Così come abbiamo dimostrato che, al fine di correggere questa sottostima (in media) per le misure ripetute, occorre dividere la somma dei quadrati degli scarti per $N-1$ invece che per N, si potrebbe analogamente dimostrare che una corretta stima dell'errore dei punti misurati si ha, in media, dividendo l'analoga somma per $N-2$; in definitiva, che la corretta stima di σ_y è data dalla formula

$$\sigma_y{}^2 = \frac{\sum_{i=1}^{N}\left[(a+bx_i)-y_i\right]^2}{N-2} \quad .$$

In essa a numeratore compare la somma dei quadrati dei residui, cioè delle "distanze" dei punti misurati (x_i, y_i) dalla retta interpolante di equazione $a+bx$ calcolate secondo la direzione parallela all'asse delle ordinate. Questa formula[2] permette una corretta stima dell'errore dei dati interpolati, qualora sia impossibile (o scomodo) determinarli per altra via; l'errore è stimato dai residui dei dati sperimentali, ed è quindi scientificamente affidabile.

Il fatto che la corretta stima dell'errore si ottenga dividendo per $N-2$ invece che per N deve essere messo in relazione con il fatto che gli scarti, invece che rispetto al valore vero, sono calcolati rispetto ad un valore stimato che dipende da *due* parametri, che sono a loro volta stati preventivamente determinati sulla base dei dati sperimentali: cioè i due coefficienti a e b dell'equazione della retta.

Nell'analogo caso della stima dell'errore quadratico medio di una variabile casuale unidimensionale, gli scarti erano calcolati rispetto ad un valore che, unica grandezza necessaria, veniva preventivamente determinato sulla base delle misure: appunto la media aritmetica.

In generale, disponendo di N dati sperimentali dai quali possiamo determinare un valore dell'errore quadratico medio che dipende da M parametri che debbano essere preventivamente derivati dai dati stessi, la modifica da apportare alla formula per ottenere una corretta valutazione dell'errore della popolazione consiste nel dividere la somma dei quadrati degli scarti per un fattore $N-M$.

11.4.3 Interpolazione con una retta per l'origine

Se conosciamo altri vincoli cui debba soddisfare la legge che mette in relazione i valori delle variabili misurate x e y, possiamo imporre che la retta corrispondente appartenga ad un particolare sottoinsieme delle rette del piano; ad esempio, un caso che si può presentare è che la retta sia vincolata

[2] Una formula equivalente (ma più semplice) per il calcolo di σ_y si può trovare nell'equazione (C.8) alla pagina 264.

a passare per una posizione particolare, che supporremo qui essere l'origine degli assi coordinati.

Una generica retta per l'origine ha equazione $y = mx$; ammesso ancora che gli errori commessi riguardino soltanto la misura della y e non quella della x, che tutti i vari y_i abbiano errori distribuiti secondo la legge normale e tra loro uguali, e che le misure siano tra loro indipendenti, il problema dell'interpolazione lineare si riduce a trovare tra le infinite rette passanti per l'origine quella che rende massima la funzione di verosimiglianza

$$\mathcal{L}(x_1, y_1, x_2, y_2, \ldots, x_N, y_N; m) \;=\; \prod_{i=1}^{N} \frac{1}{\sigma_y\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{mx_i-y_i}{\sigma_y}\right)^2} \quad .$$

Passando al logaritmo naturale di $\mathcal{L}$, è facile vedere che la soluzione ricercata è sempre quella che rende minima la somma dei quadrati dei residui dai punti misurati: che cioè minimizza la funzione

$$f(m) = \sum_{i=1}^{N} (mx_i - y_i)^2 \quad .$$

La derivata prima di f vale

$$\frac{\mathrm{d}f}{\mathrm{d}m} \;=\; 2\sum_{i=1}^{N}(mx_i - y_i)\,x_i \;=\; 2\left(m\sum_{i=1}^{N} x_i{}^2 \;-\; \sum_{i=1}^{N} x_i y_i\right)$$

e, imponendo che essa sia nulla, l'estremante si ha per

$$m = \frac{\sum_i x_i y_i}{\sum_i x_i{}^2}$$

e corrisponde in effetti ad un minimo. La legge di propagazione degli errori è esatta anche in questo caso, perché m è una combinazione lineare delle variabili affette da errore (le y_i); il coefficiente di y_i nella combinazione vale

$$\frac{x_i}{\sum_k x_k{}^2}$$

e quindi

$$\sigma_m{}^2 \;=\; \sum_{i=1}^{N}\left(\frac{x_i}{\sum_k x_k{}^2}\right)^2 \sigma_y{}^2 \;=\; \frac{\sigma_y{}^2}{\left(\sum_k x_k{}^2\right)^2}\sum_{i=1}^{N} x_i{}^2 \;=\; \frac{\sigma_y{}^2}{\sum_k x_k{}^2}$$

e la formula per il calcolo degli errori a posteriori diventa

$$\sigma_y{}^2 = \frac{\sum_i (mx_i - y_i)^2}{N-1}$$

visto che il parametro da cui l'errore quadratico medio dipende e che deve essere stimato sulla base dei dati è uno soltanto: m.

11.4.4 Interpolazione lineare nel caso generale

Le condizioni 1) e 2) sugli errori delle grandezze misurate x e y date nel paragrafo 11.4.1 non potranno ovviamente mai essere verificate esattamente; come ci si deve comportare quando nemmeno in prima approssimazione le possiamo considerare vere?

Se gli errori quadratici medi delle y_i sono tra loro diversi, non è più possibile raccogliere a fattore comune $1/\sigma^2$ nell'espressione del logaritmo della verosimiglianza; e ciascun addendo sarà diviso per il corrispondente errore σ_i. In definitiva la retta più verosimile si trova cercando il minimo della funzione

$$f(a,b) = \sum_{i=1}^{N} \left[\frac{(a + bx_i) - y_i}{\sigma_i} \right]^2 \quad .$$

Questo avviene quando

$$\begin{cases} a = \dfrac{1}{\Delta} \left[\left(\sum_i \dfrac{x_i^2}{\sigma_i^2} \right) \cdot \left(\sum_i \dfrac{y_i}{\sigma_i^2} \right) - \left(\sum_i \dfrac{x_i}{\sigma_i^2} \right) \cdot \left(\sum_i \dfrac{x_i y_i}{\sigma_i^2} \right) \right] \\ b = \dfrac{1}{\Delta} \left[\left(\sum_i \dfrac{1}{\sigma_i^2} \right) \cdot \left(\sum_i \dfrac{x_i y_i}{\sigma_i^2} \right) - \left(\sum_i \dfrac{x_i}{\sigma_i^2} \right) \cdot \left(\sum_i \dfrac{y_i}{\sigma_i^2} \right) \right] \end{cases}$$

in cui si è posto

$$\Delta = \left(\sum_i \frac{1}{\sigma_i^2} \right) \cdot \left(\sum_i \frac{x_i^2}{\sigma_i^2} \right) - \left(\sum_i \frac{x_i}{\sigma_i^2} \right)^2 \quad .$$

Le varianze di a e di b saranno poi date dalle

$$\begin{cases} {\sigma_a}^2 = \dfrac{1}{\Delta} \sum_i \dfrac{x_i^2}{\sigma_i^2} \\ {\sigma_b}^2 = \dfrac{1}{\Delta} \sum_i \dfrac{1}{\sigma_i^2} \end{cases}$$

Si deve tuttavia osservare che per applicare questo metodo è necessario conoscere, per altra via e preventivamente, tutte le N varianze ${\sigma_i}^2$. Ciò può essere molto laborioso o addirittura impossibile, e non risulta conveniente rinunciare ad una stima unica e ragionevole ${\sigma_y}^2$ di queste varianze per tener conto di una variazione, generalmente debole, delle σ_i in un intervallo limitato di valori della x.

Volendo tener conto dell'errore su entrambe le variabili x ed y, non è generalmente possibile usare un metodo, descritto in alcuni testi, consistente nel cercare la retta che rende minima la somma dei quadrati delle distanze

dai punti, misurate però *ortogonalmente* alla retta stessa: a prescindere dalla complicazione della soluzione di un sistema di equazioni non lineari, resta il fatto che se x ed y sono due grandezze fisiche diverse, o anche soltanto misurate con strumenti e metodi diversi, i loro errori quadratici medi sono generalmente differenti; mentre la distanza sul piano attribuisce lo stesso peso agli scarti in x ed a quelli in y.

Per applicare questo metodo si dovrebbe conoscere, per via indipendente, almeno il rapporto tra σ_x e σ_y; e rappresentare i valori misurati (x_i, y_i) non già sul piano $\{x, y\}$, bensì su quello delle variabili ridotte $\{x/\sigma_x\ ,\ y/\sigma_y\}$.

Per solito nella pratica si preferisce considerare affetta da errore una soltanto delle variabili, ad esempio la y, la scelta cadendo generalmente su quella determinata in maniera *più indiretta*, e che risente perciò degli errori di tutte le altre grandezze misurate direttamente; così, in un diagramma velocità-tempo trascorso o velocità-spazio percorso, si assumerà affetta da errore la sola velocità.

Un eventuale errore sulla x si *propagherà* attraverso la relazione funzionale anche alla y, e, se l'errore quadratico medio σ_y è stimato dai dati sperimentali, esso congloberà anche l'indeterminazione dovuta alla x.

Per meglio chiarire il concetto, consideriamo la legge $v = v(t) = v_0 + gt$ che descrive la caduta di un grave, e pensiamo di misurare la sua velocità in un certo istante: nell'ipotesi originale t sarebbe determinabile esattamente, ma l'imprecisione nella misura delle velocità ci darebbe valori di v compresi in un intervallo di ampiezza non nulla (dipendente dall'errore quadratico medio σ_v).

Se delle due grandezze, al contrario, fosse la velocità ad essere conoscibile esattamente, l'impossibilità di determinare con precisione l'istante t in cui essa deve essere misurata ci darebbe ugualmente valori di v distribuiti in un intervallo di ampiezza non nulla (legata stavolta a $g \cdot \sigma_t$).

Indicando, insomma, con σ_x e σ_y gli errori (sempre supposti costanti) di ognuna delle determinazioni (sempre supposte indipendenti) x_i e y_i, la formula dell'errore a posteriori ci permette di ricavare dai dati una ragionevole stima non tanto del solo σ_y quanto, piuttosto, della combinazione (quadratica) dell'errore intrinseco delle ordinate e di quello intrinseco delle ascisse propagato sulle ordinate:

$$\sigma^2 \approx {\sigma_y}^2 + (b^* \ \sigma_x)^2$$

(ove b^* è il valore vero della pendenza della retta).

11.4.5 Interpolazione non lineare

Formule analoghe a quelle trovate si possono ricavare per risolvere il problema dell'interpolazione di curve di ordine superiore al primo (parabole, cubiche, polinomiali in genere) ad un insieme di dati sperimentali, sempre usando il metodo della massima verosimiglianza.

