate almeno 200 misure nel caso più favorevole e quasi 20˙000 in quello più sfavorevole.

Appendice C

Covarianza e correlazione

C.1 La covarianza

Per due variabili casuali x ed y si definisce la *covarianza*, che si indica con uno dei due simboli $\mathrm{Cov}(x, y)$ o K_{xy}, nel seguente modo:

$$\begin{aligned}\mathrm{Cov}(x, y) &= E\Big\{[x - E(x)]\,[y - E(y)]\Big\} \\ &= E(xy) - E(x) \cdot E(y) \quad .\end{aligned}$$

Per provare l'equivalenza delle due forme, basta osservare che[1]

$$\begin{aligned}\mathrm{Cov}(x, y) &= E\Big\{[x - E(x)]\,[y - E(y)]\Big\} \\ &= \textstyle\sum_{ij} P_{ij}\,[x_i - E(x)]\,[y_j - E(y)] \\ &= \textstyle\sum_{ij} P_{ij}\,x_i\,y_j \;-\; E(x)\sum_{ij} P_{ij}\,y_j \;-\; E(y)\sum_{ij} P_{ij}\,x_i \;+ \\ &\qquad + E(x)\,E(y)\textstyle\sum_{ij} P_{ij} \\ &= E(xy) \;-\; E(x)\textstyle\sum_j q_j\,y_j \;-\; E(y)\sum_i p_i\,x_i \;+\; E(x)\,E(y) \\ &= E(xy) \;-\; E(x) \cdot E(y)\end{aligned}$$

ricordando alcune relazioni già ricavate nel capitolo 5, e valide per variabili casuali *qualunque*: in particolare, anche *non* statisticamente indipendenti.

[1] Nel seguito useremo per la varianza, per le probabilità di ottenere i vari valori x_i o y_j e così via, le stesse notazioni già introdotte nel capitolo 5.

È chiaro come per variabili *statisticamente indipendenti* la covarianza sia nulla: infatti per esse vale la

$$E(xy) \;=\; \sum_{ij} P_{ij}\, x_i\, y_j \;=\; \sum_{ij} p_i\, q_j\, x_i\, y_j \;=\; E(x)\cdot E(y) \;\;.$$

Non è però vero l'inverso: consideriamo ad esempio le due variabili casuali x ed $y = x^2$, ovviamente dipendenti l'una dall'altra: la loro covarianza vale

$$\mathrm{Cov}(x,y) \;=\; E(xy) - E(x)\cdot E(y) \;=\; E(x^3) - E(x)\cdot E(x^2)$$

ed è chiaramente nulla per qualunque variabile casuale x con distribuzione simmetrica rispetto allo zero; quindi l'annullarsi della covarianza è condizione *necessaria ma non sufficiente* per l'indipendenza statistica di due variabili casuali.

Possiamo ora calcolare la varianza delle combinazioni lineari di due variabili casuali qualunque, estendendo la formula già trovata nel capitolo 5 nel caso particolare di variabili statisticamente indipendenti; partendo ancora da due variabili x e y con media zero per semplificare i calcoli, per la loro combinazione lineare $z = ax + by$ valgono le:

$$\begin{aligned}
E(z) \;&=\; a\,E(x) + b\,E(y) \;=\; 0 \\
\mathrm{Cov}(x,y) \;&=\; E(xy) - E(x)\cdot E(y) \;=\; E(xy) \\
\mathrm{Var}(z) \;&=\; E\left\{\left[z - E(z)\right]^2\right\} \\
&=\; E(z^2) \\
&=\; E\left[(ax+by)^2\right] \\
&=\; \sum_{ij} P_{ij}(a\,x_i + b\,y_j)^2 \\
&=\; a^2 \sum_{ij} P_{ij}\, x_i^2 \;+\; b^2 \sum_{ij} P_{ij}\, y_j^2 \;+\; 2ab \sum_{ij} P_{ij}\, x_i\, y_j \\
&=\; a^2 \sum_{i} p_i x_i^2 \;+\; b^2 \sum_{j} q_j y_j^2 \;+\; 2ab\, E(xy)
\end{aligned}$$

ed infine

$$\mathrm{Var}(z) \;=\; a^2\,\mathrm{Var}(x) + b^2\,\mathrm{Var}(y) + 2ab\,\mathrm{Cov}(x,y) \;\;. \qquad \text{(C.1)}$$

Questa si estende immediatamente a variabili casuali con media qualsiasi: introducendo ancora le variabili ausiliarie

$$\xi = x - E(x) \qquad \text{ed} \qquad \eta = y - E(y)$$

per le quali già sappiamo che vale la

$$E(\xi) \;=\; E(\eta) \;=\; 0$$

con le

$$\mathrm{Var}(\xi) = \mathrm{Var}(x) \qquad \text{e} \qquad \mathrm{Var}(\eta) = \mathrm{Var}(y) \;\;;$$

basta osservare infatti che vale anche la

$$\mathrm{Cov}(x, y) \;=\; E\Big\{[x - E(x)]\,[y - E(y)]\Big\} \;=\; \mathrm{Cov}(\xi, \eta) \;\;.$$

La (C.1) si può poi generalizzare, per induzione completa, ad una variabile z definita come combinazione lineare di un numero qualsiasi N di variabili casuali: si trova che, se

$$z = \sum_{i=1}^{N} a_i\, x_i$$

risulta

$$\mathrm{Var}(z) = \sum_{i} a_i{}^2\, \mathrm{Var}(x_i) + \sum_{\substack{i,j \\ j>i}} 2\, a_i\, a_j\, \mathrm{Cov}(x_i, x_j) \;\;. \tag{C.2}$$

Per esprimere in modo compatto la (C.2), si ricorre in genere ad una notazione che usa la cosiddetta *matrice delle covarianze* delle variabili x; ovverosia una matrice quadrata $\boldsymbol{V}$ di ordine N, in cui il generico elemento V_{ij} è uguale alla covarianza delle variabili casuali x_i e x_j:

$$V_{ij} \;=\; \mathrm{Cov}(x_i, x_j) \;=\; E(x_i \cdot x_j) - E(x_i) \cdot E(x_j) \;\;. \tag{C.3}$$

La matrice è ovviamente *simmetrica* ($V_{ij} = V_{ji}$); e, in particolare, gli elementi diagonali V_{ii} valgono

$$V_{ii} \;=\; E(x_i{}^2) - \big[E(x_i)\big]^2 \;\equiv\; \mathrm{Var}(x_i) \;\;.$$

Consideriamo poi le a_i come le N componenti di un *vettore* $\boldsymbol{A}$ di dimensione N (che possiamo concepire come una matrice rettangolare di N righe ed una colonna); ed introduciamo la matrice *trasposta* di $\boldsymbol{A}$, $\widetilde{\boldsymbol{A}}$, che è una matrice rettangolare di una riga ed N colonne i cui elementi valgono

$$\widetilde{A}_i = A_i \;\;.$$

Possiamo allora scrivere la (C.2) nella forma

$$\mathrm{Var}(z) = \sum_{i,j} \widetilde{A}_i \, V_{ij} \, A_j$$

(la simmetria di $\boldsymbol{V}$ e quella tra $\widetilde{\boldsymbol{A}}$ ed $\boldsymbol{A}$ produce, nello sviluppo delle sommatorie, il fattore 2 che moltiplica le covarianze); o anche, ricordando le regole del prodotto tra matrici,

$$\boxed{\mathrm{Var}(z) \;=\; \widetilde{\boldsymbol{A}} \, \boldsymbol{V} \, \boldsymbol{A}}$$

Si può poi facilmente dimostrare il seguente teorema, che ci sarà utile più avanti:

> TEOREMA: *due differenti combinazioni lineari delle stesse variabili casuali sono sempre correlate.*

Infatti, dette A e B le due combinazioni lineari:

$$A = \sum_{i=1}^{N} a_i \, x_i \qquad \Rightarrow \qquad E(A) = \sum_{i=1}^{N} a_i \, E(x_i)$$

$$B = \sum_{j=1}^{N} b_j \, x_j \qquad \Rightarrow \qquad E(B) = \sum_{j=1}^{N} b_j \, E(x_j)$$

abbiamo che la covarianza di A e B vale

$$\begin{aligned}
\mathrm{Cov}(A,B) &= E\Big\{\big[A - E(A)\big]\big[B - E(B)\big]\Big\} \\
&= E\Big\{\sum_{i,j} a_i \, b_j \big[x_i - E(x_i)\big]\big[x_j - E(x_j)\big]\Big\} \\
&= \sum_{i,j} a_i \, b_j \, E\Big\{\big[x_i - E(x_i)\big]\big[x_j - E(x_j)\big]\Big\} \\
&= \sum_{i} a_i \, b_i \; \mathrm{Var}(x_i) + \sum_{\substack{i,j \\ i \neq j}} a_i \, b_j \; \mathrm{Cov}(x_i, x_j)
\end{aligned}$$

e non è in genere nulla. In forma matriciale e con ovvio significato dei simboli,

$$\mathrm{Cov}(A,B) = \widetilde{\boldsymbol{A}} \, \boldsymbol{V} \, \boldsymbol{B} \;\; .$$

È da notare come A e B siano di norma sempre correlate *anche se le variabili di partenza x_i sono tutte tra loro statisticamente indipendenti*: in questo caso infatti tutti i termini non diagonali della matrice delle covarianze si annullano, e risulta

$$\mathrm{Cov}(A,B) = \sum_{i=1}^{N} a_i\, b_i\, \mathrm{Var}(x_i) \quad . \tag{C.4}$$

C.2 La correlazione lineare

Per due variabili casuali qualunque si definisce poi il *coefficiente di correlazione lineare* $\mathrm{Corr}(x,y)$ (anche indicato col simbolo r_{xy}, o semplicemente come r) nel modo seguente:

$$r_{xy} \equiv \mathrm{Corr}(x,y) = \frac{\mathrm{Cov}(x,y)}{\sqrt{\mathrm{Var}(x)\,\mathrm{Var}(y)}} = \frac{\mathrm{Cov}(x,y)}{\sigma_x\,\sigma_y} \quad .$$

Il coefficiente di correlazione di due variabili è ovviamente adimensionale; è nullo quando le variabili stesse sono statisticamente indipendenti (visto che è zero la loro covarianza); ed è comunque compreso tra i due limiti -1 e $+1$. Che valga quest'ultima proprietà si può dimostrare calcolando dapprima la varianza di una variabile casuale ausiliaria z definita attraverso la relazione $z = \sigma_y\, x - \sigma_x\, y$, ed osservando che essa deve essere una quantità non negativa:

$$\begin{aligned}
\mathrm{Var}(z) &= {\sigma_y}^2\,\mathrm{Var}(x) + {\sigma_x}^2\,\mathrm{Var}(y) - 2\,\sigma_x\,\sigma_y\,\mathrm{Cov}(x,y) \\
&= 2\,\mathrm{Var}(x)\,\mathrm{Var}(y) - 2\,\sigma_x\,\sigma_y\,\mathrm{Cov}(x,y) \\
&\geq 0 \quad ;
\end{aligned}$$

da cui

$$\mathrm{Corr}(x,y) \leq 1 \quad .$$

Poi, compiendo analoghi passaggi su un'altra variabile definita stavolta come $z = \sigma_y\, x + \sigma_x\, y$, si troverebbe che deve essere anche $\mathrm{Corr}(x,y) \geq -1$.

Se il coefficiente di correlazione lineare raggiunge uno dei due valori estremi ± 1, risulta $\mathrm{Var}(z) = 0$; e dunque deve essere

$$z = \sigma_y\, x \mp \sigma_x\, y = \text{costante}$$

cioè x ed y devono essere legati da una relazione funzionale *di tipo lineare*.

Vale anche l'inverso: partendo infatti dall'ipotesi che le due variabili siano legate da una relazione lineare data da $y = a + bx$, con b finito e non nullo, ne consegue che:

$$
\begin{aligned}
E(y) &= a + b\,E(x) \\
\mathrm{Var}(y) &= b^2\ \mathrm{Var}(x) \\
E(xy) &= E(a\,x + b\,x^2) \\
&= a\,E(x) + b\,E(x^2) \\
\mathrm{Cov}(x,y) &= E(xy) - E(x)\cdot E(y) \\
&= a\,E(x) + b\,E(x^2) - E(x)[a + b\,E(x)] \\
&= b\left\{E(x^2) - [E(x)]^2\right\} \\
&= b\cdot \mathrm{Var}(x) \\
\mathrm{Corr}(x,y) &= \frac{\mathrm{Cov}(x,y)}{\sqrt{\mathrm{Var}(x)\ \mathrm{Var}(y)}} \\
&= \frac{b\ \mathrm{Var}(x)}{\sqrt{b^2[\mathrm{Var}(x)]^2}} \\
&= \frac{b}{|b|} \\
&= \pm 1\ .
\end{aligned}
$$

Il segno del coefficiente di correlazione è quello del coefficiente angolare della retta. Sono da notare due cose: innanzi tutto il rapporto $b/|b|$ perde significato quando $b = 0$ o quando $b = \infty$, cioè quando la retta è parallela ad uno degli assi coordinati: in questi casi (x = costante o y = costante) una delle due grandezze non è in realtà una variabile casuale, e l'altra è dunque indipendente da essa; è facile vedere che tanto il coefficiente di correlazione tra x e y quanto la covarianza valgono zero, essendo $E(xy) \equiv E(x)\cdot E(y)$ in questo caso.

Anche quando esiste una relazione funzionale esatta tra x e y, se questa non è rappresentata da una funzione lineare il coefficiente di correlazione non raggiunge i valori estremi ± 1; per questa ragione appunto esso si chiama più propriamente "coefficiente di correlazione *lineare*".

C.3 Propagazione degli errori per variabili correlate

Vediamo ora come si può ricavare una formula di propagazione per gli errori (da usare in luogo dell'equazione (10.2) che abbiamo incontrato a pagina 164) se le grandezze fisiche misurate direttamente non sono tra loro statisticamente indipendenti; nel corso di questo paragrafo continueremo ad usare la notazione già introdotta nel capitolo 10.

Consideriamo una funzione F di N variabili, $F = F(x_1, x_2, \ldots, x_N)$; ed ammettiamo che sia lecito svilupparla in serie di Taylor nell'intorno del punto $(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_N)$ trascurando i termini di ordine superiore al primo (questo avviene, come sappiamo, o se gli errori di misura sono piccoli o se F è lineare rispetto a tutte le variabili). Tenendo presente il teorema di pagina 52, ed applicando alla formula dello sviluppo

$$F(x_1, x_2, \ldots, x_N) \approx F(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_N) + \sum_{i=1}^{N} \frac{\partial F}{\partial x_i} (x_i - \bar{x}_i)$$

l'equazione (C.2), otteniamo

$$\mathrm{Var}(F) \approx \sum_i \left(\frac{\partial F}{\partial x_i}\right)^2 \mathrm{Var}(x_i) + 2 \sum_{\substack{i,j \\ j>i}} \frac{\partial F}{\partial x_i} \frac{\partial F}{\partial x_j} \mathrm{Cov}(x_i, x_j) \ . \qquad \text{(C.5)}$$

Per esprimere in modo compatto la (C.5), si può ricorrere ancora alla matrice delle covarianze $\boldsymbol{V}$ delle variabili x_i; ricordandone la definizione (data dall'equazione (C.3) a pagina 257) ed introducendo poi un vettore $\boldsymbol{F}$ di dimensione N di componenti

$$F_i = \frac{\partial F}{\partial x_i}$$

ed il suo trasposto $\widetilde{\boldsymbol{F}}$, la (C.5) si può riscrivere nella forma

$$\mathrm{Var}(F) = \sum_{i,j} \widetilde{F}_i \, V_{ij} \, F_j$$

ossia

$$\boxed{\mathrm{Var}(F) = \widetilde{\boldsymbol{F}} \, \boldsymbol{V} \, \boldsymbol{F}}$$

C.4 Applicazioni all'interpolazione lineare

Riprendiamo adesso il problema dell'interpolazione lineare, già discusso nel capitolo 11: si sia cioè compiuto un numero N di misure indipendenti di coppie di valori di due grandezze fisiche x e y, tra le quali si ipotizza che esista una relazione funzionale di tipo lineare data da $y = a + bx$. Supponiamo inoltre che siano valide le ipotesi esposte nel paragrafo 11.4.1; in particolare che le x_i siano prive di errore, e che le y_i siano affette da errori normali e tutti uguali tra loro.

C.4.1 Riscrittura delle equazioni dei minimi quadrati

Sebbene i valori della x siano scelti dallo sperimentatore e privi di errore, e non siano pertanto variabili casuali in senso stretto; e sebbene la variabilità delle y sia dovuta non solo agli errori casuali di misura ma anche alla variazione della x, introduciamo ugualmente (in maniera *puramente formale*) le medie e le varianze degli N valori x_i e y_i, date dalle espressioni

$$\bar{x} = \frac{\sum_i x_i}{N} \qquad \text{e} \qquad \mathrm{Var}(x) = \frac{\sum_i (x_i - \bar{x})^2}{N} = \frac{\sum_i {x_i}^2}{N} - \bar{x}^2$$

(e simili per la y); e la covarianza di x e y, data dalla

$$\mathrm{Cov}(x, y) = \frac{\sum_i x_i y_i}{N} - \bar{x}\bar{y} \ .$$

Queste grandezze permettono di riscrivere le equazioni (11.9) risolutive del problema dell'interpolazione lineare per un insieme di dati, che abbiamo già incontrato a pagina 181, nella forma

$$\begin{cases} a + b\bar{x} = \bar{y} \\ a\bar{x} + b\left[\mathrm{Var}(x) + \bar{x}^2\right] = \mathrm{Cov}(x, y) + \bar{x}\bar{y} \end{cases}$$

La prima equazione intanto implica che la retta interpolante deve passare per il punto $(\bar{x}, \bar{y})$ le cui coordinate sono le medie dei valori misurati delle due variabili in gioco; poi, ricavando da essa $a = \bar{y} - b\bar{x}$ e sostituendo nella seconda equazione, dopo aver semplificato alcuni termini si ottiene la soluzione per l'altra incognita:

$$b = \frac{\mathrm{Cov}(x, y)}{\mathrm{Var}(x)} \equiv \mathrm{Corr}(x, y)\sqrt{\frac{\mathrm{Var}(y)}{\mathrm{Var}(x)}} \tag{C.6}$$

e la retta interpolante ha quindi equazione

$$y = a + bx = (\bar{y} - b\bar{x}) + bx$$

o anche

$$(y - \bar{y}) \;=\; b\,(x - \bar{x})$$

(in cui b ha il valore (C.6)). Introduciamo ora le due variabili casuali ausiliarie $\xi = x - \bar{x}$ e $\eta = y - \bar{y}$, per le quali valgono le

$$\bar{\xi} \;=\; 0 \qquad\qquad \text{e} \qquad\qquad \mathrm{Var}(\xi) \;=\; \mathrm{Var}(x)$$

(con le analoghe per η ed y), ed inoltre la

$$\mathrm{Cov}(\xi, \eta) = \mathrm{Cov}(x, y)$$

ed indichiamo poi con $\hat{y}_i$ il valore della y sulla retta interpolante in corrispondenza dell'ascissa x_i:

$$\hat{y}_i \;=\; a + bx_i \;=\; \bar{y} + b\,(x_i - \bar{x}) \tag{C.7}$$

e con δ_i la differenza $\hat{y}_i - y_i$. Le differenze δ_i prendono il nome di *residui*, e di essi ci occuperemo ancora più avanti; risulta che

$$\begin{aligned}
\textstyle\sum_i \delta_i{}^2 &= \textstyle\sum_i \Big\{[\bar{y} + b\,(x_i - \bar{x})] - y_i\Big\}^2 \\
&= \textstyle\sum_i (b\xi_i - \eta_i)^2 \\
&= b^2 \textstyle\sum_i \xi_i{}^2 + \textstyle\sum_i \eta_i{}^2 - 2b \textstyle\sum_i \xi_i\eta_i \\
&= N\,b^2\ \mathrm{Var}(\xi) + N\ \mathrm{Var}(\eta) - 2Nb\ \mathrm{Cov}(\xi, \eta) \\
&= N\,b^2\ \mathrm{Var}(x) + N\ \mathrm{Var}(y) - 2Nb\ \mathrm{Cov}(x, y) \\
&= N\ \mathrm{Var}(x) \left[\frac{\mathrm{Cov}(x, y)}{\mathrm{Var}(x)}\right]^2 + N\ \mathrm{Var}(y) - 2N\,\frac{\mathrm{Cov}(x, y)}{\mathrm{Var}(x)}\ \mathrm{Cov}(x, y) \\
&= N \left\{\mathrm{Var}(y) - \frac{[\mathrm{Cov}(x, y)]^2}{\mathrm{Var}(x)}\right\} \\
&= N\ \mathrm{Var}(y)\ (1 - r^2)
\end{aligned}$$

in cui r è il coefficiente di correlazione lineare calcolato usando, sempre solo formalmente, i campioni dei valori misurati delle x e delle y.

Visto che quest'ultimo, nel calcolo dell'interpolazione lineare fatto con le calcolatrici da tasca, viene in genere dato come sottoprodotto dell'algoritmo,

la sua conoscenza permette (usando questa formula) di ottenere facilmente l'errore a posteriori sulle ordinate interpolate (studiato nel paragrafo 11.4.2):

$$\boxed{\mu_y = \sqrt{\frac{\sum_i \delta_i^2}{N-2}} = \sqrt{\frac{N}{N-2}\,\mathrm{Var}(y)\,(1-r^2)}} \tag{C.8}$$

oltre a fornire una grossolana stima dell'allineamento dei punti; quanto più infatti esso è rigoroso, tanto più r si avvicina a ± 1.

Il valore μ_y dell'errore rappresenta sempre anche una stima dell'allineamento dei punti, a differenza del coefficiente di correlazione lineare, per cui $r = \pm 1$ implica allineamento perfetto, ma *non* inversamente: potendo essere ad esempio (punti su di una retta parallela all'asse x) $r = 0$ e $\mathrm{Var}(y) = 0$ ancora con allineamento perfetto.

È opportuno qui osservare che r *non* è il coefficiente di correlazione fra gli *errori* delle grandezze x ed y; infatti, per ipotesi, la x non è affetta da errore e, se pur lo fosse, la correlazione fra gli errori sarebbe nulla qualora ciascuna x_i fosse misurata indipendentemente dalla corrispondente y_i o, altrimenti, sarebbe tanto più prossima ai valori estremi ± 1 quanto maggiore fosse il numero di cause d'errore comuni alle misure di x e di y.

Invece r è il coefficiente di correlazione per l'insieme dei punti aventi coordinate date dalle coppie di valori misurati, e nell'ipotesi di effettiva dipendenza lineare delle grandezze x ed y sarebbe sempre rigorosamente uguale a ± 1 se non intervenissero gli errori sperimentali.

C.4.2 Verifica di ipotesi sulla correlazione lineare

Non è pensabile di poter svolgere una teoria completa della correlazione in queste pagine; tuttavia talvolta un fisico si trova nella necessità di verificare delle ipotesi statistiche sul coefficiente di correlazione lineare ricavato sperimentalmente da un insieme di N coppie di valori misurati $\{x_i, y_i\}$. Nel seguito riassumiamo, senza darne alcuna dimostrazione, alcune proprietà di questi coefficienti:

- Se si vuole verificare che la correlazione tra x ed y sia significativamente differente da zero, si può calcolare il valore della variabile casuale
$$t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}}$$
che è distribuita secondo Student con $(N-2)$ gradi di libertà, e controllare la sua compatibilità con lo zero.

- Rispetto al valore vero ρ della correlazione lineare tra le due variabili, il valore r ricavato da un insieme di N coppie $\{x_i, y_i\}$ estratte a caso dalle rispettive popolazioni è tale che la variabile casuale

$$Z(r) \;=\; \ln\sqrt{\frac{1+r}{1-r}} \;=\; \frac{1}{2}\left[\ln(1+r) - \ln(1-r)\right] \tag{C.9}$$

(detta *variabile di Fisher*) segue una distribuzione approssimativamente normale con valore medio e varianza date da

$$E(Z) = Z(\rho) \qquad \text{e} \qquad \mathrm{Var}(Z) = {\sigma_Z}^2 = \frac{1}{N-3}$$

rispettivamente; la trasformazione inversa della (C.9) è la

$$r \;=\; \frac{e^{2Z}-1}{e^{2Z}+1} \;.$$

Quindi:

  - per verificare se il valore vero della correlazione può essere una quantità prefissata ρ, si controlla la compatibilità con la distribuzione normale $N(Z(\rho), \sigma_Z)$ del valore ottenuto $Z(r)$;
  - per calcolare un intervallo di valori corrispondente ad un certo livello di confidenza, si usano i corrispondenti intervalli per la distribuzione normale con deviazione standard σ_Z;
  - per verificare se due coefficienti di correlazione lineare r_1 ed r_2, ricavati da N_1 ed N_2 coppie di valori $\{x_i, y_i\}$ rispettivamente, siano o meno significativamente differenti, si calcola la variabile casuale

$$\delta \;=\; \frac{Z_1 - Z_2}{\sqrt{\mathrm{Var}(Z_1) + \mathrm{Var}(Z_2)}} \;=\; \frac{Z_1 - Z_2}{\sqrt{\dfrac{1}{N_1-3} + \dfrac{1}{N_2-3}}}$$

(ove Z_1 e Z_2 sono le variabili di Fisher ricavate dalla (C.9) per i due campioni; δ segue asintoticamente la distribuzione normale con media $E(Z_1) - E(Z_2)$ e varianza 1) e si verifica se il risultato ottenuto è compatibile con lo zero.