Nel caso poi ci si trovasse di fronte ad una curva di equazione diversa da un polinomio, in parecchi casi è possibile *linearizzare* la relazione cambiando variabile: così, ad esempio, se due grandezze hanno tra loro una relazione di tipo esponenziale, il logaritmo naturale ne avrà una di tipo lineare:

$$y = Ke^{-bx} \qquad \Longleftrightarrow \qquad \ln y \;=\; \ln K - bx \;=\; a - bx \;\; .$$

11.5 Altre applicazioni della stima di massima verosimiglianza

Per concludere il capitolo, presentiamo altre tre applicazioni del metodo della massima verosimiglianza: la stima delle probabilità ignote di un insieme di modalità esclusive ed esaurienti cui può dar luogo un fenomeno casuale; la stima sia della media che della varianza di una popolazione normale; e la stima del range di una popolazione uniforme.

11.5.1 Stima di probabilità

Supponiamo che un fenomeno casuale possa dare origine ad un numero finito M di eventualità, ognuna delle quali sia associata ad un valore p_i ignoto della probabilità; se, eseguite N prove indipendenti, indichiamo con n_i la frequenza assoluta con cui ognuna delle M eventualità si è presentata nel corso di esse, quale è la stima di massima verosimiglianza, $\hat{p}_i$, per le incognite probabilità p_i?

La funzione di verosimiglianza è, visto che la generica delle M eventualità, di probabilità p_i, si è presentata n_i volte, data[3] da

$$\mathcal{L}(\boldsymbol{n};\boldsymbol{p}) = \prod_{i=1}^{M} {p_i}^{n_i}$$

[3] A meno di un fattore moltiplicativo costante, corrispondente al numero di modi in cui N oggetti si possono ripartire tra M gruppi in modo che ogni gruppo sia composto da n_i oggetti; numero delle *partizioni ordinate* (vedi in proposito il paragrafo A.7).

(in cui abbiamo indicato sinteticamente con due vettori, $\boldsymbol{n}$ e $\boldsymbol{p}$, entrambi di dimensione M, l'insieme degli M valori n_i e quello degli M valori p_i rispettivamente); ed il suo logaritmo da

$$\ln \mathcal{L}(\boldsymbol{n};\boldsymbol{p}) = \sum_{i=1}^{M} n_i \ln p_i \;\; . \tag{11.12}$$

Il problema della ricerca del massimo della (11.12) è complicato dal fatto che i valori delle p_i non sono liberi, ma *vincolati* dalla condizione

$$\sum_{i=1}^{M} p_i = 1 \;\; . \tag{11.13}$$

Usiamo quindi il metodo dei moltiplicatori di Lagrange, costruendo la funzione

$$f(\boldsymbol{n};\boldsymbol{p}) \;=\; \sum_{i=1}^{M} n_i \ln p_i - \lambda \left(\sum_{i=1}^{M} p_i - 1 \right) \tag{11.14}$$

e risolvendo il sistema delle $M+1$ equazioni, nelle $M+1$ incognite p_i e λ, composto dalla (11.13) e dalle altre M ottenute derivando la (11.14) rispetto ad ognuna delle p_k:

$$\frac{\partial f}{\partial p_k} \;=\; n_k \frac{1}{p_k} - \lambda \;=\; 0 \qquad\qquad (k = 1, 2, \ldots, M) \;\; .$$

Da quest'ultima si ricava

$$\hat{p}_k = \frac{n_k}{\lambda}$$

e, sostituendo nella (11.13),

$$\sum_{i=1}^{M} \hat{p}_i \;=\; \frac{1}{\lambda} \sum_{i=1}^{M} n_i \;=\; \frac{N}{\lambda} \;=\; 1$$

si ottiene

$$\lambda = N$$

per cui in definitiva la soluzione di massima verosimiglianza è (cosa non sorprendente) data dalle

$$\hat{p}_i = \frac{n_i}{N} \;\; .$$

11.5.2 Media e varianza di una popolazione normale

Abbiamo già visto nel paragrafo 11.3 che, *ammessa nota la varianza* σ^2 di una popolazione normale, il suo valore medio μ ha come stima di massima verosimiglianza la media aritmetica $\bar{x}$ di un campione di stime indipendenti; vogliamo ora stimare *contemporaneamente sia* μ *che* σ dai dati, usando sempre il metodo della massima verosimiglianza.

La densità di probabilità vale

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

ed il suo logaritmo

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

Il logaritmo della funzione di verosimiglianza è

$$\ln\mathcal{L}(\boldsymbol{x};\mu,\sigma) \;=\; \sum_{i=1}^{N} \ln f(x_i;\mu,\sigma)$$

e dunque

$$\ln\mathcal{L}(\boldsymbol{x};\mu,\sigma) \;=\; -N\ln\sigma - N\,\ln\sqrt{2\pi} - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\mu)^2 \quad ;$$

e le sue derivate parziali prime sono

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

e

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

Il sistema ottenuto annullando le due derivate parziali prime ha l'unica soluzione (in effetti un massimo) data da

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

Questo era già noto: entrambe le stime, come sappiamo, sono consistenti; però la seconda non è imparziale (ma può essere resa tale moltiplicandola per un opportuno fattore di correzione).

In sostanza il fatto che la varianza della popolazione abbia un determinato valore (come assunto nel paragrafo 11.3) non cambia il fatto che la nostra migliore stima del valore medio della popolazione sia comunque data dalla media aritmetica del campione: vedremo poi nel paragrafo 12.1 che il valore medio del campione e la sua varianza sono variabili casuali *statisticamente indipendenti tra loro.*

11.5.3 Range di una popolazione uniforme

Sia una variabile casuale x distribuita uniformemente tra un estremo inferiore noto, che senza perdere in generalità possiamo supporre sia lo zero, ed un estremo superiore ignoto A; in questo caso dobbiamo innanzi tutto osservare sia che il dominio di definizione della funzione di frequenza $f(x)$ della x *dipende dal parametro* che dobbiamo stimare, sia che $f(x)$ e la sua derivata prima hanno dei punti di discontinuità: e non possiamo in conseguenza garantire *a priori* né la consistenza, né la massima efficienza asintotica del metodo usato.

Comunque, introducendo la cosiddetta *funzione gradino* (o *step function*) $S(x)$, definita attraverso la

$$\begin{cases} S(x) = 0 & (x < 0) \\ S(x) = 1 & (x \geq 0) \end{cases}$$

la funzione di frequenza $f(x)$ si può anche scrivere

$$f(x) \;=\; \frac{1}{A}\,S(x)\,S(A-x)$$

e la funzione di verosimiglianza

$$\mathcal{L}(x_1,x_2,\ldots,x_N;A) = \frac{1}{A^N}\,S(x_{\min})\,S(A-x_{\max}) \quad .$$

Come sappiamo, ammesso noto il valore del parametro A essa rappresenta la densità di probabilità di ottenere gli N valori $x_i \in [-\infty,+\infty]$; se invece si considera A come l'unica variabile e si ammettono noti gli N valori x_i, rappresenta la densità di probabilità che un dato A abbia prodotto i dati osservati. Ma in quest'ultimo caso $S(x_{\min}) \equiv 1$, e la funzione di verosimiglianza si riduce alla

$$\mathcal{L}(A) = \frac{1}{A^N}\,S(A-x_{\max}) \tag{11.15}$$

che è nulla per $A < x_{\max}$ e monotona strettamente decrescente per $A \geq x_{\max}$; ed ammette quindi un unico massimo all'estremo del dominio, che vale

$$\hat{A} = x_{\max} \ . \tag{11.16}$$

Valore medio e varianza della stima valgono, come già sappiamo dal paragrafo 8.1.3,

$$E(\hat{A}) \ = \ A^* - \frac{A^*}{N+1} \ = \ \frac{N}{N+1} A^*$$

e

$$\mathrm{Var}(\hat{A}) \ = \ \frac{N}{(N+1)^2 \, (N+2)} \, (A^*)^2$$

e quindi la stima è consistente, ma non imparziale; una stima imparziale è invece

$$\bar{A} \ = \ \frac{N+1}{N} x_{\max} \ = \ x_{\max} \left(1 + \frac{1}{N}\right) \ ,$$

di varianza ovviamente superiore per un fattore $(1 + 1/N)^2$. È anche ovvio, dalla forma sia della (11.15) che della (11.16), che $\hat{A}$ è una stima sufficiente di A^*.

11.5.4 Stima della vita media di una particella

Nel processo di decadimento di una particella instabile, indichiamo con τ l'incognita vita media e con t i tempi (propri) di decadimento osservati; tempi che (come sappiamo) seguono la distribuzione esponenziale:

$$f(t;\tau) = \frac{1}{\tau} e^{-\frac{t}{\tau}} \qquad \Longrightarrow \qquad \begin{cases} E(t) = \tau \\ \mathrm{Var}(t) = \tau^2 \end{cases}$$

Ammettendo per semplicità che l'osservazione avvenga con una efficienza unitaria, o, in altre parole, che tutti i decadimenti vengano osservati, la funzione di verosimiglianza si scrive

$$\mathcal{L} \ = \ \prod_{k=1}^{N} f(t_k;\tau) \ = \ \frac{1}{\tau^N} e^{-\frac{1}{\tau}\sum_k t_k} \ ,$$

ed il suo logaritmo vale

$$\ln(\mathcal{L}) \ = \ -N \ln \tau - \frac{1}{\tau} \sum_{k=1}^{N} t_k \ = \ -N \left(\frac{\bar{t}}{\tau} + \ln \tau \right) \ .$$

Derivando rispetto al parametro e cercando gli estremanti,

$$\frac{\mathrm{d} \ln(\mathcal{L})}{\mathrm{d}\tau} \ = \ \frac{N}{\tau^2} (\bar{t} - \tau) \ = \ 0 \ ;$$

e quindi l'unico estremante della funzione di verosimiglianza si ha per

$$\hat{\tau} = \bar{t} \ .$$

Se calcoliamo la derivata seconda,

$$\frac{\mathrm{d}^2 \ln(\mathcal{L})}{\mathrm{d}\tau^2} = -\frac{N}{\tau^3} (2\bar{t} - \tau)$$

essa, calcolata per $t = \bar{t}$ è negativa; quindi l'unico estremante è effettivamente un punto di massimo.

La soluzione di massima verosimiglianza $\hat{\tau} = \bar{t}$ è *consistente ed imparziale* (essendo il valore medio del campione); *di varianza minima* (per il teorema di Cramér–Rao); inoltre la stima è *sufficiente* (riassume insomma tutta l'informazione del campione).

Normalmente l'efficienza non è però unitaria; ad esempio il nostro rivelatore può avere dimensioni confrontabili col cammino medio delle particelle, che possono quindi uscirne prima di decadere. In questo caso, visto che i decadimenti possono essere stati osservati solo essendo avvenuti all'interno di un intervallo compreso tra un valore temporale minimo (eventualmente nullo) ed uno massimo (ad esempio dipendente dalla posizione del decadimento, dalla direzione di emissione dei suoi prodotti, dalle dimensioni del rivelatore, ...) — intervallo differente per ognuno dei decadimenti — dovremo costruire la funzione di verosimiglianza considerando le probabilità di osservazione *condizionate* dall'essere il decadimento i-esimo avvenuto tra un certo $(t_{\min})_i$ ed un certo $(t_{\max})_i$:

$$\mathcal{L} = \prod_{i=1}^{N} \left[\frac{\frac{1}{\tau} e^{-\frac{t_i}{\tau}}}{e^{-\frac{(t_{\min})_i}{\tau}} - e^{-\frac{(t_{\max})_i}{\tau}}} \right] \tag{11.17}$$

Il denominatore della (11.17) rappresenta infatti la probabilità di decadere tra il tempo $t_{\min}$ e quello $t_{\max}$, come è immediato ricavare dalla funzione di distribuzione della densità di probabilità esponenziale, che vale

$$F(t) \;=\; \int_0^t \frac{1}{\tau} e^{-\frac{x}{\tau}}\, \mathrm{d}x \;=\; 1 - e^{-\frac{t}{\tau}} \;;$$

dalla (11.17) si ricava poi

$$\ln(\mathcal{L}) = -N \ln \tau + \sum_{i=1}^{N} \left\{ -\frac{t_i}{\tau} - \ln\left[e^{-\frac{(t_{\min})_i}{\tau}} - e^{-\frac{(t_{\max})_i}{\tau}} \right] \right\}$$

e, posto per brevità

$$\varphi_i(\tau) = \frac{(t_{\min})_i^2 \cdot e^{-\frac{(t_{\min})_i}{\tau}} - (t_{\max})_i^2 \cdot e^{-\frac{(t_{\max})_i}{\tau}}}{e^{-\frac{(t_{\min})_i}{\tau}} - e^{-\frac{(t_{\max})_i}{\tau}}}$$

si arriva alla

$$\frac{\mathrm{d}\,\ln(\mathcal{L})}{\mathrm{d}\tau} \;=\; \frac{1}{\tau^2} \left\{ \sum_{i=1}^{N} [t_i - \varphi_i(\tau)] - N\tau \right\} \;=\; 0$$

che bisogna risolvere in modo numerico. Non si può inoltre in questo caso garantire che le proprietà precedentemente delineate per $\hat{\tau}$ (consistenza, normalità, efficienza, ...) siano ancora valide, almeno per N finito. Può darsi che la funzione di verosimiglianza ammetta più di un massimo, e non si sa a priori quale di essi convergerà verso τ^*; e, per finire, l'errore della stima deve essere ricavato dalla concavità della funzione di verosimiglianza, supposta approssimativamente normale.

Capitolo 12

La verifica delle ipotesi (I)

Una volta eseguita una misura, si può voler controllare se i nostri risultati possono confermare o rigettare una determinata ipotesi riguardante il fenomeno fisico che li ha prodotti; naturalmente, visto che risultati di una misura comunque lontani dal valore vero sono sempre possibili (anche se con probabilità sempre più piccole al crescere dello scarto), una qualunque ipotesi sulla grandezza fisica misurata potrà essere confermata o rigettata dai dati solo ad un certo livello di probabilità.

Qui ci occuperemo inoltre di alcune funzioni di frequenza collegate a quella di Gauss, ossia della distribuzione del χ^2, di quella di Student[1] e di quella di Fisher; e dell'uso che di esse si può fare per la verifica di ipotesi statistiche: quali ad esempio quella che un campione di dati sperimentali provenga da una popolazione descritta da una densità di probabilità nota a priori; o quella che il valore vero della grandezza misurata coincida con un valore determinato, noto anch'esso a priori.

12.1 La distribuzione del χ^2

Se le N variabili casuali x_i, tra loro statisticamente indipendenti, sono variabili normali standardizzate (ovverosia distribuite secondo la legge normale con media 0 e varianza 1), si può dimostrare che la nuova variabile

[1] "Student" è lo pseudonimo con cui vennero pubblicati i lavori statistici di William Gosset, scienziato inglese vissuto dal 1876 al 1937. Uno dei pionieri di questo ramo della matematica, svolse le sue ricerche essendo dipendente (prima come chimico, poi come dirigente) della Guinness Brewery di Dublino.

casuale

$$X = \sum_{i=1}^{N} {x_i}^2$$

(ovviamente non negativa) è distribuita con una densità di probabilità data dalla

$$\frac{\mathrm{d}p}{\mathrm{d}X} = f(X;N) = K_N X^{\left(\frac{N}{2}-1\right)} e^{-\frac{X}{2}} \tag{12.1}$$

(*distribuzione del chi quadro*); la costante K_N viene fissata dalla condizione di normalizzazione, ed il parametro N prende il nome di *numero di gradi di libertà* della distribuzione.

La funzione caratteristica della X si può trovare facilmente considerando che, se la x è una variabile normale standardizzata, il suo quadrato $y = x^2$ ha una funzione caratteristica

$$\begin{aligned}
\phi_y(t) &= E(e^{ity}) \\
&= E\left(e^{itx^2}\right) \\
&= \int_{-\infty}^{+\infty} e^{itx^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \,\mathrm{d}x \\
&= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}(1-2it)} \,\mathrm{d}x \\
&= \frac{1}{\sqrt{1-2it}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \,\mathrm{d}u \\
&= (1-2it)^{-\frac{1}{2}}
\end{aligned}$$

(si è eseguita la sostituzione di variabile $u = x\sqrt{1-2it}$; l'integrale definito è quello di una distribuzione normale $N(u;0,1)$ e vale dunque 1). Di conseguenza, applicando l'equazione (6.11), la funzione caratteristica della X vale

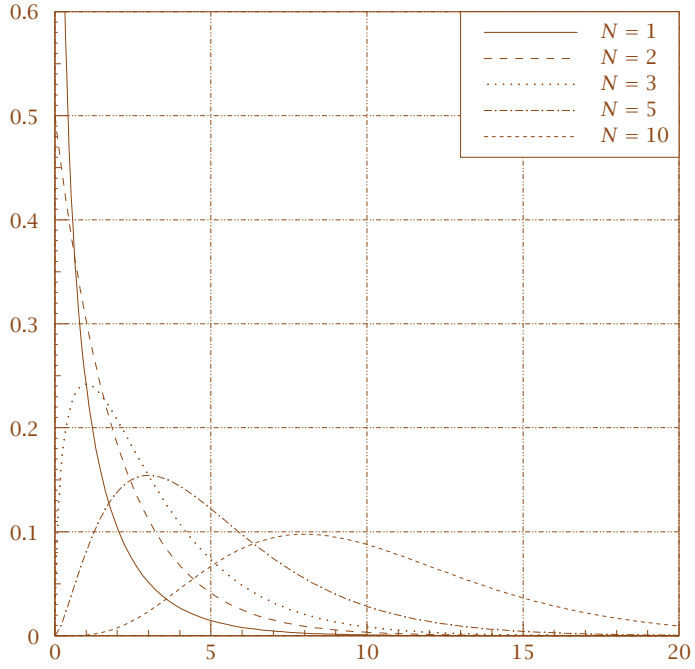
$$\phi_X(t) = (1-2it)^{-\frac{N}{2}} \ . \tag{12.2}$$

Per dimostrare che la funzione di frequenza della X è effettivamente la (12.1), si parte poi dall'espressione (12.2) della funzione caratteristica e le si applica la trasformazione inversa di Fourier già definita nella (6.10).

Con simili passaggi si potrebbe ricavare la funzione generatrice dei momenti, che vale

$$M_X(t) = (1-2t)^{-\frac{N}{2}}$$

FIGURA 12a - La distribuzione del χ^2 per alcuni valori del parametro N.



e, da queste, si ottiene infine che il valore medio e la varianza di una variabile casuale distribuita come il χ^2 a N gradi di libertà sono

$$E(X) = N \qquad \text{e} \qquad \mathrm{Var}(X) = 2N$$

mentre i coefficienti di asimmetria e di curtosi valgono

$$\gamma_1 = 2\sqrt{\frac{2}{N}} \qquad \text{e} \qquad \gamma_2 = \frac{12}{N} \ .$$

La distribuzione del χ^2 tende asintoticamente ad una distribuzione normale con la stessa media N e la stessa varianza $2N$; infatti la funzione caratteristica della variabile standardizzata

$$y = \frac{X-N}{\sqrt{2N}} = \frac{X}{\sqrt{2N}} - \frac{N}{\sqrt{2N}}$$

vale, ricordando la (6.17),

$$\phi_y(t) = e^{-\frac{iNt}{\sqrt{2N}}}\left[1 - \frac{2it}{\sqrt{2N}}\right]^{-\frac{N}{2}} \ .$$

Passando ai logaritmi naturali,

$$\ln\phi_y(t) = -\frac{iNt}{\sqrt{2N}} - \frac{N}{2}\ln\left(1 - \frac{2it}{\sqrt{2N}}\right)$$

e, sviluppando in serie di McLaurin il logaritmo,

$$\begin{aligned}\ln\phi_y(t) &= -\frac{iNt}{\sqrt{2N}} - \frac{N}{2}\left[-\frac{2it}{\sqrt{2N}} - \frac{1}{2}\left(\frac{2it}{\sqrt{2N}}\right)^2 + \mathcal{O}\left(N^{-\frac{3}{2}}\right)\right] \\ &= -\frac{t^2}{2} + \mathcal{O}\left(N^{-\frac{1}{2}}\right)\end{aligned}$$

da cui

$$\lim_{N\to\infty}\phi_y(t) = e^{-\frac{t^2}{2}}$$

che è appunto la funzione caratteristica di una distribuzione normale standardizzata.

In definitiva:

- Quando N assume valori sufficientemente grandi, la distribuzione del χ^2 è ben approssimata da una distribuzione normale avente la stessa media N e la stessa varianza $2N$; tale approssimazione si può ritenere in pratica già buona quando N è superiore a 30.

- Inoltre si potrebbe analogamente dimostrare che la variabile casuale $\sqrt{2X}$, anche per valori relativamente piccoli di N, ha una distribuzione che è assai bene approssimata da una funzione normale con media $\sqrt{2N-1}$ e varianza 1; l'approssimazione è già buona per $N \gtrsim 8$.

Dalla definizione (o dalla funzione caratteristica (12.2)) discende immediatamente la cosiddetta *regola di somma del* χ^2: ossia che, se X ed Y sono due variabili casuali statisticamente indipendenti entrambe distribuite come il χ^2, con N ed M gradi di libertà rispettivamente, la loro somma $Z = X + Y$ è una variabile casuale ancora distribuita come il χ^2; però con $N + M$ gradi di libertà.

Ovviamente, se le x_i (con $i = 1, \ldots, N$) sono N variabili casuali statisticamente indipendenti tra loro e provenienti da una stessa distribuzione normale con media μ e varianza σ^2, discende da quanto detto che la nuova variabile casuale

$$X' = \sum_{i=1}^{N}\left(\frac{x_i - \mu}{\sigma}\right)^2$$

è distribuita come il χ^2 a N gradi di libertà. Indichiamo ora, al solito, con $\bar{x}$ la media aritmetica delle x_i: vogliamo dimostrare che la variabile casuale

$$X'' = \sum_{i=1}^{N}\left(\frac{x_i - \bar{x}}{\sigma}\right)^2$$

è distribuita *ancora come il* χ^2*, ma con* $N-1$ *gradi di libertà.*

A questo scopo facciamo dapprima alcune considerazioni, indipendenti dalle ipotesi prima fatte sulle x_i e che risultano quindi valide per variabili casuali qualunque: supponiamo di definire N nuove variabili y_i come generiche combinazioni lineari delle x_j, con coefficienti che indicheremo col simbolo A_{ij}; in modo insomma che risulti

$$y_i = \sum_{j=1}^{N} A_{ij}\, x_j \ .$$

La somma dei quadrati delle y_i è data da

$$\sum_{i=1}^{N} {y_i}^2 = \sum_{i=1}^{N}\left(\sum_{j=1}^{N} A_{ij}\, x_j\right)\left(\sum_{k=1}^{N} A_{ik}\, x_k\right) = \sum_{jk} x_j\, x_k \sum_i A_{ij}\, A_{ik} \ ;$$

è possibile che questa somma risulti uguale alla somma dei quadrati delle x_i *qualunque* sia il valore di queste ultime? Ovviamente questo avviene se e

solo se vale la

$$\sum_i A_{ij} A_{ik} = \delta_{jk} = \begin{cases} 0 & \text{per } j \neq k \\ 1 & \text{per } j = k \end{cases} \tag{12.3}$$

(il simbolo δ_{jk}, che assume il valore 1 quando gli indici sono uguali e 0 quando sono invece diversi, si chiama *simbolo di Kronecker* o *delta di Kronecker*).