C.4.3 La correlazione tra i coefficienti della retta

Notiamo anche che l'intercetta e la pendenza a e b della retta interpolante un insieme di punti sperimentali, essendo ottenute come combinazioni

lineari delle stesse variabili casuali, sono tra loro correlate (come osservato nel paragrafo C.1); le equazioni (11.10) possono essere riscritte (l'abbiamo già visto nel paragrafo 11.4.1) come

$$a = \sum_i a_i\, y_i \qquad \text{e} \qquad b = \sum_i b_i\, y_i$$

una volta che si sia posto

$$\begin{cases} a_i = \dfrac{1}{\Delta}\left[\sum_j {x_j}^2 - \left(\sum_j x_j\right) x_i\right] \\ b_i = \dfrac{1}{\Delta}\left[N\, x_i - \sum_j x_j\right] \end{cases}$$

Ricordando la definizione di Δ, possiamo esprimerlo come

$$\begin{aligned} \Delta &= N\left(\sum_j {x_j}^2\right) - \left(\sum_j x_j\right)^2 \\ &= N^2\left[\frac{\sum_j {x_j}^2}{N} - \left(\frac{\sum_j x_j}{N}\right)^2\right] \end{aligned}$$

ed infine come

$$\Delta = N^2\ \mathrm{Var}(x)\ . \tag{C.10}$$

Visto che le y_i sono per ipotesi statisticamente indipendenti tra loro, possiamo applicare la (C.4); ne ricaviamo, sostituendovi la (C.10) e ricordando che gli errori sono tutti uguali, che

$$\begin{aligned} \mathrm{Cov}(a,b) &= \sum_i a_i\, b_i\ \mathrm{Var}(y_i) \\ &= \frac{{\sigma_y}^2}{\Delta^2}\sum_i\left[\sum_j {x_j}^2 - \left(\sum_j x_j\right) x_i\right]\left[N x_i - \sum_j x_j\right] \\ &= \frac{{\sigma_y}^2}{\Delta^2}\sum_i\Bigg[\left(\sum_j {x_j}^2\right) N x_i - \left(\sum_j {x_j}^2\right)\left(\sum_j x_j\right) - \\ &\qquad - \left(\sum_j x_j\right) N {x_i}^2 + \left(\sum_j x_j\right)^2 x_i\Bigg] \\ &= \frac{{\sigma_y}^2}{\Delta^2}\Bigg[N\left(\sum_j {x_j}^2\right)\left(\sum_i x_i\right) - N\left(\sum_j {x_j}^2\right)\left(\sum_j x_j\right) - \\ &\qquad - N\left(\sum_j x_j\right)\left(\sum_i {x_i}^2\right) + \left(\sum_j x_j\right)^3\Bigg] \\ &= \frac{{\sigma_y}^2}{\Delta^2}\left(\sum_j x_j\right)\left[\left(\sum_j x_j\right)^2 - N\left(\sum_j {x_j}^2\right)\right] \end{aligned}$$

ed infine, vista la definizione di Δ,

$$\mathrm{Cov}(a,b) = -\frac{\sum_j x_j}{\Delta}\,{\sigma_y}^2 \ ; \tag{C.11}$$

o anche

$$\mathrm{Cov}(a,b) = -\frac{\bar{x}}{N}\,\frac{{\sigma_y}^2}{\mathrm{Var}(x)} \ ,$$

diversa da zero se $\bar{x} \neq 0$; inoltre il segno della correlazione tra a e b è opposto a quello di $\bar{x}$. Questo non sorprende: la retta interpolante deve necessariamente passare per il punto $(\bar{x}, \bar{y})$, come abbiamo notato nel paragrafo C.4.1: se $\bar{x}$ è positivo, aumentando la pendenza della retta deve diminuire il valore della sua intercetta; e viceversa per $\bar{x} < 0$.

C.4.4 Stima puntuale mediante l'interpolazione lineare

Bisogna tener presente che le formule dei minimi quadrati ci danno l'equazione della retta che meglio approssima la relazione tra le due variabili nella parte di piano in cui esistono punti misurati; ma non ci dicono nulla sulla dipendenza tra le variabili stesse in zone in cui non siano state effettuate delle osservazioni.

In altre parole, non si può mai escludere sulla base delle misure che la $y = f(x)$ sia una funzione comunque complessa ma approssimativamente lineare *solamente* nell'intervallo in cui abbiamo investigato; per questo motivo bisogna evitare per quanto possibile di usare l'equazione della retta interpolante per ricavare valori stimati $\hat{y}$ della variabile indipendente (dalla $\hat{y} = a + bx$) in corrispondenza di valori della x non compresi nell'intorno delle misure, e questo tanto più rigorosamente quanto più x è distante da tale intorno: *non è lecito usare l'interpolazione per estrapolare* su regioni lontane da quelle investigate.

A questo proposito, se lo scopo primario dell'interpolazione non è tanto quello di ottenere una stima di a o b quanto quello di ricavare il valore $\hat{y}$ assunto dalla y in corrispondenza di un particolare valore $\hat{x}$ della variabile indipendente x, applicando la formula di propagazione degli errori (C.5) alla $\hat{y} = a + b\hat{x}$ ricaviamo:

$$\mathrm{Var}(\hat{y}) \ = \ \mathrm{Var}(a) + \hat{x}^2\,\mathrm{Var}(b) + 2\,\hat{x}\,\mathrm{Cov}(a,b) \ .$$

Sostituendo nell'equazione precedente le espressioni (11.11) per le varianze di a e b e quella (C.11) della loro covarianza, si ha poi

$$\mathrm{Var}(\hat{y}) \ = \ \frac{\sum_i {x_i}^2}{\Delta}\,{\sigma_y}^2 + \hat{x}^2\,\frac{N}{\Delta}\,{\sigma_y}^2 - 2\,\hat{x}\,\frac{\sum_i x_i}{\Delta}\,{\sigma_y}^2$$

e, introducendo nell'equazione precedente sia l'espressione (C.10) per Δ che la

$$\sum_i x_i^2 = N\left[\mathrm{Var}(x) + \bar{x}^2\right]$$

e la

$$\sum_i x_i = N\bar{x}$$

otteniamo

$$\begin{aligned}\mathrm{Var}(\hat{y}) &= \frac{\sigma_y^2}{\Delta}\left[\sum_i x_i^2 - 2\hat{x}\sum_i x_i + N\hat{x}^2\right] \\ &= \frac{\sigma_y^2}{N\,\mathrm{Var}(x)}\left[\mathrm{Var}(x) + \bar{x}^2 - 2\hat{x}\bar{x} + \hat{x}^2\right]\end{aligned}$$

ed infine la

$$\mathrm{Var}(\hat{y}) = \frac{\sigma_y^2}{N}\left[1 + \frac{(\hat{x}-\bar{x})^2}{\mathrm{Var}(x)}\right] \quad . \tag{C.12}$$

Si vede immediatamente dalla (C.12) che l'errore sul valore stimato $\hat{y}$ (che è funzione di $\hat{x}$) è minimo quando $\hat{x} = \bar{x}$: quindi, per ricavare una stima della y con il più piccolo errore casuale possibile, bisogna che il corrispondente valore della x sia nel centro dell'intervallo in cui si sono effettuate le misure (questo in accordo con le considerazioni qualitative precedenti a riguardo di interpolazione ed estrapolazione).

C.4.5 Verifica di ipotesi nell'interpolazione lineare

Nel paragrafo 11.4.1 abbiamo ricavato le formule (11.10) dei minimi quadrati e le formule (11.11) per l'errore dei coefficienti della retta interpolata: queste ultime richiedono la conoscenza dell'errore comune sulle ordinate σ_y che, di consueto, viene ricavato a posteriori dai dati attraverso l'equazione (C.8).

Assai di frequente è necessario verificare delle ipotesi statistiche sui risultati dell'interpolazione lineare; una volta ricavata, ad esempio, la pendenza della retta interpolante, si può o voler confrontare la stima ottenuta con un valore noto a priori, o voler costruire attorno ad essa un intervallo corrispondente ad un certo livello di confidenza; o, ancora, si può voler effettuare il confronto (o calcolare l'ampiezza dell'intervallo di confidenza) per un valore $\hat{y} = a + b\hat{x}$ della y stimato sulla base dell'interpolazione, e la cui varianza è data dalla (C.12).

È naturale pensare di sfruttare per questo scopo le tabelle della distribuzione normale; ma questo implicitamente richiede che il numero di coppie di dati a disposizione *sia sufficientemente elevato* perché la stima di σ_y ottenuta a posteriori dai dati si possa considerare esatta. In realtà quando l'errore è ricavato a posteriori *tutte* le grandezze precedentemente citate *non* seguono la distribuzione normale *ma la distribuzione di Student* con $(N-2)$ gradi di libertà.

C.4.6 Adeguatezza dell'interpolazione lineare o polinomiale in genere

Talvolta non si sa a priori che la relazione tra due variabili x ed y è di tipo lineare; ma, una volta eseguite le misure, si nota un loro approssimativo disporsi lungo una linea retta. In questo caso si può cercare di inferire una legge fisica dai dati sperimentali; ma come si può essere ragionevolmente sicuri che la relazione tra le due grandezze sia effettivamente lineare, anche limitandoci all'intervallo in cui sono distribuite le osservazioni?

Una possibilità è quella di eseguire interpolazioni successive con più curve polinomiali di grado M crescente, del tipo

$$y = P_M(x) = \sum_{k=0}^{M} a_k x^k \tag{C.13}$$

e di osservare l'andamento del *valore* dei residui in funzione di M: al crescere del grado del polinomio questi diminuiranno, dapprima in modo rapido per poi assestarsi su valori, sempre decrescenti, ma più o meno dello stesso ordine di grandezza; ed infine si annulleranno quando $M = N-1$. Il valore di M che segna la transizione tra questi due comportamenti ci dà il grado della curva polinomiale che descrive in modo soddisfacente la relazione tra le variabili senza per questo seguire le fluttuazioni casuali di ogni singola misura.

Nel caso in cui ognuno degli N valori y_i ha errore noto σ_i (e le y_i non sono correlate), la somma dei quadrati dei residui pesati in maniera inversamente proporzionale ai quadrati degli errori,

$$S_M = \sum_{i=1}^{N} \frac{{\delta_i}^2}{{\sigma_i}^2} = \sum_{i=1}^{N} \frac{(y_i - \hat{y}_i)^2}{{\sigma_i}^2}$$

(le $\hat{y}_i$ si intendono calcolate usando il polinomio interpolante $y = P_M(x)$ (C.13), i cui $M+1$ coefficienti siano stati stimati dai dati) è distribuita come il χ^2 a $N-M-1$ gradi di libertà; la probabilità di ottenere un determinato valore

di S_M sotto l'ipotesi $y = P_M(x)$ può dunque essere stimata dalle tabelle della distribuzione, e ci dà una misura statisticamente corretta dell'"accordo complessivo" tra i punti misurati ed un polinomio di grado M.

Per valutare numericamente la significatività della variazione di questo "accordo complessivo" dovuta ad un aumento di grado del polinomio interpolante, la regola di somma del χ^2 ci dice che la variazione della somma pesata dei quadrati dei residui, $S_{M-1} - S_M$, deve essere distribuita come il χ^2 ad un grado di libertà; normalmente, in luogo di usare direttamente $S_{M-1} - S_M$, si considera l'altra variabile casuale

$$F = \frac{S_{M-1} - S_M}{\dfrac{S_M}{N - M - 1}}$$

(che rappresenta una sorta di "variazione relativa dell'accordo complessivo"), e la si confronta con la funzione di frequenza di Fisher a 1 e $(N - M - 1)$ gradi di libertà; un valore di F elevato, e che con piccola probabilità potrebbe essere ottenuto da quella distribuzione, implicherà che l'aumento di grado è significativo.

C.4.7 Il *run test* per i residui

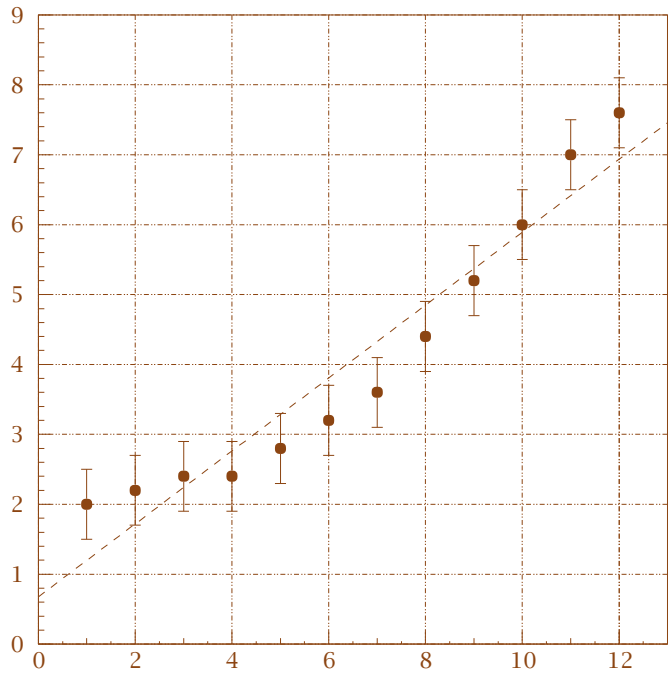
Un'altra tecnica che ci permette di capire se una funzione di primo grado è o meno adeguata a rappresentare un insieme di dati è quella che consiste nell'osservare l'andamento, in funzione della x, del solo segno dei residui δ_i differenza tra i valori misurati e quelli stimati della y:

$$\delta_i = y_i - \hat{y}_i = y_i - (a + b\,x_i) \ .$$

Per meglio chiarire questo concetto, osserviamo la figura C1 tratta dal paragrafo 8.3.2 del testo di Barlow citato nella bibliografia (appendice H, a pagina 323). È evidente come l'andamento dei dati sperimentali non suggerisca affatto l'ipotesi di una dipendenza lineare del tipo $y = A + Bx$; questo anche se l'entità degli errori assunti sulle y_i fa sì che l'accordo tra i dati stessi e la retta interpolante, se valutato con il calcolo del χ^2, risulti comunque accettabile: infatti il metodo citato usa come stima la somma dei quadrati dei *rapporti tra i residui e gli errori stimati*, ovviamente piccola se questi ultimi sono stati sopravvalutati.

Quello che è in grado di suggerire il sospetto di un andamento non lineare della legge $y = y(x)$, in casi come questo, è un altro tipo di controllo basato appunto *sul solo segno dei residui e non sul loro valore* (come il calcolo del χ^2, o dell'errore a posteriori, o del coefficiente di correlazione lineare, o

Figura C1 - Esempio di interpolazione lineare per un insieme di 12 punti.



dalla somma pesata dei quadrati dei residui). Segni che siano (come nell'esempio di figura C1) per piccole x tutti positivi, poi tutti negativi, ed infine tutti positivi per i valori più grandi delle x suggeriranno che si sta tentando di approssimare con una retta una funzione che in realtà è una curva più complessa (ad esempio una parabola) avente concavità rivolta verso l'alto.

Cominciamo con l'osservare che il valore medio $\bar{\delta}$ dei residui è identicamente nullo: infatti dalla (C.7) ricaviamo immediatamente

$$\bar{\delta} = \frac{1}{N} \sum_{i=1}^{N} \Big[(y_i - \bar{y}) - b\,(x_i - \bar{x}) \Big] \equiv 0 \ .$$

Questo è dovuto al fatto che sia la somma degli scarti $x_i - \bar{x}$ che quella degli scarti $y_i - \bar{y}$ sono identicamente nulle in conseguenza della (4.2), ed è vero quindi indipendentemente dal valore di b: questa proprietà vale insomma per residui calcolati rispetto a *qualunque* retta del fascio di centro $(\bar{x}, \bar{y})$ cui sappiamo che la retta interpolante *deve* appartenere.

Continuiamo osservando che i residui δ_i e le coordinate x_i hanno tra loro

covarianza nulla:

$$\begin{aligned}
\mathrm{Cov}(\delta, x) &= \frac{1}{N}\sum_{i=1}^{N}\left[\left(\delta_i - \bar{\delta}\right)(x_i - \bar{x})\right] \\
&= \frac{1}{N}\sum_{i=1}^{N}\left[(y_i - \bar{y})(x_i - \bar{x}) - b\,(x_i - \bar{x})^2\right] \\
&= \mathrm{Cov}(x, y) - b\,\mathrm{Var}(x) \\
&\equiv 0
\end{aligned}$$

(si è sfruttata, alla fine, la (C.6)). Questa condizione non è sufficiente, come si sa, ad assicurare l'indipendenza statistica tra i residui δ e le coordinate x dei punti interpolati; in effetti queste due variabili casuali *non* sono tra loro indipendenti, essendo ad esempio impossibile che i residui si presentino in una sequenza crescente all'aumentare delle ascisse x_i.

Però, quando il numero N dei dati è grande, delle $N!$ sequenze possibili di residui assai poche sono quelle escluse a priori dalla natura della loro origine; mentre la probabilità delle altre sequenze (che decresce in maniera inversamente proporzionale a $N!$) è comunque assai piccola: e si può in prima approssimazione assumere che *tutte* le sequenze di residui siano *possibili ed equiprobabili*.

Tornando alla figura C1, i residui sono (muovendosi nel senso delle x crescenti) dapprima positivi, poi negativi, poi ancora positivi; sono composti insomma da una sequenza di tre sottoinsiemi di valori aventi tutti lo stesso segno (o, con parola anglosassone, da una sequenza di tre *runs*). Se possiamo assumere equiprobabili tutte le possibili sequenze dei residui è intuitivo capire come un numero così basso di runs si debba presentare con piccola probabilità sulla base di fluttuazioni unicamente casuali; per cui l'osservazione di tale evento può essere attribuita invece alla falsità dell'ipotesi che ha prodotto i residui, ovvero alla non linearità della dipendenza funzionale $y = y(x)$.

Nell'ipotesi di avere un insieme composto da N_+ numeri positivi (corrispondenti, nel caso dei residui, a punti al di sopra della retta interpolante) e da $N_- = N - N_+$ numeri negativi (residui di punti al di sotto della retta interpolante), è possibile calcolare quante delle loro permutazioni producono un prefissato numero di runs N_r; se è accettabile l'approssimazione cui abbiamo appena accennato, il rapporto tra N_r ed il numero totale di permutazioni possibili ci darà la probabilità di ottenere un certo N_r.

I calcoli dettagliati si possono trovare nel citato paragrafo 8.3.2 del Bar-

low o, più in breve, nel paragrafo 11.3.1 di Eadie *et al.* (testo sempre citato nella bibliografia); qui ricordiamo solo come:

1. se $N_+ = 0$ o $N_- = 0$ (caso impossibile questo per i residui di un'interpolazione lineare) l'unico valore possibile è $N_r = 1$.

2. Se $N_+ > 0$ e $N_- > 0$, ed indicando con $m = \min(N_+, N_-)$ il più piccolo di questi due valori, il numero di runs N_r è compreso tra i seguenti estremi:

$$\begin{cases} 2 \le N_r \le 2m & \text{se } N_+ = N_- = m\,; \\ 2 \le N_r \le 2m+1 & \text{se } N_+ \ne N_-\,. \end{cases}$$

   Il massimo valore di N_r corrisponde, nel primo caso, ad un alternarsi di valori positivi e negativi; nel secondo, a singoli valori del segno che si è presentato con minore frequenza che separino sequenze di uno o più valori dell'altro segno.

3. Se N_r è pari, con $N_r = 2s$, la probabilità di avere N_r runs è data da

$$\Pr(N_r) = 2\,\frac{\binom{N_+ - 1}{s-1}\binom{N_- - 1}{s-1}}{\binom{N}{N_+}}\;; \tag{C.14}$$

4. se N_r è dispari, con $N_r = 2s - 1$ (ed ovviamente $s \ge 2$ essendo $N_+ > 0$ e $N_- > 0$), la probabilità di avere N_r runs è data da

$$\Pr(N_r) = \frac{\binom{N_+ - 1}{s-2}\binom{N_- - 1}{s-1} + \binom{N_+ - 1}{s-1}\binom{N_- - 1}{s-2}}{\binom{N}{N_+}}\;. \tag{C.15}$$

5. In ogni caso, valore medio e varianza di N_r valgono rispettivamente

$$E(N_r) = 1 + 2\,\frac{N_+ N_-}{N}$$

   e

$$\mathrm{Var}(N_r) = 2\,\frac{N_+ N_- \,(2N_+ N_- - N)}{N^2\,(N-1)}\;.$$

Nel caso della figura C1 ($N_+ = N_- = 6$) la probabilità di ottenere casualmente $N_r \le 3$, calcolata applicando direttamente le formule (C.14) e (C.15), vale appena l'1.3%; è insomma lecito (almeno ad un livello di confidenza del 98.7%) rigettare l'ipotesi di un andamento lineare della y in funzione di x.

C.5 Applicazioni alla stima di parametri

Se la densità di probabilità $f(x;\theta)$ di una variabile casuale x dipende da un parametro di valore vero ignoto θ^*, abbiamo visto nel capitolo 11 che una stima $\hat{\theta}$ di tale valore può essere ottenuta col metodo della massima verosimiglianza; e che è possibile ricavare dalla derivata seconda della funzione di verosimiglianza (che ne misura la concavità nel punto di massimo) l'errore di questa stima, attraverso l'equazione (11.5).

Il metodo si può ovviamente estendere alla stima contemporanea di più parametri (e lo abbiamo in effetti usato, ad esempio, per ricavare i due coefficienti della retta interpolante nel paragrafo 11.4.1): se x è la variabile misurata, di densità di probabilità $f(x;\theta_1,\theta_2,\ldots,\theta_M)$ dipendente da M parametri θ_k, le stime di massima verosimiglianza $\hat{\theta}_k$ si ricavano risolvendo il sistema ottenuto annullando contemporaneamente ognuna delle derivate parziali prime (ed esaminando poi ognuna delle eventuali soluzioni per controllare se corrisponde ad un massimo).

I punti enunciati nel paragrafo 11.2 continuano a rimanere validi anche nel caso multidimensionale: in particolare, ognuna delle stime $\hat{\theta}_k$ è asintoticamente normale; e si troverebbe che le derivate seconde della funzione di verosimiglianza nel punto di minimo sono legate all'inversa della matrice delle covarianze delle M stime attraverso la

$$\left(\boldsymbol{V}^{-1}\right)_{ij} = -N \cdot E\left\{\frac{\partial^2 \ln f(x;\theta_1,\theta_2,\ldots,\theta_M)}{\partial\theta_i\,\partial\theta_j}\right\} \tag{C.16}$$

che si può pensare come la generalizzazione a più stime contemporanee dell'equazione (11.5).

Come esempio, abbiamo già visto nel paragrafo 11.5 come stimare *contemporaneamente* i due parametri μ e σ di una popolazione normale da un campione di N determinazioni indipendenti col metodo della massima verosimiglianza; e vogliamo ora ricavare gli errori di quelle stime dalla (C.16).