Consideriamo gli A_{ij} come gli elementi di una matrice quadrata $\boldsymbol{A}$ di ordine N; gli x_j e le y_i si possono invece considerare come le componenti di due *vettori* $\boldsymbol{X}$ ed $\boldsymbol{Y}$ definiti in uno spazio N-dimensionale — ossia come gli elementi di due matrici rettangolari con N righe ed 1 colonna.

La trasformazione che muta $\boldsymbol{X}$ in $\boldsymbol{Y}$ si può scrivere, in forma matriciale, come $\boldsymbol{Y} = \boldsymbol{AX}$; la somma dei quadrati delle x_j o delle y_i altro non è se non il prodotto scalare, di $\boldsymbol{X}$ ed $\boldsymbol{Y}$ rispettivamente, per loro stessi: ovverosia la loro *norma*, il quadrato della loro lunghezza nello spazio a N dimensioni. Quella che abbiamo ricavato adesso è la condizione perché una *trasformazione lineare* applicata ad un vettore ne conservi la lunghezza: occorre e basta che la matrice $\boldsymbol{A}$ sia *ortogonale*. Infatti la (12.3) si può scrivere

$$\tilde{\boldsymbol{A}}\boldsymbol{A} = \boldsymbol{1} \qquad \text{ossia} \qquad \tilde{\boldsymbol{A}} = \boldsymbol{A}^{-1}$$

($\tilde{\boldsymbol{A}}$ è la matrice trasposta di $\boldsymbol{A}$, di elementi $\tilde{\boldsymbol{A}}_{ij} = A_{ji}$; $\boldsymbol{1}$ è la matrice unità, di elementi $\boldsymbol{1}_{ij} = \delta_{ij}$; $\boldsymbol{A}^{-1}$ è la matrice inversa di $\boldsymbol{A}$; ed una matrice per cui $\tilde{\boldsymbol{A}} = \boldsymbol{A}^{-1}$ si dice, appunto, ortogonale).

Consideriamo adesso una trasformazione lineare definita dalle seguenti relazioni:

$$\begin{cases} y_1 = \dfrac{1}{\sqrt{N}}(x_1 + x_2 + \cdots + x_N) \\ y_2 = \dfrac{1}{\sqrt{2}}(x_1 - x_2) \\ y_3 = \dfrac{1}{\sqrt{6}}(x_1 + x_2 - 2x_3) \\ \cdots \\ y_N = \dfrac{1}{\sqrt{N(N-1)}}\left[x_1 + x_2 + \cdots + x_{N-1} - (N-1)x_N\right] \end{cases} \tag{12.4}$$

e per la quale la matrice di trasformazione abbia, insomma, elementi A_{ij}

definiti come

$$A_{ij} \equiv \begin{cases} i = 1: & \dfrac{1}{\sqrt{N}} \\ i > 1: & \begin{cases} j < i: & \dfrac{1}{\sqrt{i(i-1)}} \\ j = i: & -\dfrac{i-1}{\sqrt{i(i-1)}} \\ j > i: & 0 \end{cases} \end{cases}$$

Non è difficile controllare che la matrice $\boldsymbol{A}$ è ortogonale; inoltre la prima riga è stata scelta in modo tale che

$$y_1 = \sum_{i=1}^{N} \frac{1}{\sqrt{N}} x_i = \frac{1}{\sqrt{N}} \cdot N\bar{x} = \sqrt{N}\,\bar{x}$$

e quindi

$$\sum_{i=1}^{N} x_i^2 = \sum_{i=1}^{N} y_i^2 = N\bar{x}^2 + \sum_{i=2}^{N} y_i^2 \ .$$

Inoltre risulta (per $i > 1$)

$$\sum_{j=1}^{N} A_{ij} = \sum_{j=1}^{i-1} \frac{1}{\sqrt{i(i-1)}} - \frac{i-1}{\sqrt{i(i-1)}} = 0 \tag{12.5}$$

e, per ogni i,

$$\sum_{j=1}^{N} A_{ij}^2 = \left(\boldsymbol{A}\tilde{\boldsymbol{A}}\right)_{ii} = \delta_{ii} = 1 \ . \tag{12.6}$$

Tornando al nostro problema, supponiamo ora che tutte le x_j siano variabili aventi distribuzione normale; che abbiano tutte valore medio μ e varianza σ^2; ed inoltre che siano tra loro tutte statisticamente indipendenti. Una qualsiasi loro combinazione lineare, quindi anche ognuna delle y_i legate alle x_j da quella particolare matrice di trasformazione (12.4) che abbiamo prima definita, è anch'essa distribuita secondo la legge normale; inoltre risulta

$$\frac{1}{\sigma^2}\sum_{i=1}^{N}(x_i-\bar{x})^2 = \frac{1}{\sigma^2}\left(\sum_{i=1}^{N}{x_i}^2 - N\bar{x}^2\right)$$
$$= \frac{1}{\sigma^2}\left(N\bar{x}^2 + \sum_{i=2}^{N}{y_i}^2 - N\bar{x}^2\right)$$
$$= \sum_{i=2}^{N}\frac{{y_i}^2}{\sigma^2} \ .$$

Applicando alle $y_i = \sum_j A_{ij}x_j$ le formule per la media e la varianza delle combinazioni lineari di variabili casuali statisticamente indipendenti già ricavate nel capitolo 5, si trova facilmente (tenendo presenti la (12.5) e la (12.6)) che la varianza di ognuna di esse è ancora σ^2; e che, per $i \neq 1$, il loro valore medio è 0. Di conseguenza, per $i \geq 2$ le y_i/σ sono variabili casuali normali aventi media 0 e varianza 1: e questo implica che

$$X'' = \sum_{i=1}^{N}\left(\frac{x_i-\bar{x}}{\sigma}\right)^2 \tag{12.7}$$

sia effettivamente distribuita come il χ^2 a $N-1$ gradi di libertà.

È interessante confrontare questo risultato con quello precedentemente ricavato, e riguardante la stessa espressione — in cui però gli scarti erano calcolati rispetto alla media della popolazione μ. Nel primo caso la distribuzione era ancora quella del χ^2, ma con N gradi di libertà: riferendoci invece alla media aritmetica del campione, i gradi di libertà diminuiscono di una unità. Questo è conseguenza di una legge generale, secondo la quale il numero di gradi di libertà da associare a variabili che seguono la distribuzione del χ^2 è dato dal numero di contributi *indipendenti*: ovvero il numero di termini con distribuzione normale sommati in quadratura (qui N, uno per ogni determinazione x_i) diminuito del numero di parametri che compaiono nella formula e che sono stati stimati dai dati stessi (qui uno: appunto la media della popolazione, stimata usando la media aritmetica delle misure).

Un'ultima notevole conseguenza del fatto che la variabile casuale X'' definita dalla (12.7) sia distribuita come il χ^2 a $N-1$ gradi di libertà è la seguente: la stima della varianza della popolazione ottenuta dal campione, s^2, vale

$$s^2 = X''\,\frac{\sigma^2}{N-1} \tag{12.8}$$

e, essendo proporzionale a X'', è anch'essa distribuita come il χ^2 a $N-1$ gradi di libertà; quindi la sua densità di probabilità è data dalla (12.1) e dipende

solamente da N; non dipende, in particolare, dalla media del campione $\bar{x}$. Quindi:

> *Il valore medio $\bar{x}$ e la varianza campionaria s^2, calcolati su valori estratti a caso da una stessa popolazione normale, sono due variabili casuali* ***statisticamente indipendenti*** *tra loro.*

Questo risulta anche intuitivamente comprensibile; se infatti ci è noto che un certo campione di dati ha una dispersione più o meno grande, questo non deve alterare la probabilità che il suo valore medio abbia un valore piuttosto che un altro; né, viceversa, il fatto che il campione sia centrato attorno ad un certo valore deve permetterci di prevedere in qualche modo la sua dispersione.

12.2 Verifiche basate sulla distribuzione del χ^2

12.2.1 Compatibilità dei dati con una distribuzione

Supponiamo di avere dei dati raccolti in un istogramma, e di voler verificare l'ipotesi che i dati provengano da una certa distribuzione; ad esempio, dalla distribuzione normale. Ora, per una misura, la probabilità p_i di cadere nell'intervallo i-esimo (di ampiezza prefissata Δx e corrispondente alla generica classe di frequenza usata per la realizzazione dell'istogramma) è data dal valore medio della funzione densità di probabilità nell'intervallo stesso moltiplicato per Δx.

Il numero di misure effettivamente ottenute in una classe di frequenza su N prove deve obbedire poi alla distribuzione binomiale: il loro valore medio è quindi Np_i, e la loro varianza $N\,p_i\,(1-p_i)$; quest'ultimo termine si può approssimare ancora con Np_i se si ammette che le classi di frequenza siano sufficientemente ristrette da poter trascurare i termini in ${p_i}^2$ rispetto a quelli in p_i (cioè se $p_i \ll 1$).

In questo caso il numero di misure in ciascuna classe segue approssimativamente la distribuzione di Poisson; questa è infatti la funzione di frequenza che governa il presentarsi, su un grande numero di osservazioni, di eventi aventi probabilità trascurabile di verificarsi singolarmente in ognuna: distribuzione nella quale l'errore quadratico medio è effettivamente dato dalla radice quadrata del valore medio, $\sigma = \sqrt{N\,p_i\,(1-p_i)} \simeq \sqrt{Np_i}$.

Nei limiti in cui il numero di misure attese in una classe è sufficientemente elevato da poter confondere la relativa funzione di distribuzione con

la funzione normale, la quantità

$$X = \sum_{i=1}^{M} \frac{(n_i - Np_i)^2}{Np_i} = \sum_{i=1}^{M} \frac{(O_i - A_i)^2}{A_i} \tag{12.9}$$

cioè la somma, su tutte le classi di frequenza (il cui numero abbiamo supposto sia M), del quadrato della differenza tra il numero di misure ivi *attese* ($A_i = Np_i$) ed ivi *effettivamente osservate* ($O_i = n_i$), diviso per la varianza del numero di misure attese (approssimata da $Np_i = A_i$), ha *approssimativamente* la distribuzione del χ^2, con $M - 1$ gradi di libertà; il motivo di quest'ultima affermazione è che esiste un vincolo sulle O_i, quello di avere per somma il numero totale di misure effettuate N (che viene usato nella formula (12.9), mediante la quale abbiamo definito X, per calcolare il numero A_i di misure attese in ogni intervallo).