Ricordiamo dal paragrafo 11.5 che il logaritmo della densità di probabilità vale

$$\ln f(x;\mu,\sigma) \;=\; -\ln\sigma - \ln\sqrt{2\pi} - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \quad ;$$

e che le due stime di massima verosimiglianza per μ e σ sono

$$\hat{\mu} = \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \text{e} \qquad \hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i-\hat{\mu})^2$$

rispettivamente. Le derivate prime di $\ln f$ sono

$$\frac{\partial}{\partial\mu}\ln f = \frac{x-\mu}{\sigma^2} \qquad\qquad \frac{\partial}{\partial\sigma}\ln f = -\frac{1}{\sigma} + \frac{(x-\mu)^2}{\sigma^3}$$

e le derivate seconde

$$\frac{\partial^2}{\partial\mu^2}\ln f = -\frac{1}{\sigma^2} \qquad \frac{\partial^2}{\partial\mu\,\partial\sigma}\ln f = -2\,\frac{x-\mu}{\sigma^3}$$

$$\frac{\partial^2}{\partial\sigma\,\partial\mu}\ln f = -2\,\frac{x-\mu}{\sigma^3} \qquad \frac{\partial^2}{\partial\sigma^2}\ln f = \frac{1}{\sigma^2} - 3\,\frac{(x-\mu)^2}{\sigma^4}$$

di valori medi

$$E\left(\frac{\partial^2}{\partial\mu^2}\ln f\right) = -\frac{1}{\sigma^2} \qquad E\left(\frac{\partial^2}{\partial\mu\,\partial\sigma}\ln f\right) = 0$$

$$E\left(\frac{\partial^2}{\partial\sigma\,\partial\mu}\ln f\right) = 0 \qquad E\left(\frac{\partial^2}{\partial\sigma^2}\ln f\right) = -\frac{2}{\sigma^2}\ ;$$

per cui, dalla (C.16), l'inverso della matrice delle covarianze (che è diagonale) è

$$\boldsymbol{V}^{-1} = \begin{Vmatrix} \dfrac{N}{\hat\sigma^2} & 0 \\ 0 & \dfrac{2N}{\hat\sigma^2} \end{Vmatrix}$$

e la matrice $\boldsymbol{V}$ stessa vale

$$\boldsymbol{V} = \begin{Vmatrix} \dfrac{\hat\sigma^2}{N} & 0 \\ 0 & \dfrac{\hat\sigma^2}{2N} \end{Vmatrix}\ .$$

Insomma, oltre alla consueta espressione della varianza della media

$$\mathrm{Var}(\hat\mu) = \frac{\hat\sigma^2}{N}$$

abbiamo ottenuto quella della varianza di $\hat\sigma$

$$\mathrm{Var}(\hat\sigma) = \frac{\hat\sigma^2}{2N}$$

da confrontare con la (B.1), in cui però $\hat\sigma$ era già stato corretto, moltiplicandolo per un fattore $N/(N-1)$, per eliminare la distorsione della stima; e la riconferma del fatto, già visto nel paragrafo 12.1 a pagina 203, che valore medio e varianza di un campione di stime indipendenti sono variabili casuali statisticamente indipendenti tra loro (le covarianze infatti sono nulle).

Appendice D

Il modello di Laplace e la funzione di Gauss

Pensiamo di eseguire una misura di una grandezza fisica (il cui valore vero indicheremo con il simbolo x^*), e sia x il risultato ottenuto; in generale x è diverso da x^* per la presenza degli errori di misura, che supporremo siano di natura puramente casuale.

Questi errori casuali di misura possono essere schematizzati come un insieme estremamente grande, al limite infinito, di disturbi contemporanei molto piccoli, al limite infinitesimi, ognuno dei quali tende ad alterare di pochissimo il risultato della misura; si considerino in particolare le seguenti ipotesi (*modello semplificato di Laplace*[1] *per gli errori di misura*):

1. *Ognuna delle singole cause di disturbo presenti introdurrà nella misura una variazione rispetto al valore vero di modulo fisso ϵ, con uguale probabilità in difetto o in eccesso.*

2. *Ognuna delle variazioni nella misura dovute a queste cause di disturbo è statisticamente indipendente dalle altre.*

[1] Pierre Simon de Laplace visse in Francia dal 1749 al 1827; famoso matematico, fisico ed astronomo, provò la stabilità del sistema solare, sviluppò la teoria delle equazioni differenziali e dei potenziali, contribuì allo studio del calore e dei fenomeni capillari oltre a gettare le basi matematiche per una teoria dell'elettromagnetismo. Durante la rivoluzione francese fu uno degli ideatori del sistema metrico decimale; per quel che riguarda la statistica, nel 1812 pubblicò il trattato "Théorie Analytique des Probabilités" che contiene, tra l'altro, studi sulla distribuzione normale e la derivazione della regola dei minimi quadrati.

Ognuna delle N cause indipendenti di disturbo produce quindi la variazione $+\epsilon$ con probabilità $p = 0.5$ oppure $-\epsilon$ con probabilità $q = 1 - p = 0.5$; se M tra le N perturbazioni sono positive (e le altre $N - M$ negative), il valore osservato sarà

$$x \;=\; x^* + M\epsilon - (N - M)\epsilon \;=\; x^* + (2M - N)\epsilon \;\; .$$

La probabilità di un dato valore di M sulle N prove è data dalla distribuzione binomiale (vedi il paragrafo 8.4, ed in particolare l'equazione (8.7)), e vale

$$P(M, N) \;=\; \frac{N!}{M!\,(N - M)!}\, p^M q^{N-M} \;\; .$$

Il valore medio di M è dato da Np, e la sua varianza da Npq; indichiamo poi con il simbolo λ lo scarto di M dal suo valore medio

$$M \;=\; Np + \lambda \;\; .$$

In corrispondenza al variare di M tra 0 ed N, λ varia tra i limiti $-Np$ e $+Nq$; risulta poi anche

$$N - M \;=\; N - Np - \lambda \;=\; Nq - \lambda$$

e la probabilità di ottenere un certo valore di λ su N prove vale

$$P(\lambda, N) \;=\; \frac{N!}{(Np + \lambda)!\,(Nq - \lambda)!}\, p^{Np+\lambda}\, q^{Nq-\lambda} \;\; .$$

Valore medio e varianza di λ valgono poi

$$E(\lambda) \;=\; E(M) - Np \;\equiv\; 0$$

e

$$\mathrm{Var}(\lambda) \;=\; \mathrm{Var}(M) \;=\; Npq \;\; .$$

L'andamento generale della probabilità in funzione di M si può trovare considerando il rapporto tra i valori di P che corrispondono a due valori successivi di M:

$$\begin{aligned} \frac{P(M+1, N)}{P(M, N)} &= \frac{N!\, p^{M+1}\, q^{N-M-1}}{(M+1)!\,(N-M-1)!}\, \frac{M!\,(N-M)!}{N!\, p^M\, q^{N-M}} \\ &= \frac{N-M}{M+1}\, \frac{p}{q} \end{aligned}$$

e $P(M,N)$ risulterà minore, uguale o maggiore di $P(M+1,N)$ a seconda che $(M+1)q$ risulti minore, uguale o maggiore di $(N-M)p$; ossia, essendo $p+q=1$, a seconda che M sia minore, uguale o maggiore di $Np-q$.

Insomma, chiamato $\mu = \lceil Np-q \rceil$ il più piccolo intero non minore di $Np-q$, la sequenza di valori $P(0,N), P(1,N), \ldots, P(\mu,N)$ è crescente, mentre quella dei valori $P(\mu+1,N), P(\mu+2,N), \ldots, P(N,N)$ è decrescente. Il massimo valore della probabilità si ha in corrispondenza ad un intero μ che soddisfi la

$$Np-q \;\le\; \mu \;\le\; Np-q+1 \;=\; Np+p$$

e che è unico, salvo il caso che i due estremi dell'intervallo siano entrambi numeri interi: in questo caso si hanno due valori massimi, uguali, in corrispondenza di entrambi. Concludendo: il caso più probabile è che l'evento E si presenti in una sequenza di N prove Np volte, ed il valore di λ con la massima probabilità di presentarsi è 0.

Cerchiamo ora di determinare se esiste e quanto vale il limite della probabilità di ottenere un certo risultato al crescere indefinito del numero delle prove. Per ottenere questo, introduciamo la *formula approssimata di de Moivre e Stirling*[2] per il fattoriale:

$$N! \;=\; N^N e^{-N}\sqrt{2\pi N}\,(1+\epsilon_N) \;\approx\; \sqrt{2\pi}\,N^{\left(N+\frac{1}{2}\right)}e^{-N}$$

con

$$0 \;\le\; \epsilon_N \;<\; \frac{1}{11\cdot N}\;.$$

È lecito trascurare il resto ϵ_N quando l'argomento del fattoriale è elevato: per $N=10$ l'errore commesso è già inferiore all'1%. Per usare la formula di de Moivre e Stirling nel nostro caso, sviluppiamo

$$\begin{aligned}(Np+\lambda)! &\approx \sqrt{2\pi}\,(Np+\lambda)^{\left(Np+\lambda+\frac{1}{2}\right)}e^{(-Np-\lambda)}\\ &= \sqrt{2\pi}\left(1+\frac{\lambda}{Np}\right)^{\left(Np+\lambda+\frac{1}{2}\right)} e^{(-Np-\lambda)}(Np)^{\left(Np+\lambda+\frac{1}{2}\right)}\end{aligned}$$

e, similmente,

$$(Nq-\lambda)! \approx \sqrt{2\pi}\left(1-\frac{\lambda}{Nq}\right)^{\left(Nq-\lambda+\frac{1}{2}\right)} e^{(-Nq+\lambda)}(Nq)^{\left(Nq-\lambda+\frac{1}{2}\right)}\;.$$

[2]Per la dimostrazione, vedi ad esempio: G. Castelnuovo – Calcolo delle probabilità (Zanichelli), in appendice. La formula è dovuta al solo Stirling, che la pubblicò nel suo libro "Methodus Differentialis" del 1730; ma non divenne nota nel mondo scientifico fino a quando de Moivre non la usò — da qui il nome comunemente adottato.

Queste approssimazioni sono valide quando gli argomenti dei fattoriali, $Np + \lambda$ e $Nq - \lambda$, sono abbastanza grandi: cioè quando λ non è vicino ai valori limite $-Np$ e Nq; accettata la loro validità (e ritorneremo su questo punto tra poco), sostituendo si ha

$$P(\lambda, N) = \frac{1}{\sqrt{2\pi Npq}} \left(1 + \frac{\lambda}{Np}\right)^{-\left(Np+\lambda+\frac{1}{2}\right)} \left(1 - \frac{\lambda}{Nq}\right)^{-\left(Nq-\lambda+\frac{1}{2}\right)} .$$

Questa espressione è certamente valida quando $|\lambda|$ non è troppo grande, e per $\lambda = 0$ fornisce la probabilità del valore medio di M ($M = Np$), che risulta

$$P(0, N) = \frac{1}{\sqrt{2\pi Npq}} .$$

Questa probabilità tende a zero come $1/\sqrt{N}$ al crescere di N; dato che la somma delle probabilità relative a tutti i casi possibili deve essere 1, si deve concludere che il numero di valori di λ per cui la probabilità non è trascurabile rispetto al suo massimo deve divergere come $\sqrt{N}$ al crescere di N, sebbene il numero di tutti i possibili valori (che è $N + 1$) diverga invece come N.

L'espressione approssimata di $P(\lambda, N)$ non è valida per valori di λ prossimi agli estremi $\lambda = -Np$ e $\lambda = Nq$ (è infatti divergente); tuttavia tali valori hanno probabilità infinitesime di presentarsi al crescere di N. Infatti $P(-Np, N) = q^N$ e $P(Nq, N) = p^N$, ed entrambi tendono a zero quando N tende all'infinito essendo sia p che q inferiori all'unità.

Concludendo: la formula approssimata da noi ricavata è valida già per valori relativamente piccoli di N, e per N molto grande si può ritenere esatta per tutti i valori dello scarto λ con probabilità non trascurabile di presentarsi, valori che sono mediamente dell'ordine dell'errore quadratico medio $\sqrt{Npq}$ e che quindi divergono solo come $\sqrt{N}$. Consideriamo ora il fattore

$$\kappa = \left(1 + \frac{\lambda}{Np}\right)^{-\left(Np+\lambda+\frac{1}{2}\right)} \left(1 - \frac{\lambda}{Nq}\right)^{-\left(Nq-\lambda+\frac{1}{2}\right)}$$

che nell'espressione approssimata di $P(\lambda, N)$ moltiplica il valore massimo $P(0, N)$, e se ne prenda il logaritmo naturale:

$$\ln \kappa = -\left(Np + \lambda + \frac{1}{2}\right) \ln\left(1 + \frac{\lambda}{Np}\right) - \left(Nq - \lambda + \frac{1}{2}\right) \ln\left(1 - \frac{\lambda}{Nq}\right) .$$

Ora, poiché sia λ/Np che λ/Nq sono in modulo minori dell'unità (salvi i due casi estremi, di probabilità come sappiamo infinitesima), si possono

sviluppare i due logaritmi in serie di McLaurin:

$$\ln(1+x) \;=\; x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots \quad .$$

Il primo termine di $\ln\kappa$ diventa

$$\begin{aligned}
-\left(Np+\lambda+\frac{1}{2}\right)\left(\frac{\lambda}{Np} - \frac{\lambda^2}{2\,N^2p^2} + \frac{\lambda^3}{3\,N^3p^3} - \cdots\right) \;&=\\
= -\lambda + \left(\frac{\lambda^2}{2\,Np} - \frac{\lambda^2}{Np}\right) - \left(\frac{\lambda^3}{3\,N^2p^2} - \frac{\lambda^3}{2\,N^2p^2} + \frac{\lambda}{2\,Np}\right) + \cdots\\
= -\,\lambda - \frac{\lambda^2}{2\,Np} - \frac{\lambda}{2\,Np} + \frac{\lambda^3}{6\,N^2p^2} + \cdots
\end{aligned}$$

ed il secondo

$$\begin{aligned}
-\left(Nq-\lambda+\frac{1}{2}\right)\left(-\frac{\lambda}{Nq} - \frac{\lambda^2}{2\,N^2q^2} - \frac{\lambda^3}{3\,N^3q^3} - \cdots\right) \;&=\\
= \lambda - \frac{\lambda^2}{2\,Nq} + \frac{\lambda}{2\,Nq} - \frac{\lambda^3}{6\,N^2q^2} - \cdots
\end{aligned}$$

e sommando si ottiene

$$\ln\kappa = -\frac{\lambda^2}{2\,Npq} - \frac{\lambda}{2\,N}\left(\frac{1}{p} - \frac{1}{q}\right) + \frac{\lambda^3}{6\,N^2}\left(\frac{1}{p^2} - \frac{1}{q^2}\right) + \cdots \quad .$$

Da questo sviluppo risulta che il solo termine che si mantiene finito al divergere di N, e per valori di λ dell'ordine di $\sqrt{Npq}$, è il primo; gli altri due scritti convergono a zero come $1/\sqrt{N}$, e tutti gli altri omessi almeno come $1/N$. In conclusione, per valori dello scarto per cui la probabilità non è trascurabile (grosso modo $|\lambda| < 3\sqrt{Npq}$), al divergere di N il logaritmo di κ è bene approssimato da

$$\ln\kappa \;\approx\; -\frac{\lambda^2}{2\,Npq}$$

e la probabilità dello scarto dalla media λ da

$$P(\lambda) \;\approx\; \frac{1}{\sqrt{2\pi Npq}}\, e^{-\frac{1}{2}\frac{\lambda^2}{Npq}} \quad ;$$

per la variabile M sarà invece

$$P(M) \;\approx\; \frac{1}{\sqrt{2\pi Npq}}\, e^{-\frac{1}{2}\frac{(M-Np)^2}{Npq}} \quad .$$

Nel caso particolare del modello semplificato di Laplace per gli errori di misura, $p = q = 0.5$ e pertanto i termini di ordine $1/\sqrt{N}$ sono identicamente nulli: l'approssimazione è già buona per $N \geq 25$; nel caso generale $p \neq q$, essa è invece accettabile per $Npq \geq 9$. Introducendo lo scarto quadratico medio di M e di λ

$$\sigma = \sqrt{Npq}$$

l'espressione si può scrivere

$$P(\lambda) \approx \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\lambda^2}{2\sigma^2}}$$

che è la celebre *legge normale* o *legge di Gauss*.

Tornando ancora al modello semplificato di Laplace per gli errori di misura, il risultato x ha uno scarto dal valore vero che vale

$$x - x^* = \epsilon\,(2M - N) = \epsilon\,(2Np + 2\lambda - N) = 2\epsilon\lambda$$

e possiede varianza

$$\sigma_x{}^2 \equiv \mathrm{Var}\,(x - x^*) = 4\,\epsilon^2\,\mathrm{Var}(\lambda) = 4\epsilon^2\sigma^2 \ .$$

La probabilità di un certo risultato $x = x^* + 2\epsilon\lambda$ vale infine

$$P(x) = P(\lambda) \approx \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{\lambda^2}{\sigma^2}} = \frac{2\epsilon}{\sigma_x\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-x^*}{\sigma_x}\right)^2} \ .$$

La x è una grandezza discreta che varia per multipli di ϵ; nel limite su accennato diventa una variabile continua, e $P(x)$ è infinitesima con ϵ perdendo così significato; si mantiene invece finita la densità di probabilità, che si ottiene dividendo $P(x)$ per l'ampiezza 2ϵ dell'intervallo che separa due valori contigui di x:

$$f(x) = \frac{P(x)}{2\epsilon} = \frac{1}{\sigma_x\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-x^*}{\sigma_x}\right)^2}$$

ed ha infatti le dimensioni fisiche di $1/\sigma_x$, ovvero di $1/x$.

Al medesimo risultato per $f(x)$ si perverrebbe anche nell'ipotesi più generale che gli errori elementari siano distribuiti comunque, ed anche diversamente l'uno dall'altro, purché ciascuno abbia una varianza dello stesso ordine di grandezza degli altri ed infinitesima al divergere del numero delle cause di errore.

Appendice E

La funzione di verosimiglianza

Si supponga di aver compiuto N osservazioni indipendenti relative ad una grandezza fisica x, e di aver trovato i valori x_i, con $i = 1, 2, \ldots, N$. Ciascuna delle variabili casuali x_i abbia poi densità di probabilità data da una funzione nota $f_i(x_i; \theta)$; funzione che supponiamo dipenda da un parametro θ di valore vero θ^* ignoto, e definita in un intervallo dell'asse reale delle x_i con estremi indipendenti da θ (che potremo assumere essere $\pm\infty$ ponendo eventualmente $f_i(x_i; \theta) \equiv 0$ esternamente all'intervallo di definizione).

Una *stima* di una generica funzione nota del parametro, $\tau(\theta)$, che supporremo con derivata non nulla, è una funzione dei soli valori osservati $t(x_1, x_2, \ldots, x_N)$; dunque a sua volta una variabile casuale, con associata una funzione densità di probabilità che indicheremo con $g(t; \theta)$. La stima si dice *imparziale* (o *indistorta*) quando il suo valore medio

$$\begin{aligned} E(t) &= \int_{-\infty}^{+\infty} t\, g(t; \theta)\, \mathrm{d}t \\ &= \int_{-\infty}^{+\infty} \mathrm{d}x_1\, f_1(x_1; \theta) \cdots \int_{-\infty}^{+\infty} \mathrm{d}x_N\, f_N(x_N; \theta)\, t(x_1, x_2, \ldots, x_N) \end{aligned}$$

è uguale al rispettivo valore vero:

$$E(t) = \tau(\theta) \ .$$

Il caso particolare della stima *del parametro stesso* corrisponde alla funzione $\tau(\theta) = \theta$, che soddisfa evidentemente alla richiesta di possedere derivata prima non nulla $\tau'(\theta) = 1$.

Una importante proprietà della stima t è la sua varianza, data (se essa è imparziale) da

$$\begin{aligned}\sigma_t{}^2 &= \int_{-\infty}^{+\infty} [t - \tau(\theta)]^2 g(t;\theta)\,\mathrm{d}t \\ &= \int_{-\infty}^{+\infty} \mathrm{d}x_1\, f_1(x_1;\theta) \cdots \int_{-\infty}^{+\infty} \mathrm{d}x_N\, f_N(x_N;\theta)\, [t(x_1, x_2, \ldots, x_N) - \tau(\theta)]^2\end{aligned}$$

perché *la minima varianza* sarà il nostro criterio di scelta fra diverse stime di $\tau(\theta)$.

Il teorema che segue (**teorema di Cramér–Rao**) mostra che esiste un limite inferiore per la varianza di una stima. Osserviamo per prima cosa che la densità di probabilità per la N-pla $(x_1, x_2, \ldots, x_N)$ risulta

$$\prod_{i=1}^{N} f_i(x_i;\theta^*)$$

per il teorema della probabilità composta; se in luogo del valore vero θ^* si pone il parametro variabile θ, si ottiene la *funzione di verosimiglianza*

$$\mathcal{L}(x_1, x_2, \ldots, x_N;\theta) = \prod_{i=1}^{N} f_i(x_i;\theta) \quad .$$

La condizione di normalizzazione di ciascuna f_i comporta che l'integrale della verosimiglianza su tutti i domini delle variabili x_i valga 1:

$$\begin{aligned}\int_{-\infty}^{+\infty} \mathrm{d}x_1 \int_{-\infty}^{+\infty} \mathrm{d}x_2 \cdots \int_{-\infty}^{+\infty} \mathrm{d}x_N\, \mathcal{L}(x_1, x_2, \ldots, x_N;\theta) &= \\ &= \int_{-\infty}^{+\infty} \mathrm{d}x_1\, f_1(x_1;\theta) \int_{-\infty}^{+\infty} \mathrm{d}x_2\, f_2(x_2;\theta) \cdots \int_{-\infty}^{+\infty} \mathrm{d}x_N\, f_N(x_N;\theta) \\ &= \prod_{i=1}^{N} \int_{-\infty}^{+\infty} \mathrm{d}x_i\, f_i(x_i;\theta) \\ &\equiv 1\end{aligned}$$

indipendentemente dal valore di θ. Derivando sotto il segno di integrale rispetto a θ, dato che i domini delle $f_i(x_i;\theta)$ non dipendono da detta variabile si ottiene

$$\int_{-\infty}^{+\infty} \mathrm{d}x_1 \int_{-\infty}^{+\infty} \mathrm{d}x_2 \cdots \int_{-\infty}^{+\infty} \mathrm{d}x_N\, \frac{\partial \mathcal{L}}{\partial \theta} = 0$$

da cui, dividendo e moltiplicando l'integrando per $\mathcal{L}$, risulta

$$\begin{aligned}
\int_{-\infty}^{+\infty} \mathrm{d}x_1 \int_{-\infty}^{+\infty} \mathrm{d}x_2 \cdots \int_{-\infty}^{+\infty} \mathrm{d}x_N \, \mathcal{L} \left(\frac{1}{\mathcal{L}} \frac{\partial \mathcal{L}}{\partial \theta} \right) &= \\
&= \int_{-\infty}^{+\infty} \mathrm{d}x_1 \int_{-\infty}^{+\infty} \mathrm{d}x_2 \cdots \int_{-\infty}^{+\infty} \mathrm{d}x_N \, \mathcal{L} \, \frac{\partial \left(\ln \mathcal{L} \right)}{\partial \theta} \\
&= \int_{-\infty}^{+\infty} \mathrm{d}x_1 \, f_1(x_1; \theta) \cdots \int_{-\infty}^{+\infty} \mathrm{d}x_N \, f_N(x_N; \theta) \, \frac{\partial \left(\ln \mathcal{L} \right)}{\partial \theta} \\
&= 0
\end{aligned}$$

ossia

$$\boxed{E \left\{ \frac{\partial \left(\ln \mathcal{L} \right)}{\partial \theta} \right\} = 0} \tag{E.1}$$

Se t è imparziale

$$E(t) = \int_{-\infty}^{+\infty} \mathrm{d}x_1 \cdots \int_{-\infty}^{+\infty} \mathrm{d}x_N \, t(x_1, x_2, \ldots, x_N) \, \mathcal{L}(x_1, x_2, \ldots, x_N; \theta) = \tau(\theta)$$

da cui, derivando ambo i membri rispetto a θ,

$$\int_{-\infty}^{+\infty} \mathrm{d}x_1 \int_{-\infty}^{+\infty} \mathrm{d}x_2 \cdots \int_{-\infty}^{+\infty} \mathrm{d}x_N \, t \, \frac{\partial \mathcal{L}}{\partial \theta} = \tau'(\theta) \ .$$

Dividendo e moltiplicando poi l'integrando per la verosimiglianza $\mathcal{L}$, risulta

$$\begin{aligned}
\int_{-\infty}^{+\infty} \mathrm{d}x_1 \int_{-\infty}^{+\infty} \mathrm{d}x_2 \cdots \int_{-\infty}^{+\infty} \mathrm{d}x_N \, t \, \frac{\partial \mathcal{L}}{\partial \theta} &= \\
&= \int_{-\infty}^{+\infty} \mathrm{d}x_1 \cdots \int_{-\infty}^{+\infty} \mathrm{d}x_N \, t \, \mathcal{L} \left(\frac{1}{\mathcal{L}} \frac{\partial \mathcal{L}}{\partial \theta} \right) \\
&= \int_{-\infty}^{+\infty} \mathrm{d}x_1 \, f_1(x_1; \theta) \cdots \int_{-\infty}^{+\infty} \mathrm{d}x_N \, f_N(x_N; \theta) \, t \, \frac{\partial \left(\ln \mathcal{L} \right)}{\partial \theta} \\
&= E \left\{ t \, \frac{\partial \left(\ln \mathcal{L} \right)}{\partial \theta} \right\}
\end{aligned}$$

e, in definitiva,

$$\boxed{E \left\{ t \, \frac{\partial \left(\ln \mathcal{L} \right)}{\partial \theta} \right\} = \tau'(\theta)}$$