La condizione enunciata si può in pratica supporre verificata se le A_i in ogni intervallo sono almeno pari a 5; o, meglio, se il numero di classi di frequenza in cui ci si aspetta un numero di misure minore di 5 è trascurabile rispetto al totale (meno del 10%). In realtà, se le classi di frequenza si possono scegliere arbitrariamente, la cosa migliore consiste nel definirle di ampiezze differenti: in modo tale che quegli intervalli dove cadono poche misure vengano riuniti assieme in un'unica classe più ampia, ove n_i valga almeno 5 (ma nemmeno troppo ampia, per soddisfare al vincolo di avere ${p_i}^2 \ll p_i$; in genere si cerca di riunire assieme più classi in modo da avere degli $n_i \sim 5 \div 10$).

Tornando al problema iniziale, per la verifica dell'ipotesi statistica che i dati vengano dalla distribuzione usata per il calcolo delle A_i basta:

- fissare arbitrariamente un livello di probabilità che rappresenti il confine tra eventi ammissibili nell'ipotesi della pura casualità ed eventi invece tanto improbabili da far supporre che il loro verificarsi sia dovuto non a fluttuazioni statistiche, ma al non essere verificate le ipotesi fatte in partenza (il provenire i dati dalla distribuzione nota a priori): ad esempio il 95% o il 99%.

- Cercare nelle apposite tabelle[2] il valore di taglio corrispondente alla coda superiore della distribuzione del χ^2 ad $M - 1$ gradi di libertà avente area pari al livello di confidenza desiderato; ossia quell'ascissa ξ che lascia alla propria sinistra, sotto la curva della distribuzione del χ^2 ad $M - 1$ gradi di libertà, un'area pari a tale valore.

- Calcolare X; ed infine rigettare l'ipotesi (al livello di confidenza prescelto) perché incompatibile con i dati raccolti, se X risultasse superiore a ξ (o, altrimenti, considerare l'ipotesi compatibile con i dati al livello di confidenza prescelto e quindi accettarla).

[2] Alcuni valori numerici di questo tipo sono tabulati nell'appendice G. È bene anche ricordare che quando il numero di gradi di libertà N è superiore a 30 si può far riferimento alla distribuzione normale con media N ed errore quadratico medio $\sqrt{2N}$; e che, già per piccoli N, $\sqrt{2\chi^2}$ è approssimativamente normale con media $\sqrt{2N-1}$ e varianza 1.

Quanto detto a proposito della particolare distribuzione del χ^2 da usare per il la verifica della nostra ipotesi, però, è valido solo se le caratteristiche della distribuzione teorica con cui confrontare i nostri dati sono note a priori; se, invece, R parametri da cui essa dipende fossero stati stimati a partire dai dati, il numero di gradi di libertà sarebbe inferiore e pari ad $M - R - 1$.

Così se le p_i sono state ricavate integrando sulle classi di frequenza una distribuzione normale la cui media e la cui varianza siano state a loro volta ottenute dal campione istogrammato, il numero di gradi di libertà, essendo $R = 2$, sarebbe pari a $M - 3$.

Per dare un'idea dei valori del χ^2 che corrispondono al rigetto di una ipotesi (ad un certo livello di confidenza), e senza ricorrere alle tabelle numeriche, nella figura 12b sono riportati in grafico i valori P dell'integrale da x a $+\infty$ della funzione di frequenza del χ^2 (ovvero il complemento ad uno della funzione di distribuzione), per alcuni valori del parametro N.

Le curve della figura 12c permettono invece di identificare (per differenti scelte del livello di confidenza ε) i corrispondenti valori di taglio del χ^2 *ridotto* — ovvero del rapporto χ^2/N tra esso ed il numero di gradi di libertà N. Insomma, ogni punto di queste curve al di sopra di un'ascissa (intera) N ha come ordinata un numero X/N tale che l'integrale da X a $+\infty$ della funzione di frequenza del χ^2 ad N gradi di libertà sia uguale ad ε.

12.2.2 Il metodo del minimo χ^2

Supponiamo di sapere a priori che i nostri dati istogrammati debbano seguire una data distribuzione, ma che essa dipenda da R parametri incogniti che dobbiamo stimare a partire dai dati stessi; visto che l'accordo tra i dati e la distribuzione è dato dalla X definita nella (12.9), ed è tanto migliore quanto più il valore ottenuto per essa è basso, un metodo plausibile di stima potrebbe essere quello di trovare per quali valori dei parametri stessi la X è minima (*metodo del minimo* χ^2).

Indicando con α_k ($k = 1, \ldots, R$) i parametri da stimare, ognuna delle p_i sarà esprimibile in funzione delle α_k; ed imponendo che le derivate prime della X rispetto ad ognuna delle α_k siano tutte nulle contemporaneamente,

FIGURA 12b - L'integrale da x a $+\infty$ della funzione di frequenza del χ^2, per alcuni valori del parametro N.

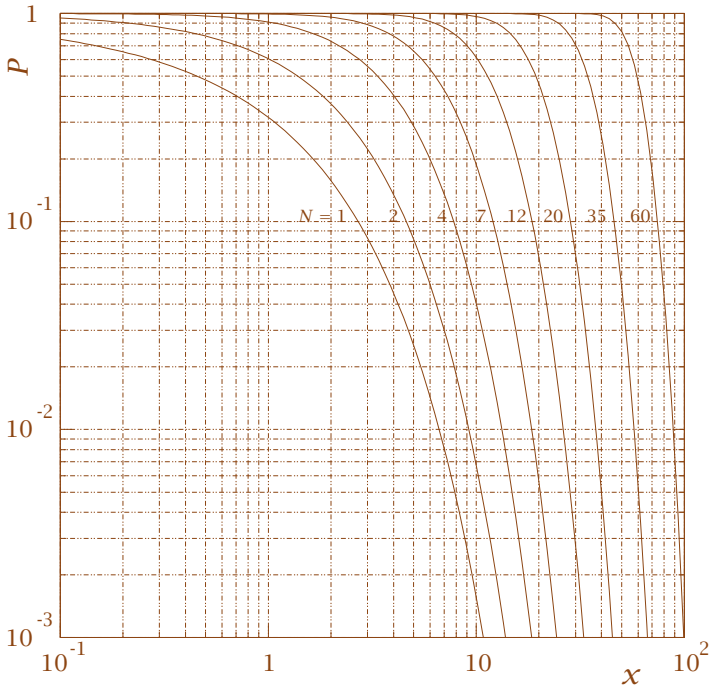
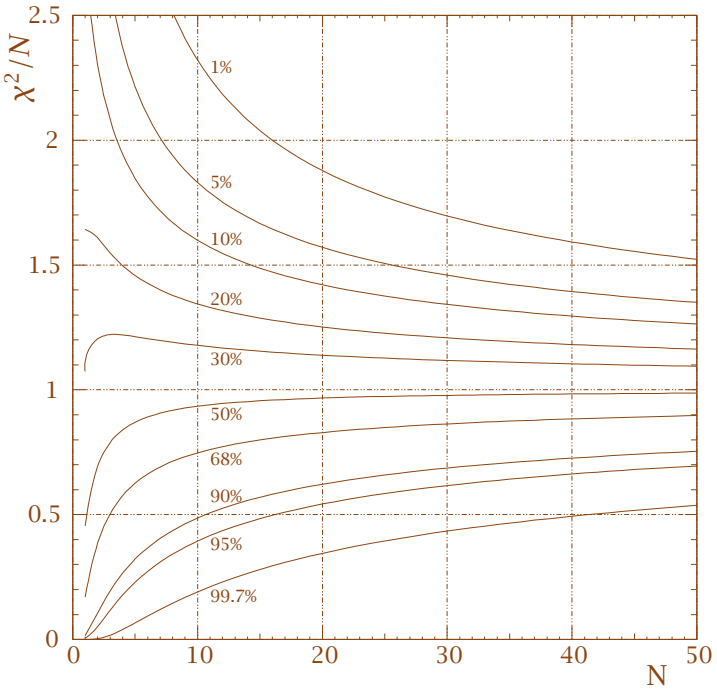


FIGURA 12c - I valori del χ^2 ridotto (χ^2/N) che corrispondono, per differenti gradi di libertà N, ad un certo livello di confidenza.



otteniamo

$$\frac{\partial X}{\partial \alpha_k} = \sum_{i=1}^{M} \frac{-2\,(n_i - Np_i)\,N^2 p_i - N\,(n_i - Np_i)^2}{N^2 {p_i}^2} \frac{\partial p_i}{\partial \alpha_k} = 0 \ ,$$

ossia

$$-\frac{1}{2}\frac{\partial X}{\partial \alpha_k} = \sum_{i=1}^{M} \left[\frac{n_i - Np_i}{p_i} + \frac{(n_i - Np_i)^2}{2N{p_i}^2} \right] \frac{\partial p_i}{\partial \alpha_k} = 0 \ . \tag{12.10}$$

L'insieme delle (12.10) costituisce un sistema di R equazioni, nelle R incognite α_k, che ci permetterà di stimarne i valori (salvo poi, nel caso il sistema delle (12.10) abbia più di una soluzione, controllare quali di esse corrispondono in effetti ad un minimo e quale tra queste ultime corrisponde al minimo assoluto); le condizioni sotto le quali il metodo è applicabile sono quelle già enunciate in precedenza[3], ossia ${p_i}^2 \ll p_i$ e $n_i \gtrsim 5$.

In genere però si preferisce servirsi, in luogo delle equazioni (12.10), di una forma semplificata, ottenuta trascurando il secondo termine nella parentesi quadra: che, si può dimostrare, è molto inferiore al primo per grandi N (infatti il rapporto tra i due termini vale

$$\frac{(n_i - Np_i)^2}{2N{p_i}^2} \frac{p_i}{n_i - Np_i} = \frac{n_i - Np_i}{2Np_i} = \frac{1}{2p_i}\left(\frac{n_i}{N} - p_i\right)$$

e converge ovviamente a zero all'aumentare di N); e risolvere, insomma, il sistema delle

$$\sum_{i=1}^{M} \left(\frac{n_i - Np_i}{p_i}\right) \frac{\partial p_i}{\partial \alpha_k} = 0 \tag{12.11}$$

(metodo *semplificato* del minimo χ^2).

Si può dimostrare che le soluzioni $\bar{\alpha}_k$ del sistema delle (12.11) tendono stocasticamente ai valori veri α_k^* (in assenza di errori sistematici) al crescere di N; inoltre il valore di X calcolato in corrispondenza dei valori ricavati $\bar{\alpha}_k$ dà, se rapportato alla distribuzione del χ^2 con $M - R - 1$ gradi di libertà, una misura della bontà della soluzione stessa.