Infine, sottraendo membro a membro da questa equazione la precedente (E.1) moltiplicata per $\tau(\theta)$, si ottiene

$$E\left\{t\,\frac{\partial\,(\ln\mathcal{L})}{\partial\theta}\right\} - \tau(\theta)\cdot E\left\{\frac{\partial\,(\ln\mathcal{L})}{\partial\theta}\right\} = \tau'(\theta)$$

ovvero

$$E\left\{[t-\tau(\theta)]\cdot\frac{\partial\,(\ln\mathcal{L})}{\partial\theta}\right\} = \tau'(\theta)\ .$$

Se ora si definiscono il rapporto

$$R(\theta) \;=\; \frac{E\left\{[t-\tau(\theta)]\cdot\dfrac{\partial\,(\ln\mathcal{L})}{\partial\theta}\right\}}{E\left\{\left[\dfrac{\partial\,(\ln\mathcal{L})}{\partial\theta}\right]^2\right\}} \;=\; \frac{\tau'(\theta)}{E\left\{\left[\dfrac{\partial\,(\ln\mathcal{L})}{\partial\theta}\right]^2\right\}}$$

(che è una costante dipendente da θ; osserviamo anche che deve risultare $R(\theta)\neq 0$) e la variabile casuale

$$z = [t-\tau(\theta)] - R(\theta)\,\frac{\partial\,(\ln\mathcal{L})}{\partial\theta}$$

il cui quadrato risulta essere

$$z^2 = [t-\tau(\theta)]^2 - 2\,R(\theta)\cdot[t-\tau(\theta)]\,\frac{\partial\,(\ln\mathcal{L})}{\partial\theta} + R^2(\theta)\cdot\left[\frac{\partial\,(\ln\mathcal{L})}{\partial\theta}\right]^2$$

prendendo il valore medio di z^2 si ottiene

$$\begin{aligned} E(z^2) \;=\; & E\left\{[t-\tau(\theta)]^2\right\} - 2\,R(\theta)\cdot E\left\{[t-\tau(\theta)]\cdot\frac{\partial\,(\ln\mathcal{L})}{\partial\theta}\right\} + \\ & + R^2(\theta)\cdot E\left\{\left[\frac{\partial\,(\ln\mathcal{L})}{\partial\theta}\right]^2\right\} \end{aligned}$$

ossia

$$\begin{aligned} E(z^2) \;=\; & {\sigma_t}^2 - 2\,\frac{\tau'(\theta)}{E\left\{\left[\dfrac{\partial\,(\ln\mathcal{L})}{\partial\theta}\right]^2\right\}}\,\tau'(\theta) + \\ & + \left\{\frac{\tau'(\theta)}{E\left\{\left[\dfrac{\partial\,(\ln\mathcal{L})}{\partial\theta}\right]^2\right\}}\right\}^2 E\left\{\left[\frac{\partial\,(\ln\mathcal{L})}{\partial\theta}\right]^2\right\} \end{aligned}$$

ed infine

$$
\begin{aligned}
E(z^2) &= {\sigma_t}^2 - 2\,\frac{[\tau'(\theta)]^2}{E\left\{\left[\dfrac{\partial\,(\ln\mathcal{L})}{\partial\theta}\right]^2\right\}} + \frac{[\tau'(\theta)]^2}{E\left\{\left[\dfrac{\partial\,(\ln\mathcal{L})}{\partial\theta}\right]^2\right\}} \\
&= {\sigma_t}^2 - \frac{[\tau'(\theta)]^2}{E\left\{\left[\dfrac{\partial\,(\ln\mathcal{L})}{\partial\theta}\right]^2\right\}} \;.
\end{aligned}
$$

Ma il valore medio del quadrato di una qualsiasi variabile casuale non può essere negativo, e dunque

$$
0 \;\le\; E(z^2) \;=\; {\sigma_t}^2 - \frac{[\tau'(\theta)]^2}{E\left\{\left[\dfrac{\partial\,(\ln\mathcal{L})}{\partial\theta}\right]^2\right\}}
$$

ed infine

$$
{\sigma_t}^2 \;\ge\; \frac{[\tau'(\theta)]^2}{E\left\{\left[\dfrac{\partial\,(\ln\mathcal{L})}{\partial\theta}\right]^2\right\}} \;=\; [\tau'(\theta)]^2\,\frac{R(\theta)}{\tau'(\theta)} \;=\; \tau'(\theta)\cdot R(\theta)
$$

cioè:

> *Nessuna funzione dei valori osservati $t(x_1, x_2, \ldots, x_N)$, che sia stima imparziale di una funzione del parametro $\tau(\theta)$, può avere varianza inferiore ad un limite determinato.*

La varianza minima si raggiunge se e soltanto se $E(z^2)$ è nullo, il che è possibile solo se z è nulla ovunque, cioè se

$$
z \;=\; t - \tau(\theta) - R(\theta)\,\frac{\partial\,(\ln\mathcal{L})}{\partial\theta} \;\equiv\; 0
$$

o, altrimenti detto, se la derivata logaritmica della verosimiglianza è proporzionale alla variabile casuale $t - \tau(\theta)$:

$$
\boxed{\frac{\partial\,(\ln\mathcal{L})}{\partial\theta} \;=\; \frac{t - \tau(\theta)}{R(\theta)}} \tag{E.2}
$$

Nel caso particolare che tutte le x_i provengano dalla stessa popolazione, e che quindi abbiano la stessa densità di probabilità $f(x;\theta)$,

$$\frac{\partial(\ln \mathcal{L})}{\partial\theta} = \frac{\partial}{\partial\theta}\sum_{i=1}^{N}\ln f(x_i;\theta) = \sum_{i=1}^{N}\frac{\partial}{\partial\theta}\ln f(x_i;\theta)$$

$$E\left\{\frac{\partial(\ln \mathcal{L})}{\partial\theta}\right\} = \sum_{i=1}^{N}E\left\{\frac{\partial}{\partial\theta}\ln f(x_i;\theta)\right\} = N\cdot E\left\{\frac{\partial}{\partial\theta}\ln f(x;\theta)\right\}$$

e, tenuto conto della (E.1), questo implica che

$$E\left\{\frac{\partial}{\partial\theta}\ln f(x;\theta)\right\} = 0 \ . \tag{E.3}$$

Ora

$$E\left\{\left[\frac{\partial(\ln \mathcal{L})}{\partial\theta}\right]^2\right\} = E\left\{\left[\sum_{i=1}^{N}\frac{\partial}{\partial\theta}\ln f(x_i;\theta)\right]\left[\sum_{k=1}^{N}\frac{\partial}{\partial\theta}\ln f(x_k;\theta)\right]\right\}$$

$$= \sum_{i=1}^{N}E\left\{\left[\frac{\partial}{\partial\theta}\ln f(x_i;\theta)\right]^2\right\} + \sum_{\substack{i,k\\ i\neq k}}E\left\{\frac{\partial}{\partial\theta}\ln f(x_i;\theta)\cdot\frac{\partial}{\partial\theta}\ln f(x_k;\theta)\right\}$$

$$= N\cdot E\left\{\left[\frac{\partial}{\partial\theta}\ln f(x;\theta)\right]^2\right\} + \sum_{\substack{i,k\\ i\neq k}}E\left\{\frac{\partial}{\partial\theta}\ln f(x_i;\theta)\right\}\cdot E\left\{\frac{\partial}{\partial\theta}\ln f(x_k;\theta)\right\}$$

(tenendo conto del fatto che le x_i sono indipendenti); l'ultimo termine si annulla in conseguenza della (E.3), ed infine, in questo caso, il minorante della varianza della stima si può scrivere

$$\boxed{\sigma_t{}^2 \geq \frac{[\tau'(\theta)]^2}{N\cdot E\left\{\left[\frac{\partial}{\partial\theta}\ln f(x;\theta)\right]^2\right\}}}$$

Col metodo della massima verosimiglianza si assume, come stima del valore vero θ^* del parametro θ, quel valore $\hat{\theta}$ che rende massima la verosimiglianza $\mathcal{L}$ per i valori osservati delle variabili, $x_1, x_2, \ldots, x_N$.

Ora, *nel caso esista una stima di minima varianza* t per la funzione $\tau(\theta)$, tenendo conto della (E.2) la condizione perché la funzione di verosimiglianza abbia un estremante diviene

$$\frac{\partial\,(\ln \mathcal{L})}{\partial\theta} = \frac{t-\tau(\theta)}{R(\theta)} = 0$$

e le soluzioni $\hat{\theta}$ sono tutte e sole quelle dell'equazione

$$\tau(\theta) = t(x_1, x_2, \ldots, x_N) \ .$$

La derivata seconda di $\ln \mathcal{L}$ è in tal caso

$$\begin{aligned} \frac{\partial^2 (\ln \mathcal{L})}{\partial \theta^2} &= -\frac{\tau'(\theta) \cdot R(\theta) + R'(\theta) \cdot [t - \tau(\theta)]}{R^2(\theta)} \\ &= -\frac{{\sigma_t}^2 + R'(\theta) \cdot [t - \tau(\theta)]}{R^2(\theta)} \end{aligned}$$

ma se $\theta = \hat{\theta}$ è anche $t - \tau(\hat{\theta}) = 0$ e risulta

$$\left[\frac{\partial^2 (\ln \mathcal{L})}{\partial \theta^2}\right]_{\theta=\hat{\theta}} = -\frac{{\sigma_t}^2}{R^2(\hat{\theta})} < 0 \ ;$$

cioè *per tutte le soluzioni $\theta = \hat{\theta}$ la verosimiglianza è massima.*

Ora, se la funzione $\ln \mathcal{L}$ è regolare, tra due massimi deve esistere un minimo; dato che non esistono minimi, ne consegue che *il massimo è unico* ed in corrispondenza al valore della funzione τ^{-1} inversa di $\tau(\theta)$ e calcolata in $t(x_1, x_2, \ldots, x_N)$:

$$\hat{\theta} = \tau^{-1}[t(x_1, x_2, \ldots, x_N)] \ .$$

La *statistica* $t(x_1, x_2, \ldots, x_N)$ (come viene anche indicata una funzione dei dati) di minima varianza è un caso particolare di statistica *sufficiente* per il parametro θ, come è chiamata una funzione dei valori osservati, se esiste, che riassume in sé tutta l'informazione che i dati possono fornire sul valore del parametro.

Se $x_1, x_2, \ldots, x_N$ sono i valori osservati di N variabili casuali *normali* con lo stesso valore medio λ e varianze rispettive σ_i supposte note, la verosimiglianza è

$$\mathcal{L} = \prod_{i=1}^{N} \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_i - \lambda}{\sigma_i}\right)^2}$$

il suo logaritmo

$$\ln \mathcal{L} = -\frac{1}{2} \sum_{i=1}^{N} \frac{(x_i - \lambda)^2}{{\sigma_i}^2} - \sum_{i=1}^{N} \ln\left(\sigma_i \sqrt{2\pi}\right)$$

e la sua derivata rispetto al parametro λ

$$\frac{\partial\,(\ln \mathcal{L})}{\partial \lambda} \;=\; \sum_{i=1}^{N} \frac{x_i - \lambda}{{\sigma_i}^2} \;=\; \left(\sum\nolimits_i \frac{1}{{\sigma_i}^2}\right) \left(\frac{\sum\nolimits_i \dfrac{x_i}{{\sigma_i}^2}}{\sum\nolimits_i \dfrac{1}{{\sigma_i}^2}} - \lambda \right) \;.$$

Pertanto la media dei dati, pesati con coefficienti inversamente proporzionali alle varianze, è una stima di minima varianza per λ. Se le N varianze sono poi tutte uguali tra loro e di valore σ^2, risulta

$$\frac{\partial\,(\ln \mathcal{L})}{\partial \lambda} \;=\; \frac{1}{\sigma^2} \left[\left(\sum_{i=1}^{N} x_i \right) - N\lambda \right] \;=\; \frac{N}{\sigma^2}\,(\bar{x} - \lambda) \;=\; \frac{\bar{x} - \lambda}{R}$$

ed in tal caso la *media aritmetica* del campione è una stima di minima varianza per λ. Sempre in tal caso è poi

$$R(\lambda) \;\equiv\; R \;=\; \frac{\sigma^2}{N}$$

con

$$\tau(\lambda) \;\equiv\; \lambda \qquad\qquad \text{e} \qquad\qquad \tau'(\lambda) \;=\; 1$$

dunque

$$\mathrm{Var}(\bar{x}) \;=\; \tau'\,R \;=\; \frac{\sigma^2}{N}$$

come d'altra parte già si sapeva.