Ora, le equazioni (12.11) si possono scrivere anche

$$\sum_{i=1}^{M} \left(\frac{n_i - Np_i}{p_i}\right) \frac{\partial p_i}{\partial \alpha_k} = \sum_{i=1}^{M} \frac{n_i}{p_i} \frac{\partial p_i}{\partial \alpha_k} - N \sum_{i=1}^{M} \frac{\partial p_i}{\partial \alpha_k}$$

[3]Se la prima di esse non si può ritenere accettabile, delle equazioni ancora valide ma più complesse si possono ottenere dalla (12.9) sostituendo $Np_i(1 - p_i)$ al posto di Np_i nel denominatore.

e si possono ulteriormente semplificare, visto che l'ultimo termine si annulla, essendo

$$\sum_{i=1}^{M} \frac{\partial p_i}{\partial \alpha_k} = \frac{\partial}{\partial \alpha_k} \sum_{i=1}^{M} p_i = \frac{\partial}{\partial \alpha_k} 1 \equiv 0$$

se si fa l'ulteriore ipotesi che l'intervallo dei valori indagati copra, anche approssimativamente, tutti quelli in pratica permessi; per cui il sistema di equazioni da risolvere è in questo caso quello delle

$$\sum_{i=1}^{M} \frac{n_i}{p_i} \frac{\partial p_i}{\partial \alpha_k} = 0 \ . \tag{12.12}$$

Per la stima di parametri incogniti a partire da dati misurati abbiamo già affermato che teoricamente è da preferire il metodo della massima verosimiglianza, le cui soluzioni sono quelle affette, come sappiamo, dal minimo errore casuale (almeno asintoticamente); in questo caso particolare (dati in istogramma), come lo si dovrebbe applicare? Se le misure sono indipendenti, la probabilità di avere n_i eventi nella generica classe di frequenza è data da ${p_i}^{n_i}$; la funzione di verosimiglianza[4] da

$$\mathcal{L}(\alpha_1, \ldots, \alpha_R) = \prod_{i=1}^{M} {p_i}^{n_i} \tag{12.13}$$

ed il suo logaritmo da

$$\ln \mathcal{L} = \sum_{i=1}^{M} (n_i \cdot \ln p_i) \ .$$

La soluzione di massima verosimiglianza (e quindi di minima varianza) si trova cercando il massimo di $\ln \mathcal{L}$: e risolvendo quindi il sistema delle

$$\frac{\partial}{\partial \alpha_k} \ln \mathcal{L} = \sum_{i=1}^{M} n_i \frac{1}{p_i} \frac{\partial p_i}{\partial \alpha_k} = 0 \ ;$$

in questo caso, vista l'equazione (12.12) in precedenza ricavata, i due metodi (della massima verosimiglianza e del minimo χ^2 semplificato) conducono dunque *alla stessa soluzione*.

[4]Per essere precisi, la probabilità che n_1 misure si trovino nella prima classe di frequenza, n_2 nella seconda e così via, è dato dalla espressione (12.13) moltiplicata per il numero di modi differenti in cui N oggetti possono essere suddivisi in M gruppi composti da $n_1, n_2, \ldots, n_M$ oggetti rispettivamente (numero delle *partizioni ordinate*); questo vale, come mostrato nel paragrafo A.7, $N!/(n_1!\,n_2! \cdots n_M!)$, e rappresenta un fattore costante che non incide nella ricerca del massimo della (12.13).

12.2.3 Test di omogeneità per dati raggruppati

Supponiamo di avere a disposizione Q campioni di dati, indipendenti l'uno dall'altro e composti da $n_1, n_2, \ldots, n_Q$ elementi rispettivamente; e, all'interno di ognuno di tali campioni, i dati siano suddivisi nei medesimi P gruppi: indichiamo infine col simbolo ν_{ij} il numero di dati appartenenti al gruppo i-esimo all'interno del campione j-esimo.

Per fare un esempio, i campioni si potrebbero riferire alle regioni italiane e i gruppi al livello di istruzione (licenza elementare, media, superiore, laurea): così che i ν_{ij} rappresentino il numero di persone, per ogni livello di istruzione, residenti in ogni data regione; oppure (e questo è un caso che si presenta frequentemente nelle analisi fisiche) si abbiano vari istogrammi all'interno di ognuno dei quali i dati siano stati raggruppati secondo le medesime classi di frequenza: allora i ν_{ij} saranno il numero di osservazioni che cadono in una determinata classe in ogni istogramma.

Il problema che ci poniamo è quello di verificare l'ipotesi che tutti i campioni provengano dalla stessa popolazione e siano perciò compatibili tra loro (*test di omogeneità*). Indichiamo con il simbolo N il numero totale di dati a disposizione; e con m_i (con $i = 1, \ldots, P$) il numero totale di dati che cadono nell'i-esimo gruppo in tutti i campioni a disposizione.

TABELLA 12.1 - Un esempio delle cosiddette *tabelle delle contingenze*.

	Campioni					
Gruppi	ν_{11}	ν_{12}	ν_{13}	$\cdots$	ν_{1Q}	m_1
	ν_{21}	ν_{22}	$\cdots$	$\cdots$	ν_{2Q}	m_2
	ν_{31}	$\cdots$	$\cdots$	$\cdots$	$\cdots$	m_3
	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
	ν_{P1}	ν_{P2}	$\cdots$	$\cdots$	ν_{PQ}	m_P
	n_1	n_2	n_3	$\cdots$	n_Q	N

È consuetudine che dati di questo genere siano rappresentati in una tabella del tipo della 12.1, che si chiama *tabella delle contingenze*; e risulta

ovviamente

$$n_j = \sum_{i=1}^{P} \nu_{ij} \qquad (j = 1, 2, \ldots, Q) \ ;$$

$$m_i = \sum_{j=1}^{Q} \nu_{ij} \qquad (i = 1, 2, \ldots, P) \ ;$$

$$N = \sum_{j=1}^{Q} n_j = \sum_{i=1}^{P} m_i = \sum_{i,j} \nu_{ij} \ .$$

Vogliamo ora dimostrare che la variabile casuale

$$X = N \left[\sum_{i,j} \frac{(\nu_{ij})^2}{m_i \, n_j} - 1 \right] \tag{12.14}$$

è distribuita come il χ^2 a $(P-1)(Q-1)$ gradi di libertà: a questo scopo supponiamo innanzi tutto sia valida l'ipotesi che i dati provengano tutti dalla medesima popolazione, ed indichiamo con i simboli p_i e q_j le probabilità che un componente di tale popolazione scelto a caso cada rispettivamente nel gruppo i-esimo o nel campione j-esimo; e sappiamo inoltre che (ammessa però vera l'ipotesi che *tutti* i campioni provengano dalla stessa distribuzione) questi due eventi sono statisticamente indipendenti: per cui ognuno dei dati ha probabilità complessiva $p_i q_j$ di cadere in una delle caselle della tabella delle contingenze.

Possiamo stimare i P valori p_i a partire dai dati sperimentali: si tratta in realtà solo di $P-1$ stime *indipendenti*, perché, una volta ricavate le prime $P-1$ probabilità, l'ultima di esse risulterà univocamente determinata dalla condizione che la somma complessiva valga 1. Analogamente possiamo anche stimare i Q valori q_j dai dati sperimentali, e si tratterà in questo caso di effettuare $Q-1$ stime indipendenti.

Le stime di cui abbiamo parlato sono ovviamente

$$p_i = \frac{m_i}{N} \qquad \text{e} \qquad q_j = \frac{n_j}{N} \tag{12.15}$$

e, applicando le conclusioni del paragrafo precedente (l'equazione (12.9)), la

variabile

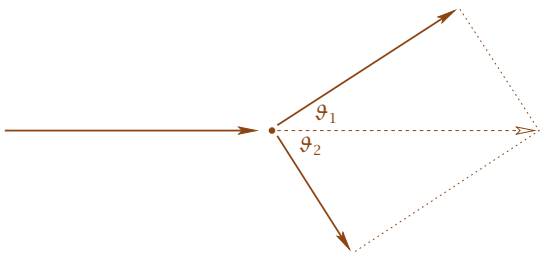
$$
\begin{aligned}
X &= \sum_{i,j} \frac{(\nu_{ij} - Np_iq_j)^2}{Np_iq_j} \\
&= \sum_{i,j} \left[\frac{(\nu_{ij})^2}{Np_iq_j} - 2\nu_{ij} + Np_iq_j \right] \\
&= \sum_{i,j} \frac{(\nu_{ij})^2}{Np_iq_j} - 2N + N \\
&= \sum_{i,j} \frac{(\nu_{ij})^2}{Np_iq_j} - N
\end{aligned}
$$

deve essere distribuita come il χ^2.

Sostituendo in quest'ultima espressione i valori (12.15) per p_i e q_j, essa si riduce alla (12.14); il numero di gradi di libertà è pari al numero di contributi sperimentali indipendenti, $PQ - 1$ (c'è il vincolo che la somma totale sia N), diminuito del numero $(P-1)+(Q-1)$ di parametri stimato sulla base dei dati: ovverosia proprio $(P-1)(Q-1)$ come anticipato.

12.2.4 Un esempio: diffusione elastica protone-protone

FIGURA 12d - Urto elastico protone-protone.



Nella figura 12d è schematicamente rappresentato un processo di urto elastico tra due particelle, una delle quali sia inizialmente ferma; dopo l'urto esse si muoveranno lungo traiettorie rettilinee ad angoli ϑ_1 e ϑ_2 rispetto alla direzione originale della particella urtante.

Gli angoli ϑ_i vengono misurati; supponendo che il processo di misura introduca errori che seguono la distribuzione normale ed abbiano una entità che (per semplificare le cose) assumiamo sia costante, nota ed indipendente dall'ampiezza dell'angolo, vogliamo verificare l'ipotesi che le due particelle coinvolte nel processo d'urto siano di massa uguale (ad esempio che siano entrambe dei protoni).

La prima cosa da fare è quella di ricavare dai dati misurati ϑ_i, che per ipotesi hanno una funzione di frequenza

$$
f(\vartheta;\vartheta^*,\sigma) \;=\; \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{\vartheta-\vartheta^*}{\sigma}\right)^2}
$$

una stima dei valori veri ϑ^*. Il logaritmo della funzione di verosimiglianza è dato da

$$
\ln \mathcal{L} \;=\; -2\ln\left(\sigma\sqrt{2\pi}\right) - \frac{1}{2}\left(\frac{\vartheta_1-\vartheta_1^*}{\sigma}\right)^2 - \frac{1}{2}\left(\frac{\vartheta_2-\vartheta_2^*}{\sigma}\right)^2 \quad ;
$$

ma le variabili ϑ_1 e ϑ_2 *non sono indipendenti*, visto che il processo deve conservare sia energia che quantità di moto. Ammessa vera l'ipotesi che le due particelle abbiano uguale massa (e restando nel limite non-relativistico), le leggi di conservazione impongono il vincolo che l'angolo tra le due particelle dopo l'urto sia di 90°(o, in radianti, $\pi/2$); usando il metodo dei moltiplicatori di Lagrange, la funzione da massimizzare è

$$
\varphi(\vartheta_1^*,\vartheta_2^*,\lambda) \;=\; -\frac{1}{2}\left(\frac{\vartheta_1-\vartheta_1^*}{\sigma}\right)^2 - \frac{1}{2}\left(\frac{\vartheta_2-\vartheta_2^*}{\sigma}\right)^2 + \lambda\left(\vartheta_1^*+\vartheta_2^*-\frac{\pi}{2}\right)
$$

e, annullando contemporaneamente le sue derivate rispetto alle tre variabili, si giunge al sistema

$$
\begin{cases}
\dfrac{\partial\varphi}{\partial\lambda} &= \vartheta_1^*+\vartheta_2^*-\dfrac{\pi}{2} &= 0 \\[2ex]
\dfrac{\partial\varphi}{\partial\vartheta_1^*} &= \dfrac{1}{\sigma^2}(\vartheta_1-\vartheta_1^*)+\lambda &= 0 \\[2ex]
\dfrac{\partial\varphi}{\partial\vartheta_2^*} &= \dfrac{1}{\sigma^2}(\vartheta_2-\vartheta_2^*)+\lambda &= 0
\end{cases}
$$

Eliminando λ dalle ultime due equazioni otteniamo

$$
\vartheta_1-\vartheta_1^* \;=\; \vartheta_2-\vartheta_2^*
$$

e, sostituendo l'espressione per ϑ_2^* ricavata dalla prima equazione,

$$
\vartheta_1-\vartheta_1^* \;=\; \vartheta_2-\left(\frac{\pi}{2}-\vartheta_1^*\right)
$$

per cui le due stime di massima verosimiglianza sono

$$\begin{cases} \hat{\vartheta}_1^* &= \vartheta_1 + \frac{1}{2}\left(\frac{\pi}{2} - \vartheta_1 - \vartheta_2\right) \\ \hat{\vartheta}_2^* &= \vartheta_2 + \frac{1}{2}\left(\frac{\pi}{2} - \vartheta_1 - \vartheta_2\right) \end{cases}$$

Ammesso che queste soluzioni siano buone stime dei valori veri, la variabile casuale

$$X = \left(\frac{\vartheta_1 - \vartheta_1^*}{\sigma}\right)^2 + \left(\frac{\vartheta_2 - \vartheta_2^*}{\sigma}\right)^2 = \frac{1}{2\sigma^2}\left(\frac{\pi}{2} - \vartheta_1 - \vartheta_2\right)^2$$

è distribuita come il χ^2 ad un grado di libertà (due contributi, un vincolo); ed il valore di X confrontato con le tabelle del χ^2 può essere usato per la verifica dell'ipotesi.

12.3 Compatibilità con un valore prefissato

Un altro caso che frequentemente si presenta è il seguente: si vuole controllare se un determinato valore numerico, a priori attribuibile alla grandezza fisica in esame, è o non è confermato dai risultati della misura; cioè se quel valore è o non è *compatibile* con i nostri risultati — più precisamente, a che livello di probabilità (o, per usare la terminologia statistica, a che *livello di confidenza*) è con essi compatibile.

Ammettiamo che gli errori di misura seguano la legge normale; sappiamo che la probabilità per il risultato di cadere in un qualunque intervallo prefissato dell'asse reale si può calcolare integrando la funzione di Gauss fra gli estremi dell'intervallo stesso. Riferiamoci per comodità alla variabile *scarto normalizzato*

$$t = \frac{x - E(x)}{\sigma}$$

che sappiamo già dal paragrafo 9.3 essere distribuita secondo una legge che è indipendente dall'entità degli errori di misura.

Se fissiamo arbitrariamente un numero positivo τ, possiamo calcolare la probabilità che si verifichi l'evento casuale consistente nell'ottenere, in una particolare misura, un valore di t che in modulo superi τ; come esempio particolare, le condizioni $|t| > 1$ o $|t| > 2$ già sappiamo che si verificano con probabilità rispettivamente del 31.73% e del 4.55%, visto che l'intervallo $-1 \le t \le 1$ corrisponde al 68.27% dell'area della curva normale, e quello $-2 \le t \le 2$ al 95.45% .

Se consideriamo poi un campione di N misure indipendenti, avente valore medio $\bar{x}$ e proveniente da questa stessa popolazione di varianza σ^2, è immediato capire come la variabile

$$t = \frac{\bar{x} - E(x)}{\frac{\sigma}{\sqrt{N}}}$$

soddisferà a queste stesse condizioni: accadrà cioè nel 31.73% dei casi che $|t|$ sia maggiore di $\tau = 1$, e nel 4.55% dei casi che $|t|$ sia superiore a $\tau = 2$.

Per converso, se fissiamo arbitrariamente un qualunque valore ammissibile P per la probabilità, possiamo calcolare in conseguenza un numero τ, tale che la probabilità di ottenere effettivamente da un particolare campione un valore dello scarto normalizzato t superiore ad esso (in modulo) sia data dal numero P. Ad esempio, fissato un valore del 5% per P, il limite per t che se ne ricava è $\tau = 1.96$: insomma

$$\int_{-1.96}^{+1.96} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt = 0.95$$

e solo nel cinque per cento dei casi si ottiene un valore di t che supera (in modulo) 1.96.

Se si fissa per convenzione un valore della probabilità che indichi il confine tra un avvenimento accettabile ed uno inaccettabile nei limiti della pura casualità, possiamo dire che l'ipotesi consistente nell'essere un certo numero ξ il valore vero della grandezza misurata sarà compatibile o incompatibile con i nostri dati a seconda che lo scarto normalizzato

$$t = \frac{\bar{x} - \xi}{\frac{\sigma}{\sqrt{N}}}$$

relativo a tale numero sia, in valore assoluto, inferiore o superiore al valore di τ che a quella probabilità corrisponde; e diremo che la compatibilità (o incompatibilità) è riferita a quel certo livello di confidenza prescelto.

La difficoltà è che tutti questi ragionamenti coinvolgono una quantità numerica (lo scarto quadratico medio) relativa *alla popolazione* e per ciò stesso in generale ignota; in tal caso, per calcolare lo scarto normalizzato relativo ad un certo valore numerico ξ non possiamo che servirci, in luogo di σ, della corrispondente stima ricavata dal campione, s:

$$t = \frac{\bar{x} - \xi}{\frac{s}{\sqrt{N}}}$$

e quindi si deve presupporre di avere un campione di dimensioni tali che questa stima si possa ritenere ragionevole, ossia sufficientemente vicina ai

corrispondenti valori relativi alla popolazione a meno di fluttuazioni casuali abbastanza poco probabili.

In generale si ammette che *almeno 30 dati* siano necessari perché questo avvenga: in corrispondenza a tale dimensione del campione, l'errore della media è circa 5.5 volte inferiore a quello dei dati; e l'errore relativo di s è approssimativamente del 13%.

Bisogna anche porre attenzione alla esatta natura dell'ipotesi che si intende verificare. Per un valore limite di $\tau = 1.96$ abbiamo visto che il 95% dell'area della curva normale è compreso tra $-\tau$ e $+\tau$: superiormente a $+\tau$ si trova il 2.5% di tale area; ed anche inferiormente a $-\tau$ se ne trova un'altra porzione pari al 2.5%.

TABELLA 12.2 - Alcuni valori della probabilità P e dei corrispondenti limiti τ sullo scarto normalizzato, per verifiche two-tailed (τ_2) o one-tailed (τ_1).

P (%)	τ_2	τ_1
10.0	1.64485	1.28155
5.0	1.95996	1.64485
2.0	2.32635	2.05375
1.0	2.57583	2.32635
0.5	2.81297	2.57583
0.2	3.09023	2.87816
0.1	3.29053	3.09023

TABELLA 12.3 - I valori della probabilità per verifiche two-tailed (P_2) ed one-tailed (P_1) che corrispondono a valori prefissati dello scarto normalizzato τ.

τ	P_2 (%)	P_1 (%)
0.5	61.708	30.854
1.0	31.731	15.866
1.5	13.361	6.681
2.0	4.550	2.275
2.5	1.242	0.621
3.0	0.270	0.135

Se l'ipotesi da verificare riguarda l'*essere differenti tra loro* due entità (il

presupposto valore vero della grandezza misurata e la media aritmetica dei nostri dati, nell'esempio precedente) quel valore di τ corrisponde in effetti ad una verifica relativa ad un livello di confidenza del 5% (usando il termine inglese, stiamo effettuando un *two-tailed test*); ma se l'ipotesi riguarda l'essere un valore numerico *superiore* (od *inferiore*) alla nostra media aritmetica (ad esempio, i dati misurati potrebbero essere relativi al rendimento di una macchina, e si vuole verificare l'ipotesi che tale rendimento misurato sia superiore ad un valore prefissato), allora un limite $\tau = 1.96$ corrisponde in effetti ad un livello di confidenza del 2.5% (*one-tailed test*): nell'esempio fatto, soltanto l'intervallo $[-\infty, -\tau]$ deve essere preso in considerazione per il calcolo della probabilità. Alcuni limiti relativi a diversi livelli di confidenza si possono trovare nelle tabelle 12.2 e 12.3; altri si possono facilmente ricavare dalle tabelle dell'appendice G.

12.4 I piccoli campioni e la distribuzione di Student

Cosa si può fare riguardo alla verifica di ipotesi statistiche come quella (considerata nel paragrafo precedente) della compatibilità del risultato delle misure con un valore noto a priori, quando si abbiano a disposizione solamente piccoli campioni? Ci riferiamo, più esattamente, a campioni costituiti da un numero di dati così esiguo da farci ritenere che non si possa ottenere da essi con ragionevole probabilità una buona stima delle varianze delle rispettive popolazioni (sempre però supposte normali).

Sia X una variabile casuale distribuita come il χ^2 ad N gradi di libertà, ed u una seconda variabile casuale, indipendente dalla prima, e avente distribuzione normale standardizzata $N(u;0,1)$; consideriamo la nuova variabile casuale t definita attraverso la

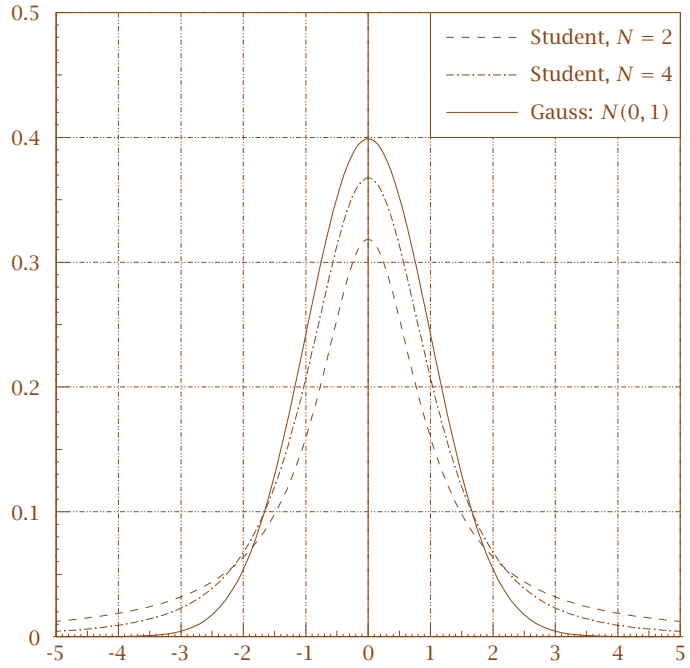
$$t = \frac{u}{\sqrt{\frac{X}{N}}} \quad . \tag{12.16}$$

Si può dimostrare che la funzione densità di probabilità relativa alla variabile casuale t è data dalla

$$f(t;N) = \frac{T_N}{\left(1 + \frac{t^2}{N}\right)^{\frac{N+1}{2}}}$$

che si chiama *distribuzione di Student ad N gradi di libertà.*

FIGURA 12e - La distribuzione di Student per $N = 2$ ed $N = 4$, confrontata con la funzione normale.



Il coefficiente T_N è una costante che viene fissata dalla condizione di normalizzazione; se N viene poi fatto tendere all'infinito il denominatore della funzione (come si potrebbe facilmente provare partendo dal limite notevole (9.9)) tende a $e^{t^2/2}$, e dunque la distribuzione di Student tende alla distribuzione normale (con media 0 e varianza 1). Anche la forma della funzione di Student ricorda molto quella della funzione di Gauss, come appare evidente dalla figura 12e; soltanto, rispetto a dati che seguano la distribuzione normale, valori elevati dello scarto sono relativamente più probabili[5].

La distribuzione di Student è simmetrica, quindi tutti i momenti di ordine dispari (compreso il valore medio λ_1) sono nulli; mentre la varianza della distribuzione è

$$\mathrm{Var}(t) = \frac{N}{N-2}$$

(se $N > 2$); ed il coefficiente di curtosi vale

$$\gamma_2 = \frac{6}{N-4}$$

(se $N > 4$).

Indicando con $\bar{x}$ la media aritmetica di un campione di dimensione N, estratto a caso da una popolazione normale avente valore medio $E(x)$ e varianza σ^2; e con s la stima della deviazione standard della popolazione ottenuta dal campione stesso, cioè

$$s^2 = \frac{\sum_i (x_i - \bar{x})^2}{N-1}$$

sappiamo, ricordando l'equazione (12.8), che la variabile casuale

$$X'' = (N-1)\,\frac{s^2}{\sigma^2}$$

è distribuita come il χ^2 ad $N-1$ gradi di libertà; inoltre, ovviamente,

$$u = \frac{\bar{x} - E(x)}{\frac{\sigma}{\sqrt{N}}}$$

segue la legge normale, con media 0 e varianza 1. Di conseguenza la variabile casuale

$$t = \frac{u}{\sqrt{\frac{X''}{N-1}}} = \frac{\bar{x} - E(x)}{\frac{s}{\sqrt{N}}} \tag{12.17}$$

[5]Per valori di $N \gtrsim 35$ la distribuzione di Student si può approssimare con la distribuzione normale a media 0 e varianza 1.

segue la distribuzione di Student ad $N-1$ gradi di libertà.

Insomma: se i campioni a disposizione non hanno dimensioni accettabili, una volta calcolato lo scarto normalizzato relativo alla differenza tra la media di un campione ed un valore prefissato occorrerà confrontare il suo valore con i limiti degli intervalli di confidenza relativi alla distribuzione di Student e non alla distribuzione normale[6].

12.5 La compatibilità di due valori misurati

Un altro caso frequente è quello in cui si hanno a disposizione due campioni di misure, e si vuole verificare l'ipotesi statistica che essi provengano da popolazioni aventi lo stesso valore medio: un caso particolare è quello dell'ipotesi consistente nell'essere i due campioni composti da misure della stessa grandezza fisica, che hanno prodotto differenti stime come effetto della presenza in entrambi degli errori; errori che assumiamo ancora seguire la legge normale.

Siano ad esempio un primo campione di N misure x_i, ed un secondo campione di M misure y_j; indichiamo con $\bar{x}$ e $\bar{y}$ le medie dei due campioni, con ${\sigma_x}^2$ e ${\sigma_y}^2$ le varianze delle popolazioni da cui tali campioni provengono, e con $\delta = \bar{x} - \bar{y}$ la differenza tra le due medie.

Sappiamo già che i valori medi e le varianze delle medie dei campioni sono legati ai corrispondenti valori relativi alle popolazioni dalle

$$E(\bar{x}) = E(x) \qquad , \qquad E(\bar{y}) = E(y)$$

e

$$\mathrm{Var}(\bar{x}) = \frac{{\sigma_x}^2}{N} \qquad , \qquad \mathrm{Var}(\bar{y}) = \frac{{\sigma_y}^2}{M}$$

per cui risulterà, se i campioni sono tra loro statisticamente indipendenti e se si ammette valida l'ipotesi (da verificare) che abbiano la stessa media,

$$E(\delta) = E(\bar{x} - \bar{y}) = E(x) - E(y) = 0$$

e

$$\mathrm{Var}(\delta) = \mathrm{Var}(\bar{x} - \bar{y}) = \frac{{\sigma_x}^2}{N} + \frac{{\sigma_y}^2}{M} \ .$$

Inoltre, essendo $\bar{x}$, $\bar{y}$ (e quindi δ) combinazioni lineari di variabili normali, seguiranno anch'esse la legge normale; e la verifica dell'ipotesi che i

[6] Per taluni più usati valori del livello di confidenza, i limiti rilevanti si possono trovare tabulati anche nell'appendice G.

campioni provengano da popolazioni aventi la stessa media si traduce nella verifica dell'ipotesi che δ abbia valore vero nullo.

Tale verifica, essendo δ distribuita secondo la legge normale, si esegue come abbiamo visto nel paragrafo precedente: si fissa arbitrariamente un valore del livello di confidenza, si determina il corrispondente valore limite degli scarti normalizzati, e lo si confronta con il valore di

$$\frac{\delta - E(\delta)}{\sigma_\delta} = \frac{\bar{x} - \bar{y}}{\sqrt{\dfrac{{\sigma_x}^2}{N} + \dfrac{{\sigma_y}^2}{M}}} \ .$$

Ovviamente vale anche qui l'osservazione fatta nel paragrafo precedente: non conoscendo le deviazioni standard delle popolazioni, σ_x e σ_y, siamo costretti ad usare in loro vece le stime ottenute dai campioni, s_x ed s_y; e questo si ammette generalmente lecito quando la dimensione di entrambi i campioni è almeno pari a 30.

In caso contrario, presupponendo cioè di avere a disposizione piccoli campioni per almeno una delle due variabili, limitiamo la nostra analisi al caso in cui si sappia con sicurezza che le due popolazioni x ed y *abbiano la stessa varianza*,

$${\sigma_x}^2 = {\sigma_y}^2 \equiv \sigma^2$$

e definiamo la grandezza S^2 (*varianza globale* dei campioni) come

$$S^2 = \frac{1}{N+M-2} \cdot \left\{ \sum_{i=1}^{N} [x_i - \bar{x}]^2 + \sum_{j=1}^{M} [y_j - \bar{y}]^2 \right\}$$

$$= \frac{(N-1)\, {s_x}^2 + (M-1)\, {s_y}^2}{N+M-2} \ .$$

Sapendo, dall'equazione (12.8), che le due variabili

$$(N-1)\,\frac{{s_x}^2}{\sigma^2} \qquad \text{e} \qquad (M-1)\,\frac{{s_y}^2}{\sigma^2}$$

sono entrambe distribuite come il χ^2, con $N-1$ ed $M-1$ gradi di libertà rispettivamente, sfruttando la regola di somma enunciata a pagina 199 si ricava che la variabile casuale

$$X = \frac{(N-1)\, {s_x}^2 + (M-1)\, {s_y}^2}{\sigma^2} = (N+M-2)\,\frac{S^2}{\sigma^2}$$

è distribuita come il χ^2 ad $N+M-2$ gradi di libertà; essendo inoltre $\delta = \bar{x} - \bar{y}$ una variabile normale con media e varianza date da

$$E(\delta) = E(x) - E(y) \qquad \text{e} \qquad {\sigma_\delta}^2 = \frac{\sigma^2}{N} + \frac{\sigma^2}{M}$$

la variabile casuale

$$u = \frac{\delta - E(\delta)}{\sigma_\delta} = \frac{(\bar{x} - \bar{y}) - [E(x) - E(y)]}{\sqrt{\sigma^2 \left(\frac{1}{N} + \frac{1}{M}\right)}}$$

è normale con media 0 e varianza 1. Per concludere,

$$t = \frac{u}{\sqrt{\frac{X}{N+M-2}}} = \frac{(\bar{x} - \bar{y}) - [E(x) - E(y)]}{\sqrt{S^2 \left(\frac{1}{N} + \frac{1}{M}\right)}} \tag{12.18}$$

deve seguire la distribuzione di Student con $N+M-2$ gradi di libertà; di conseguenza, per verificare l'ipotesi che le due popolazioni *normali* da cui i campioni provengono abbiano la stessa media *ammesso già che posseggano la stessa varianza*, basta confrontare con le apposite tabelle della distribuzione di Student il valore della t ottenuta dalla (12.18) ponendovi $E(x) - E(y) = 0$:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{S^2 \left(\frac{1}{N} + \frac{1}{M}\right)}} \quad .$$

12.6 La distribuzione di Fisher

Sia X una variabile casuale distribuita come il χ^2 ad M gradi di libertà; ed Y una seconda variabile casuale, indipendente dalla prima, distribuita ancora come il χ^2, ma con N gradi di libertà.

La variabile casuale w (sempre positiva) definita in funzione di esse attraverso la relazione

$$w = \frac{\frac{X}{M}}{\frac{Y}{N}}$$

ha una densità di probabilità che segue la cosiddetta *funzione di frequenza di Fisher* con M ed N gradi di libertà. La forma analitica della funzione di Fisher è data dalla

$$F(w; M, N) = K_{MN} \frac{w^{\frac{M}{2}-1}}{(Mw + N)^{\frac{M+N}{2}}} \tag{12.19}$$

(nella quale K_{MN} è un fattore costante determinato dalla condizione di normalizzazione).

Il valore medio e la varianza della funzione di frequenza di Fisher sono dati poi rispettivamente da

$$E(F) = \frac{N}{N-2} \qquad \text{(se } N > 2\text{)}$$

e da

$$\mathrm{Var}(F) = \frac{2N^2(M+N-2)}{M(N-2)^2(N-4)} \qquad \text{(se } N > 4\text{)} \quad .$$

Si può dimostrare che, se Y è una variabile casuale distribuita come il χ^2 ad N gradi di libertà,

$$\lim_{N \to +\infty} \frac{Y}{N} = 1$$

in senso statistico (ovverosia la probabilità che il rapporto Y/N sia differente da 1 tende a zero quando N viene reso arbitrariamente grande); per cui, indicando con $f(x; M)$ la funzione di frequenza del χ^2 ad M gradi di libertà,

$$F(w; M, \infty) \equiv \lim_{N \to +\infty} F(w; M, N) = \frac{f(w; M)}{M} \quad .$$

Allo stesso modo

$$F(w; \infty, N) \equiv \lim_{M \to +\infty} F(w; M, N) = \frac{N}{f(w; N)}$$

e quindi esiste una stretta relazione tra le distribuzioni di Fisher e del chi quadro.

Inoltre, ricordando che, se u è una variabile casuale distribuita secondo la legge normale standardizzata $N(u; 0, 1)$, l'altra variabile casuale u^2 è distribuita come il χ^2 ad un grado di libertà, il rapporto

$$w = \frac{u^2}{\frac{Y}{N}}$$

deve essere distribuito secondo $F(w; 1, N)$; ma, se definiamo

$$t = \frac{u}{\sqrt{\frac{Y}{N}}}$$

sappiamo anche dalla (12.16) che la t segue la distribuzione di Student ad N gradi di libertà. La conclusione è che il quadrato di una variabile t che segua