Qui la media del campione è un esempio di statistica sufficiente per λ; infatti non ha alcuna importanza quali siano i singoli valori x_i: ma se le medie di due diversi campioni sono uguali, le conclusioni che si possono trarre sul valore di λ sono le medesime.

Supponendo di conoscere il valore medio λ, la stima della varianza σ^2 si ottiene cercando lo zero della derivata logaritmica

$$\frac{\partial\,(\ln \mathcal{L})}{\partial \sigma} \;=\; \frac{1}{\sigma^3} \left[\sum_{i=1}^{N} (x_i - \lambda)^2 \right] - \frac{N}{\sigma} \;=\; \frac{N}{\sigma^3} \left\{ \left[\frac{1}{N} \sum_{i=1}^{N} (x_i - \lambda)^2 \right] - \sigma^2 \right\}$$

la quale ha la forma richiesta perché la soluzione

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \lambda)^2$$

sia una stima di σ^2 con minima varianza, data da

$$\mathrm{Var}\left\{\frac{1}{N}\sum_{i=1}^{N}(x_i-\lambda)^2\right\} \;=\; \tau' R \;=\; 2\,\sigma\,\frac{\sigma^3}{N} \;=\; \frac{2\sigma^4}{N}$$

essendo $R(\sigma) = \sigma^3/N$, $\tau(\sigma) = \sigma^2$ e $\tau'(\sigma) = 2\sigma$: questo risultato è lo stesso trovato nell'appendice B.

Il valore di λ tuttavia non è generalmente noto, e l'uso della media aritmetica del campione $\bar{x}$ comporta una distorsione che si corregge, come si è visto, ponendo $N-1$ in luogo di N.

Appendice F

La licenza GNU GPL (General Public License)

Questo capitolo contiene la licenza GNU GPL, sotto la quale questo libro viene distribuito, sia nella versione originale inglese[1] (la sola dotata di valore legale) che in una traduzione non ufficiale in italiano[2] che aiuti chi ha difficoltà con l'inglese legale a comprendere meglio il significato della licenza stessa.

F.1 The GNU General Public License

Version 2, June 1991
Copyright © 1989, 1991 Free Software Foundation, Inc.

59 Temple Place - Suite 330, Boston, MA 02111-1307, USA

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

Preamble

The licenses for most software are designed to take away your freedom to share and change it. By contrast, the GNU General

[1] `http://www.gnu.org/licenses/gpl.html`

[2] La traduzione è dovuta al gruppo *Pluto* (PLUTO Linux/Lumen Utentibus Terrarum Orbis, `http://www.pluto.linux.it/`).

Public License is intended to guarantee your freedom to share and change free software — to make sure the software is free for all its users. This General Public License applies to most of the Free Software Foundation's software and to any other program whose authors commit to using it. (Some other Free Software Foundation software is covered by the GNU Library General Public License instead.) You can apply it to your programs, too.

When we speak of free software, we are referring to freedom, not price. Our General Public Licenses are designed to make sure that you have the freedom to distribute copies of free software (and charge for this service if you wish), that you receive source code or can get it if you want it, that you can change the software or use pieces of it in new free programs; and that you know you can do these things.

To protect your rights, we need to make restrictions that forbid anyone to deny you these rights or to ask you to surrender the rights. These restrictions translate to certain responsibilities for you if you distribute copies of the software, or if you modify it.

For example, if you distribute copies of such a program, whether gratis or for a fee, you must give the recipients all the rights that you have. You must make sure that they, too, receive or can get the source code. And you must show them these terms so they know their rights.

We protect your rights with two steps: (1) copyright the software, and (2) offer you this license which gives you legal permission to copy, distribute and/or modify the software.

Also, for each author's protection and ours, we want to make certain that everyone understands that there is no warranty for this free software. If the software is modified by someone else and passed on, we want its recipients to know that what they have is not the original, so that any problems introduced by others will not reflect on the original authors' reputations.

Finally, any free program is threatened constantly by software patents. We wish to avoid the danger that redistributors of a free program will individually obtain patent licenses, in effect making the program proprietary. To prevent this, we have made it clear that any patent must be licensed for everyone's free use or not licensed at all.

The precise terms and conditions for copying, distribution and modification follow.

GNU General Public License

Terms and Conditions For Copying, Distribution and Modification

0. This License applies to any program or other work which contains a notice placed by the copyright holder saying it may be distributed under the terms of this General Public License. The "Program", below, refers to any such program or work, and a "work based on the Program" means either the Program or any derivative work under copyright law: that is to say, a work containing the Program or a portion of it, either verbatim or with modifications and/or translated into another language. (Hereinafter, translation is included without limitation in the term "modification".) Each licensee is addressed as "you".

   Activities other than copying, distribution and modification are not covered by this License; they are outside its scope. The act of running the Program is not restricted, and the output from the Program is covered only if its contents constitute a work based on the Program (independent of having been made by running the Program). Whether that is true depends on what the Program does.

1. You may copy and distribute verbatim copies of the Program's source code as you receive it, in any medium, provided that you conspicuously and appropriately publish on each copy an appropriate copyright notice and disclaimer of warranty; keep intact all the notices that refer to this License and to the absence of any warranty; and give any other recipients of the Program a copy of this License along with the Program.

   You may charge a fee for the physical act of transferring a copy, and you may at your option offer warranty protection in exchange for a fee.

2. You may modify your copy or copies of the Program or any portion of it, thus forming a work based on the Program, and copy and distribute such modifications or work under the terms of Section 1 above, provided that you also meet all of these conditions:

   (a) You must cause the modified files to carry prominent notices stating that you changed the files and the date of any change.

(b) You must cause any work that you distribute or publish, that in whole or in part contains or is derived from the Program or any part thereof, to be licensed as a whole at no charge to all third parties under the terms of this License.

(c) If the modified program normally reads commands interactively when run, you must cause it, when started running for such interactive use in the most ordinary way, to print or display an announcement including an appropriate copyright notice and a notice that there is no warranty (or else, saying that you provide a warranty) and that users may redistribute the program under these conditions, and telling the user how to view a copy of this License. (Exception: if the Program itself is interactive but does not normally print such an announcement, your work based on the Program is not required to print an announcement.)

These requirements apply to the modified work as a whole. If identifiable sections of that work are not derived from the Program, and can be reasonably considered independent and separate works in themselves, then this License, and its terms, do not apply to those sections when you distribute them as separate works. But when you distribute the same sections as part of a whole which is a work based on the Program, the distribution of the whole must be on the terms of this License, whose permissions for other licensees extend to the entire whole, and thus to each and every part regardless of who wrote it.

Thus, it is not the intent of this section to claim rights or contest your rights to work written entirely by you; rather, the intent is to exercise the right to control the distribution of derivative or collective works based on the Program.

In addition, mere aggregation of another work not based on the Program with the Program (or with a work based on the Program) on a volume of a storage or distribution medium does not bring the other work under the scope of this License.

3. You may copy and distribute the Program (or a work based on it, under Section 2) in object code or executable form under the terms of Sections 1 and 2 above provided that you also do one of the following:

(a) Accompany it with the complete corresponding machine-readable source code, which must be distributed under the terms of Sections 1 and 2 above on a medium customarily used for software interchange; or,

(b) Accompany it with a written offer, valid for at least three years, to give any third party, for a charge no more than your cost of physically performing source distribution, a complete machine-readable copy of the corresponding source code, to be distributed under the terms of Sections 1 and 2 above on a medium customarily used for software interchange; or,

(c) Accompany it with the information you received as to the offer to distribute corresponding source code. (This alternative is allowed only for noncommercial distribution and only if you received the program in object code or executable form with such an offer, in accord with Subsection b above.)

The source code for a work means the preferred form of the work for making modifications to it. For an executable work, complete source code means all the source code for all modules it contains, plus any associated interface definition files, plus the scripts used to control compilation and installation of the executable. However, as a special exception, the source code distributed need not include anything that is normally distributed (in either source or binary form) with the major components (compiler, kernel, and so on) of the operating system on which the executable runs, unless that component itself accompanies the executable.

If distribution of executable or object code is made by offering access to copy from a designated place, then offering equivalent access to copy the source code from the same place counts as distribution of the source code, even though third parties are not compelled to copy the source along with the object code.

4. You may not copy, modify, sublicense, or distribute the Program except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense or distribute the Program is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

5. You are not required to accept this License, since you have not signed it. However, nothing else grants you permission to modify or distribute the Program or its derivative works. These actions are prohibited by law if you do not accept this License. Therefore, by modifying or distributing the Program (or any work based on the Program), you indicate

your acceptance of this License to do so, and all its terms and conditions for copying, distributing or modifying the Program or works based on it.

6. Each time you redistribute the Program (or any work based on the Program), the recipient automatically receives a license from the original licensor to copy, distribute or modify the Program subject to these terms and conditions. You may not impose any further restrictions on the recipients' exercise of the rights granted herein. You are not responsible for enforcing compliance by third parties to this License.

7. If, as a consequence of a court judgment or allegation of patent infringement or for any other reason (not limited to patent issues), conditions are imposed on you (whether by court order, agreement or otherwise) that contradict the conditions of this License, they do not excuse you from the conditions of this License. If you cannot distribute so as to satisfy simultaneously your obligations under this License and any other pertinent obligations, then as a consequence you may not distribute the Program at all. For example, if a patent license would not permit royalty-free redistribution of the Program by all those who receive copies directly or indirectly through you, then the only way you could satisfy both it and this License would be to refrain entirely from distribution of the Program.

   If any portion of this section is held invalid or unenforceable under any particular circumstance, the balance of the section is intended to apply and the section as a whole is intended to apply in other circumstances.

   It is not the purpose of this section to induce you to infringe any patents or other property right claims or to contest validity of any such claims; this section has the sole purpose of protecting the integrity of the free software distribution system, which is implemented by public license practices. Many people have made generous contributions to the wide range of software distributed through that system in reliance on consistent application of that system; it is up to the author/donor to decide if he or she is willing to distribute software through any other system and a licensee cannot impose that choice.

   This section is intended to make thoroughly clear what is believed to be a consequence of the rest of this License.

8. If the distribution and/or use of the Program is restricted in certain countries either by patents or by copyrighted interfaces, the original copyright holder who places the Program under this License may add

an explicit geographical distribution limitation excluding those countries, so that distribution is permitted only in or among countries not thus excluded. In such case, this License incorporates the limitation as if written in the body of this License.

9. The Free Software Foundation may publish revised and/or new versions of the General Public License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.

   Each version is given a distinguishing version number. If the Program specifies a version number of this License which applies to it and "any later version", you have the option of following the terms and conditions either of that version or of any later version published by the Free Software Foundation. If the Program does not specify a version number of this License, you may choose any version ever published by the Free Software Foundation.

10. If you wish to incorporate parts of the Program into other free programs whose distribution conditions are different, write to the author to ask for permission. For software which is copyrighted by the Free Software Foundation, write to the Free Software Foundation; we sometimes make exceptions for this. Our decision will be guided by the two goals of preserving the free status of all derivatives of our free software and of promoting the sharing and reuse of software generally.

No Warranty

11. Because the program is licensed free of charge, there is no warranty for the program, to the extent permitted by applicable law. Except when otherwise stated in writing the copyright holders and/or other parties provide the program "as is" without warranty of any kind, either expressed or implied, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose. The entire risk as to the quality and performance of the program is with you. Should the program prove defective, you assume the cost of all necessary servicing, repair or correction.

12. In no event unless required by applicable law or agreed to in writing will any copyright holder, or any other party who

may modify and/or redistribute the program as permitted above, be liable to you for damages, including any general, special, incidental or consequential damages arising out of the use or inability to use the program (including but not limited to loss of data or data being rendered inaccurate or losses sustained by you or third parties or a failure of the program to operate with any other programs), even if such holder or other party has been advised of the possibility of such damages.

End of Terms and Conditions

Appendix: How to Apply These Terms to Your New Programs

If you develop a new program, and you want it to be of the greatest possible use to the public, the best way to achieve this is to make it free software which everyone can redistribute and change under these terms.

To do so, attach the following notices to the program. It is safest to attach them to the start of each source file to most effectively convey the exclusion of warranty; and each file should have at least the "copyright" line and a pointer to where the full notice is found.

<one line to give the program's name and a brief idea of what it does.>
Copyright (C) <year> <name of author>

This program is free software; you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation; either version 2 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program; if not, write to the Free Software Foundation, Inc., 59 Temple Place - Suite 330, Boston, MA 02111-1307, USA.

Also add information on how to contact you by electronic and paper mail. If the program is interactive, make it output a short notice like this when it starts in an interactive mode: