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

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

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

[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.

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).

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.

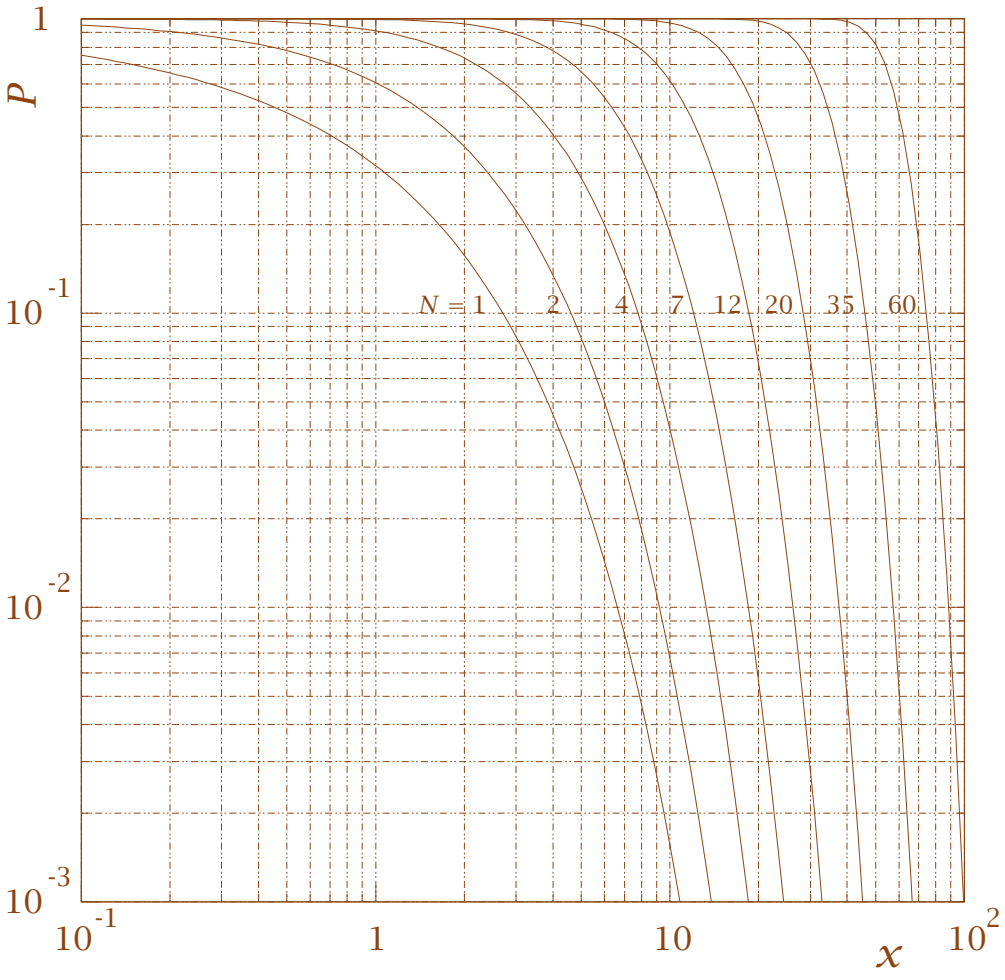
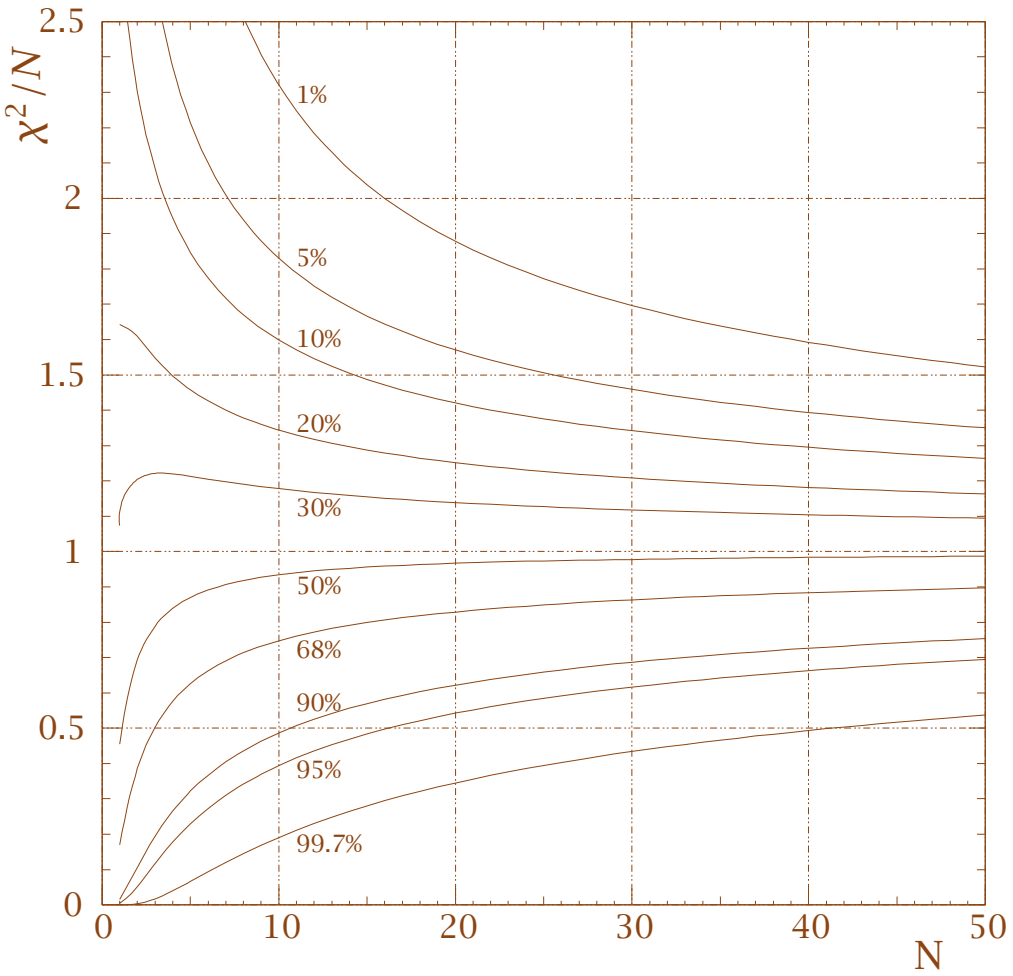


Figura 12c - I valori del χ^2 ridotto (χ^2/N) che corrispondono, per differenti gradi di libertà N, ad un certo livello di confidenza.



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

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

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

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

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

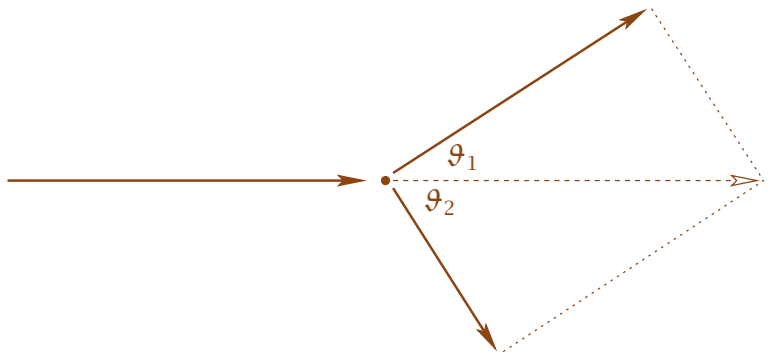
$$
\begin{aligned}
X &= \sum_{i,j} \frac{(\nu_{ij} - Np_i q_j)^2}{Np_i q_j} \\
&= \sum_{i,j} \left[\frac{(\nu_{ij})^2}{Np_i q_j} - 2\nu_{ij} + Np_i q_j \right] \\
&= \sum_{i,j} \frac{(\nu_{ij})^2}{Np_i q_j} - 2N + N \\
&= \sum_{i,j} \frac{(\nu_{ij})^2}{Np_i q_j} - N
\end{aligned}
$$

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

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

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

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

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 $-1 \leq t \leq 1$ corrisponde al 68.27% dell'area della curva normale, e quello $-2 \leq t \leq 2$ 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

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

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

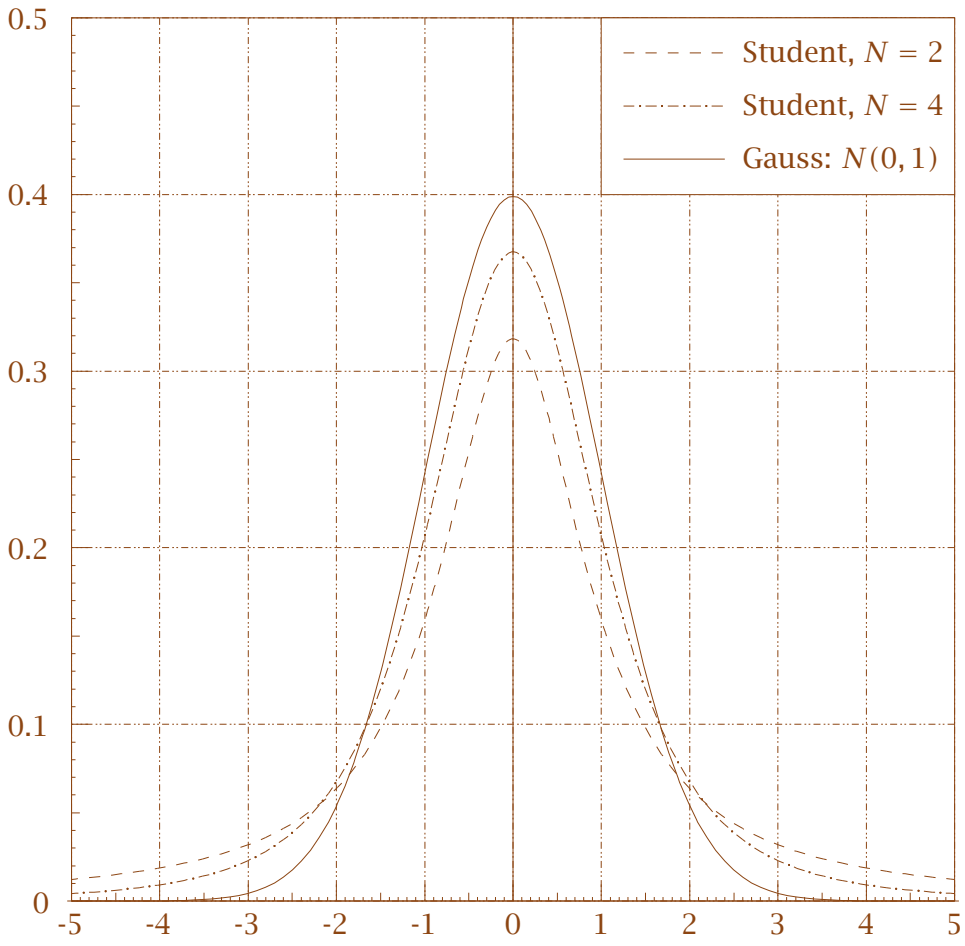
$$t = \frac{u}{\sqrt{\frac{X}{N}}} \quad . \tag{12.16}$$

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 $E(x) - E(y) = 0$:

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

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

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

e da

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

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,

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

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

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

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

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

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

la distribuzione di Student ad N gradi di libertà è a sua volta distribuito con una densità di probabilità data da $F(t^2; 1, N)$.

Per terminare, quando i due parametri M ed N (da cui la funzione di frequenza di Fisher (12.19) dipende) vengono resi arbitrariamente grandi, essa tende ad una distribuzione normale; ma la convergenza è lenta, e l'approssimazione normale alla distribuzione di Fisher si può pensare in pratica usabile quando sia M che N sono superiori a 50.

12.6.1 Confronto tra varianze

Supponiamo di avere a disposizione due campioni di misure, che ipotizziamo provenire da due differenti popolazioni che seguano delle distribuzioni normali.

Siano M ed N le dimensioni di tali campioni, e siano ${\sigma_1}^2$ e ${\sigma_2}^2$ le varianze delle rispettive popolazioni di provenienza; indichiamo poi con ${s_1}^2$ ed ${s_2}^2$ le due stime delle varianze delle popolazioni ricavate dai campioni. Vogliamo ora capire come si può verificare l'ipotesi statistica che le due popolazioni abbiano *la stessa varianza*, ossia che $\sigma_1 = \sigma_2$.

Ora sappiamo già dalla equazione (12.8) che le due variabili casuali

$$X = (M-1)\,\frac{{s_1}^2}{{\sigma_1}^2} \qquad \text{e} \qquad Y = (N-1)\,\frac{{s_2}^2}{{\sigma_2}^2}$$

sono entrambe distribuite come il χ^2, con $M-1$ ed $N-1$ gradi di libertà rispettivamente; quindi la quantità

$$w = \frac{X}{M-1}\,\frac{N-1}{Y} = \frac{{s_1}^2}{{\sigma_1}^2}\,\frac{{\sigma_2}^2}{{s_2}^2}$$

ha densità di probabilità data dalla funzione di Fisher con $M-1$ ed $N-1$ gradi di libertà.

Assunta a priori vera l'ipotesi statistica $\sigma_1 = \sigma_2$, la variabile casuale

$$w = \frac{{s_1}^2}{{s_2}^2}$$

ha densità di probabilità data dalla funzione di Fisher prima menzionata, $F(w; M-1, N-1)$; per cui, fissato un livello di confidenza al di là del quale rigettare l'ipotesi, e ricavato dalle apposite tabelle[7] il valore W che lascia alla propria sinistra, al di sotto della funzione $F(w; M-1, N-1)$, un'area pari al livello di confidenza prescelto, si può escludere che i due campioni provengano da popolazioni con la stessa varianza se $w > W$.

[7] Per un livello di confidenza pari a 0.95 o 0.99, e per alcuni valori dei due parametri M ed N, ci si può riferire ancora alle tabelle dell'appendice G; in esse si assume che sia $s_1 > s_2$, e quindi $w > 1$.

12.7 Il metodo di Kolmogorov e Smirnov

Il *test di Kolmogorov e Smirnov* è un metodo di analisi statistica che permette di confrontare tra loro un campione di dati ed una distribuzione teorica (oppure due campioni di dati) allo scopo di verificare l'ipotesi statistica che la popolazione da cui i dati provengono sia quella in esame (oppure l'ipotesi che entrambi i campioni provengano dalla stessa popolazione).

Una caratteristica interessante di questo metodo è che esso non richiede la preventiva, e più o meno arbitraria, suddivisione dei dati in classi di frequenza; definendo queste ultime in modo diverso si ottengono ovviamente, dal metodo del χ^2, differenti risultati per gli stessi campioni.

Il test di Kolmogorov e Smirnov si basa infatti sulla *frequenza cumulativa relativa* dei dati, introdotta nel paragrafo 4.1 a pagina 33; e sull'analogo concetto di *funzione di distribuzione* di una variabile continua definito nel paragrafo 6.1 a pagina 68. Per la compatibilità tra un campione ed una ipotetica legge che si ritiene possa descriverne la popolazione di provenienza, e collegata ad una funzione di distribuzione $\Phi(x)$, bisogna confrontare la frequenza cumulativa relativa $F(x)$ del campione con $\Phi(x)$ per ricavare *il valore assoluto del massimo scarto tra esse*,

$$\delta = \max\left\{ \left| F(x) - \Phi(x) \right| \right\} \ .$$

Si può dimostrare che, se l'ipotesi da verificare fosse vera, la probabilità di ottenere casualmente un valore di δ non inferiore ad una prefissata quantità (positiva) δ_0 sarebbe data da

$$\Pr\left(\delta \geq \delta_0\right) = F_{\mathrm{KS}}\left(\delta_0'\right)$$

ove F_{KS} è la serie

$$F_{\mathrm{KS}}(x) = 2 \sum_{k=1}^{\infty} (-1)^{k-1} e^{-2k^2x^2} \tag{12.20}$$

e δ_0' vale

$$\delta_0' = \left(\sqrt{N} + 0.12 + \frac{0.11}{\sqrt{N}} \right) \delta_0 \ . \tag{12.21}$$

La legge ora enunciata è approssimata, ma il test di Kolmogorov e Smirnov può essere usato già per dimensioni del campione N uguali a 5. Attenzione però che, se qualche parametro da cui la distribuzione teorica dipende è stato stimato sulla base dei dati, l'integrale della densità di probabilità

per la variabile δ di Kolmogorov e Smirnov *non segue più* la legge (12.20): non solo, ma non è più possibile ricavare teoricamente una funzione che ne descriva il comportamento in generale (in questi casi, nella pratica, la distribuzione di δ viene studiata usando metodi di Montecarlo).

Se si vogliono invece confrontare tra loro due campioni indipendenti per verificarne la compatibilità, bisogna ricavare dai dati il massimo scarto (in valore assoluto), δ, tra le due frequenze cumulative relative; e ricavare ancora dalla (12.20) la probabilità che questo possa essere avvenuto (ammessa vera l'ipotesi) per motivi puramente casuali. L'unica differenza è che la funzione (12.20) va calcolata in un'ascissa δ'_0 data dalla (12.21), nella quale N vale

$$N = \frac{1}{\frac{1}{N_1} + \frac{1}{N_2}} = \frac{N_1 N_2}{N_1 + N_2}$$

(N_1 ed N_2 sono le dimensioni dei due campioni).

Oltre al già citato vantaggio di non richiedere la creazione di più o meno arbitrarie classi di frequenza per raggrupparvi i dati, un'altra caratteristica utile del test di Kolmogorov e Smirnov è quella di essere, entro certi limiti, *indipendente dalla variabile usata* nella misura: se al posto di x si usasse, per caratterizzare il campione, $\ln(x)$ o $\sqrt{x}$, il massimo scarto tra frequenza cumulativa e funzione di distribuzione rimarrebbe invariato.

Un altrettanto ovvio svantaggio è collegato al fatto che per valori molto piccoli (o molto grandi) della variabile casuale usata, qualsiasi essa sia, *tutte* le funzioni di distribuzione e tutte le frequenze cumulative *hanno lo stesso valore* (0, o 1 rispettivamente). Per questo motivo il test di Kolmogorov e Smirnov è assai sensibile a differenze nella zona centrale dei dati (attorno al valore medio), mentre non è affatto efficace per discriminare tra due distribuzioni che differiscano significativamente tra loro solo nelle code; ad esempio che abbiano lo stesso valore medio e differente ampiezza.

Capitolo 13

La verifica delle ipotesi (II)

Nel precedente capitolo 12 abbiamo esaminato varie tecniche che ci permettono di decidere se una caratteristica del processo fisico che ha prodotto un campione di dati è o non è confermata dai dati stessi; tutte queste tecniche non sono che casi particolari di una teoria generale, di cui ora ci occuperemo, senza però scendere in profondità nei dettagli.

In sostanza, nei vari casi del capitolo 12, abbiamo formulato una certa ipotesi H_0 sulla natura di un fenomeno casuale; e, ammesso per assurdo che questa ipotesi fosse vera, abbiamo associato un ben definito valore della densità di probabilità ad ogni punto E dello spazio S degli eventi.

Se indichiamo con K un valore (arbitrariamente scelto) della probabilità, *livello di confidenza* nel linguaggio statistico, abbiamo in sostanza diviso S in due sottoinsiemi esclusivi ed esaurienti: uno $\mathcal{R}$ di eventi con probabilità complessiva $1 - K$, ed uno $\mathcal{A} = S - \mathcal{R}$ di eventi con probabilità complessiva K.

Per verificare l'ipotesi H_0 occorre scegliere a priori un valore di K da assumere come il confine che separi, da una parte, eventi che riteniamo ragionevole si possano presentare nell'ambito di pure fluttuazioni casuali se è vera H_0; e, dall'altra, eventi così improbabili (sempre ammesso che H_0 sia vera) da far sì che la loro effettiva realizzazione debba implicare la falsità dell'ipotesi.

Normalmente si sceglie $K = 0.95$ o $K = 0.997$, i valori della probabilità che corrispondono a scarti di due o tre errori quadratici medi per la distribuzione di Gauss, anche se altri valori (come ad esempio $K = 0.999$ o $K = 0.99$) sono abbastanza comuni; e, una volta fatto questo, si *rigetta l'ipotesi* H_0 se il

dato a disposizione (un evento E ottenuto dall'effettivo studio del fenomeno in esame) appartiene ad $\mathcal{R}$; e la si accetta se appartiene ad $\mathcal{A}$.

In realtà nella pratica si presenta in generale la necessità di discriminare tra *due* ipotesi, sempre mutuamente esclusive, che indicheremo con i simboli H_0 ed H_a e che, usando la terminologia della statistica, si chiamano rispettivamente *ipotesi nulla* ed *ipotesi alternativa*; i casi precedenti corrispondono al caso particolare in cui l'ipotesi alternativa coincida con il *non realizzarsi* di H_0.

Ipotesi nulla ed ipotesi alternativa possono essere entrambe eventi semplici, oppure composti (ossia somma logica di più eventualità semplici); e lo scopo di questo capitolo è quello di mostrare dei criteri sulla base dei quali si possa opportunamente definire nello spazio degli eventi una *regione di rigetto* $\mathcal{R}$ per l'ipotesi nulla (e, in corrispondenza, ovviamente, una regione $\mathcal{A} = \mathcal{S} - \mathcal{R}$ nella quale tale ipotesi viene accettata).

È chiaro che si corre sempre il rischio di sbagliare: o rigettando erroneamente ipotesi in realtà vere (*errori di prima specie*) o accettando invece ipotesi in realtà false (*errori di seconda specie*); e che, allargando o restringendo la regione di rigetto, si può diminuire la probabilità di uno di questi due tipi di errori solo per aumentare la probabilità di quelli dell'altra categoria. Se indichiamo con P_I e P_{II} le probabilità degli errori di prima e seconda specie rispettivamente, sulla base della definizione risulta

$$P_I = \Pr(E \in \mathcal{R}|H_0) \qquad \text{e} \qquad P_{II} = \Pr(E \in \mathcal{A}|H_a) \ .$$

Quello che abbiamo finora chiamato "livello di confidenza" non è altro che $1 - P_I$; P_I viene anche indicato col simbolo α e chiamato *significanza* del criterio adottato. Infine, la probabilità di *non* commettere un errore di seconda specie, ovvero la probabilità di rigettare H_0 quando l'ipotesi nulla è falsa (e quindi quella alternativa è vera) si indica col simbolo β e si chiama *potenza* del criterio adottato; essa vale quindi

$$\beta = \Pr(E \in \mathcal{R}|H_a) = 1 - P_{II} \ .$$

Per fare un esempio concreto, il fisico si trova spesso ad esaminare "eventi" sperimentali e deve decidere se essi sono del tipo desiderato (segnale) o no (fondo): in questo caso l'ipotesi nulla H_0 consiste nell'appartenenza di un evento al segnale, mentre l'ipotesi alternativa H_a corrisponde invece all'appartenenza dello stesso evento al fondo; che in genere non è l'intero insieme di eventi complementare all'ipotesi nulla, $\overline{H}_0$, ma si sa restringere ad una classe ben definita di fenomeni.

Gli errori di prima specie consistono in questo caso nello scartare eventi buoni (errori di *impoverimento* del segnale), e quelli di seconda specie nell'introduzione nel segnale di eventi di fondo (errori di *contaminazione*).

I criteri da seguire per definire una regione $\mathcal{R}$ nella quale rigettare H_0 sono dettati dalle caratteristiche del processo di generazione: se gli eventi di fondo sono preponderanti rispetto al segnale, ad esempio, bisognerà evitare gli errori di seconda specie per quanto possibile; anche al prezzo di scartare in questo modo una parte consistente del segnale.

Estendendo al caso generale il metodo seguito nei vari casi del capitolo 12 e prima delineato, se si è in grado di associare ad ogni punto dello spazio degli eventi *due* valori della probabilità (o della densità di probabilità nel caso di variabili continue), sia ammessa vera l'ipotesi nulla che ammessa invece vera l'ipotesi alternativa, si può pensare di usare *il loro rapporto* per definire la regione di rigetto.

Limitandoci al caso delle variabili continue, insomma, dopo aver definito una nuova variabile casuale λ attraverso la

$$\lambda = \frac{\mathcal{L}(\boldsymbol{x}|H_0)}{\mathcal{L}(\boldsymbol{x}|H_a)} \quad ,$$

possiamo scegliere arbitrariamente un numero reale k e decidere di accettare l'ipotesi H_0 se $\lambda \geq k$ o di rifiutarla se $\lambda < k$; in definitiva ad ogni k ammissibile è associata una differente regione di rigetto $\mathcal{R}_k$ definita da

$$\mathcal{R}_k \equiv \left\{\lambda = \frac{\mathcal{L}(\boldsymbol{x}|H_0)}{\mathcal{L}(\boldsymbol{x}|H_a)} < k\right\} \quad .$$

$\mathcal{L}$, nelle espressioni precedenti, è la funzione di verosimiglianza; che rappresenta appunto la densità di probabilità corrispondente all'ottenere (sotto una certa ipotesi) un campione di N valori $x_1, x_2, \ldots, x_N$ (qui indicato sinteticamente come un vettore $\boldsymbol{x}$ a N componenti). Ma in base a quale criterio dobbiamo scegliere k?

13.1 Un primo esempio

Cominciamo con un esempio didattico: supponiamo che i valori x_i si sappiano provenienti da una popolazione normale $N(x;\mu,\sigma)$ di varianza σ^2 nota: e che il nostro scopo consista nel discriminare tra due possibili valori μ_1 e μ_2 per μ; valori che, senza perdere in generalità, supponiamo siano 0 e 1 (potendosi sempre effettuare un opportuno cambiamento di variabile casuale che ci porti in questa situazione). Riassumendo: siano

$$\begin{cases} x \sim N(x;\mu,\sigma) & \text{(con } \sigma > 0 \text{ noto)} \\ H_0 \equiv \{\mu = 0\} \\ H_a \equiv \{\mu = 1\} \end{cases}$$

le nostre ipotesi.

La densità di probabilità della x vale

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

e, quindi, la funzione di verosimiglianza ed il suo logaritmo valgono

$$\mathcal{L}(\boldsymbol{x};\mu,\sigma) = \frac{1}{\left(\sigma\sqrt{2\pi}\right)^N} \prod_{i=1}^{N} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}$$

e, rispettivamente,

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

per cui

$$\ln \mathcal{L}(\boldsymbol{x};\mu,\sigma) = -N\,\ln(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2}\left(\sum_{i=1}^{N}{x_i}^2 - 2N\bar{x}\mu + N\mu^2\right) \quad . \tag{13.1}$$

Dalla (13.1) si ricava immediatamente

$$\ln\lambda \;=\; \ln\mathcal{L}(\boldsymbol{x};0,\sigma) - \ln\mathcal{L}(\boldsymbol{x};1,\sigma) \;=\; \frac{N}{2\sigma^2}(1-2\bar{x})$$

e la regione di rigetto $\mathcal{R}_k$ è definita dalla

$$\mathcal{R}_k \;\equiv\; \left\{\ln\lambda = \frac{N}{2\sigma^2}(1-2\bar{x}) < \ln k\right\}$$

da cui consegue, con facili passaggi,

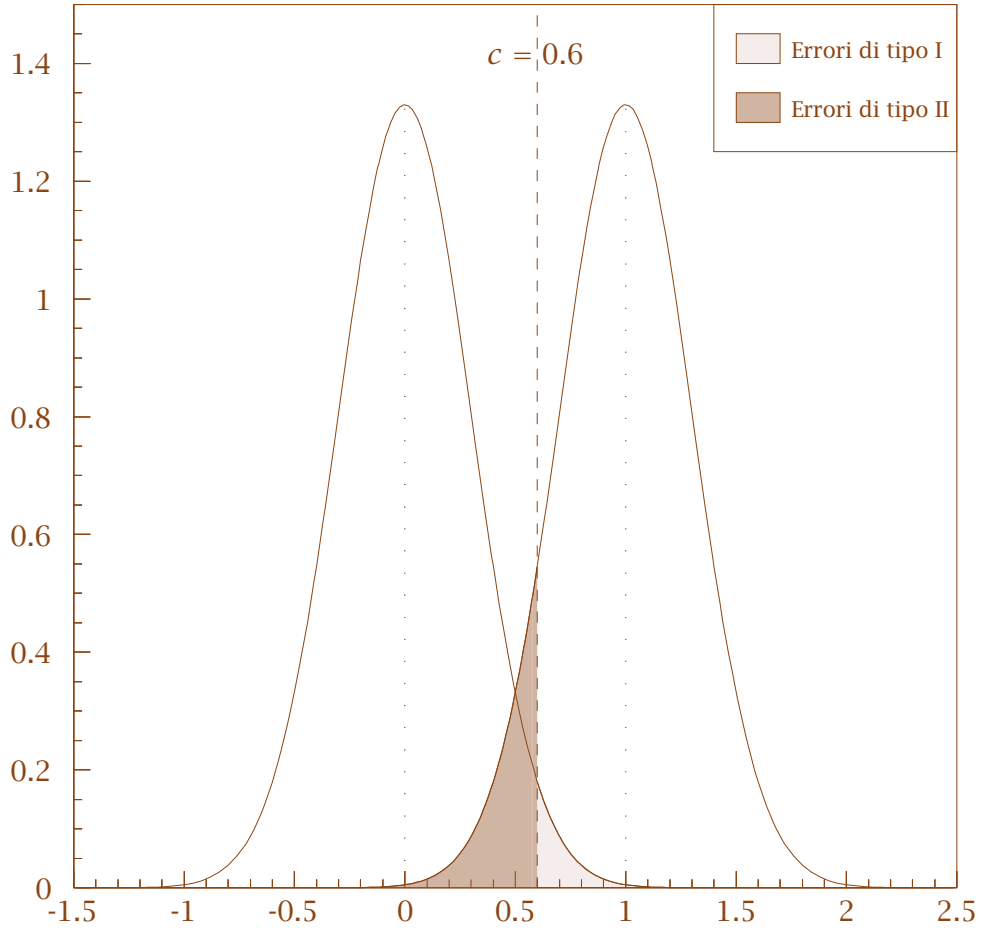
$$\mathcal{R}_k \;\equiv\; \left\{\bar{x} > \frac{N - 2\sigma^2\ln k}{2N} = c\right\} \quad ;$$

ed insomma H_0 va rigettata se la media aritmetica del campione $\bar{x}$ risulta superiore a c; ed accettata altrimenti.

Come si può scegliere un valore opportuno di k (e quindi di c)? Gli errori di prima specie (si faccia riferimento anche alla figura 13a) hanno probabilità

$$P_I \;=\; 1-\alpha \;=\; \Pr(\bar{x} > c|H_0) \;=\; \int_c^{+\infty} N\left(x;0,\frac{\sigma}{\sqrt{N}}\right)\mathrm{d}x \tag{13.2}$$

Figura 13a - L'esempio del paragrafo 13.1, con delineate (in corrispondenza ad un particolare valore di c) le probabilità degli errori di prima e seconda specie; le due curve sono $N(0, \sigma/\sqrt{N})$ e $N(1, \sigma/\sqrt{N})$.



e quelli di seconda specie

$$P_{II} \;=\; 1-\beta \;=\; \Pr(\bar{x} < c|H_a) \;=\; \int_{-\infty}^{c} N\left(x;1,\frac{\sigma}{\sqrt{N}}\right)\,\mathrm{d}x \tag{13.3}$$

per cui si hanno svariate possibilità: ad esempio, se interessa contenere gli errori di prima specie e la dimensione del campione è nota, si fissa un valore opportunamente grande per α e dalla (13.2) si ricava c; o, se interessa contenere gli errori di seconda specie e la dimensione del campione è nota, si fissa β e si ricava c dalla (13.3); o, infine, se si vogliono contenere gli errori di entrambi i tipi, si fissano sia α che β e si ricava la dimensione minima del campione necessaria per raggiungere lo scopo utilizzando entrambe le equazioni (13.2) e (13.3).

13.2 Il lemma di Neyman–Pearson

L'essere ricorsi per la definizione della regione di rigetto $\mathcal{R}$ al calcolo del *rapporto* delle funzioni di verosimiglianza non è stato casuale; esiste infatti un teorema (il cosiddetto *lemma di Neyman–Pearson*) il quale afferma che

> *Se si ha a disposizione un campione di N valori indipendenti x_i da utilizzare per discriminare tra un'ipotesi nulla ed un'ipotesi alternativa entrambe semplici, e se è richiesto un livello fisso α di significanza, la massima potenza del test (ovvero la minima probabilità di errori di seconda specie) si raggiunge definendo la regione di rigetto $\mathcal{R}_k$ attraverso una relazione del tipo*

$$\mathcal{R}_k \;\equiv\; \left\{\, \lambda = \frac{\mathcal{L}(\boldsymbol{x}|H_0)}{\mathcal{L}(\boldsymbol{x}|H_a)} < k \,\right\} \;. \tag{13.4}$$

Infatti, indicando con $f = f(x;\theta)$ la densità di probabilità della variabile x (che supponiamo dipenda da un solo parametro θ), siano $H_0 \equiv \{\theta = \theta_0\}$ e $H_a \equiv \{\theta = \theta_a\}$ le due ipotesi (semplici) tra cui decidere; la funzione di verosimiglianza vale, come sappiamo,

$$\mathcal{L}(\boldsymbol{x};\theta) = \prod_{i=1}^{N} f(x_i;\theta) \;.$$

Indichiamo con $\mathfrak{R}$ l'insieme di tutte le regioni $\mathcal{R}$ per le quali risulti

$$P_I \;=\; \int_{\mathcal{R}} \mathcal{L}(\boldsymbol{x};\theta_0)\,\mathrm{d}\boldsymbol{x} \;=\; 1-\alpha \tag{13.5}$$

con α costante prefissata (nella (13.5) abbiamo sinteticamente indicato con $\mathrm{d}\boldsymbol{x}$ il prodotto $\mathrm{d}x_1\,\mathrm{d}x_2\cdots\mathrm{d}x_N$). Vogliamo trovare quale di queste regioni rende massima

$$\beta \;=\; 1 - P_{II} \;=\; \int_{\mathcal{R}} \mathcal{L}(\boldsymbol{x};\theta_a)\,\mathrm{d}\boldsymbol{x} \;\;.$$

Ora, per una qualsiasi regione $\mathcal{R} \neq \mathcal{R}_k$, valgono sia la

$$\mathcal{R}_k \;=\; (\mathcal{R}_k \cap \mathcal{R}) \cup (\mathcal{R}_k \cap \overline{\mathcal{R}}\,)$$

che la

$$\mathcal{R} \;=\; (\mathcal{R} \cap \mathcal{R}_k) \cup (\mathcal{R} \cap \overline{\mathcal{R}}_k) \;\;;$$

e quindi, per una qualsiasi funzione $\phi(\boldsymbol{x})$, risulta sia

$$\int_{\mathcal{R}_k} \phi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \;=\; \int_{\mathcal{R}_k\cap\mathcal{R}} \phi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} + \int_{\mathcal{R}_k\cap\overline{\mathcal{R}}} \phi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

che

$$\int_{\mathcal{R}} \phi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \;=\; \int_{\mathcal{R}\cap\mathcal{R}_k} \phi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} + \int_{\mathcal{R}\cap\overline{\mathcal{R}}_k} \phi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

e, sottraendo membro a membro,

$$\int_{\mathcal{R}_k} \phi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \int_{\mathcal{R}} \phi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \;=\; \int_{\mathcal{R}_k\cap\overline{\mathcal{R}}} \phi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \int_{\mathcal{R}\cap\overline{\mathcal{R}}_k} \phi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \;\;. \tag{13.6}$$

Applicando la (13.6) alla funzione $\mathcal{L}(\boldsymbol{x}|\theta_a)$ otteniamo:

$$\begin{aligned}\int_{\mathcal{R}_k} \mathcal{L}(\boldsymbol{x}|\theta_a)\,\mathrm{d}\boldsymbol{x} - \int_{\mathcal{R}} \mathcal{L}(\boldsymbol{x}|\theta_a)\,\mathrm{d}\boldsymbol{x} \;&=\\ &= \int_{\mathcal{R}_k\cap\overline{\mathcal{R}}} \mathcal{L}(\boldsymbol{x}|\theta_a)\,\mathrm{d}\boldsymbol{x} - \int_{\mathcal{R}\cap\overline{\mathcal{R}}_k} \mathcal{L}(\boldsymbol{x}|\theta_a)\,\mathrm{d}\boldsymbol{x} \;\;;\end{aligned} \tag{13.7}$$

ma, nel primo integrale del secondo membro, essendo la regione di integrazione contenuta in $\mathcal{R}_k$, deve valere la (13.4); e quindi risultare ovunque

$$\mathcal{L}(\boldsymbol{x}|\theta_a) \;>\; \frac{1}{k}\cdot\mathcal{L}(\boldsymbol{x}|\theta_0)$$

mentre, per lo stesso motivo, nel secondo integrale

$$\mathcal{L}(\boldsymbol{x}|\theta_a) \;\leq\; \frac{1}{k}\cdot\mathcal{L}(\boldsymbol{x}|\theta_0)$$

e quindi la (13.7) implica che

$$\int_{\mathcal{R}_k} \mathcal{L}(\boldsymbol{x}|\theta_a)\,\mathrm{d}\boldsymbol{x} - \int_{\mathcal{R}} \mathcal{L}(\boldsymbol{x}|\theta_a)\,\mathrm{d}\boldsymbol{x} \;>$$
$$> \frac{1}{k}\cdot\left[\int_{\mathcal{R}_k\cap\overline{\mathcal{R}}} \mathcal{L}(\boldsymbol{x}|\theta_0)\,\mathrm{d}\boldsymbol{x} - \int_{\mathcal{R}\cap\overline{\mathcal{R}}_k} \mathcal{L}(\boldsymbol{x}|\theta_0)\,\mathrm{d}\boldsymbol{x}\right] .$$

Ricordando la (13.6),

$$\int_{\mathcal{R}_k} \mathcal{L}(\boldsymbol{x}|\theta_a)\,\mathrm{d}\boldsymbol{x} - \int_{\mathcal{R}} \mathcal{L}(\boldsymbol{x}|\theta_a)\,\mathrm{d}\boldsymbol{x} \;>\; \frac{1}{k}\cdot\left[\int_{\mathcal{R}_k} \mathcal{L}(\boldsymbol{x}|\theta_0)\,\mathrm{d}\boldsymbol{x} - \int_{\mathcal{R}} \mathcal{L}(\boldsymbol{x}|\theta_0)\,\mathrm{d}\boldsymbol{x}\right]$$

e, se $\mathcal{R} \in \mathfrak{R}$ e quindi soddisfa anch'essa alla (13.5),

$$\int_{\mathcal{R}_k} \mathcal{L}(\boldsymbol{x}|\theta_a)\,\mathrm{d}\boldsymbol{x} - \int_{\mathcal{R}} \mathcal{L}(\boldsymbol{x}|\theta_a)\,\mathrm{d}\boldsymbol{x} \;>\; \frac{1}{k}\cdot[P_I - P_I] \;=\; 0$$

che era la nostra tesi.

13.3 Tests di massima potenza uniforme

Consideriamo ora un esempio del tipo di quello del paragrafo 13.1; e sia sempre disponibile un campione di N misure indipendenti derivante da una popolazione normale di varianza nota. Assumiamo ancora come ipotesi nulla quella che la popolazione abbia un certo valore medio, che supponiamo essere 0, ma sostituiamo alla vecchia ipotesi alternativa H_a una nuova ipotesi *composta*; ovvero quella che il valore medio della popolazione sia positivo:

$$\begin{cases} x \;\sim N(x;\mu,\sigma) & \text{(con } \sigma > 0 \text{ noto)} \\ H_0 \equiv \{\mu = 0\} \\ H_a \equiv \{\mu > 0\} \end{cases}$$

(l'ipotesi alternativa è dunque somma logica di infinite ipotesi semplici del tipo $\mu = \mu_a$ con $\mu_a > 0$).

Dalla (13.1) ricaviamo immediatamente le

$$\mathcal{L}(\boldsymbol{x};0,\sigma) \;=\; -N\,\ln(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2}\sum_{i=1}^{N} {x_i}^2$$

e

$$\mathcal{L}(\boldsymbol{x};\mu_a,\sigma) \;=\; -N\,\ln(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2}\left(\sum_{i=1}^{N} {x_i}^2 - 2N\bar{x}\mu_a + N{\mu_a}^2\right)$$

(sempre con $\mu_a > 0$); e, sostituendole nella (13.4), che definisce la generica regione di rigetto $\mathcal{R}_k$, otteniamo

$$\ln \lambda \;=\; \ln \mathcal{L}(\boldsymbol{x}; 0, \sigma) - \mathcal{L}(\boldsymbol{x}; \mu_a, \sigma) \;=\; \frac{N\mu_a}{2\sigma^2}\,(\mu_a - 2\bar{x}) \;<\; \ln k$$

equivalente alla

$$\mathcal{R}_k \;\equiv\; \left\{ \bar{x} > \frac{N{\mu_a}^2 - 2\sigma^2 \ln k}{2N\mu_a} = c \right\} \;.$$

Si rigetta quindi H_0 se la media aritmetica del campione è superiore a c e la si accetta altrimenti: la probabilità di commettere errori di prima specie vale

$$P_I \;=\; 1 - \alpha \;=\; \int_c^{+\infty} N\left(x; 0, \frac{\sigma}{\sqrt{N}}\right) \mathrm{d}x$$

ed è ben definita; ma, al contrario, la probabilità di commettere errori di seconda specie dipende dal particolare valore di μ_a, e non può quindi essere calcolata.

Se interessa solo contenere gli errori di prima specie e la dimensione del campione è nota, si fissa α e si ricava il corrispondente valore di c dall'equazione precedente; altrimenti occorre fare delle ulteriori ipotesi sulla funzione di frequenza dei differenti valori di μ_a, e, ad esempio, calcolare la probabilità degli errori di seconda specie con tecniche di Montecarlo.

In ogni caso, però, osserviamo che la regione di rigetto è sempre dello stesso tipo (13.4) per *qualsiasi* $\mu_a > 0$; e quindi un confronto separato tra H_0 ed ognuna delle differenti ipotesi semplici che costituiscono H_a è comunque del tipo per cui il lemma di Neyman–Pearson garantisce la massima potenza.

Tests di questo tipo, per i quali *la significanza è costante e la potenza è massima per ognuno dei casi semplici che costituiscono l'ipotesi alternativa*, si dicono "tests di massima potenza uniforme".

13.4 Il rapporto delle massime verosimiglianze

Nel caso generale in cui sia l'ipotesi nulla che quella alternativa siano composte, la situazione è più complicata: non esiste normalmente un test di massima potenza uniforme, e, tra i vari criteri possibili per decidere tra le due ipotesi, bisogna capire quali abbiano caratteristiche (significanza e potenza) adeguate; un metodo adatto a costruire una regione di rigetto dotata asintoticamente (per grandi campioni) di caratteristiche, appunto, desiderabili, è quello seguente (*metodo del rapporto delle massime verosimiglianze*).

Sia una variabile casuale x, la cui densità di probabilità supponiamo sia una funzione $f(x;\theta_1,\theta_2,\ldots,\theta_M)$ dipendente da M parametri: indicando sinteticamente la M-pla dei valori dei parametri come un vettore $\boldsymbol{\theta}$ in uno spazio a M dimensioni (*spazio dei parametri*), consista H_0 nell'essere $\boldsymbol{\theta}$ compreso all'interno di una certa regione Ω_0 di tale spazio; mentre H_a consista nell'appartenere $\boldsymbol{\theta}$ alla regione Ω_a complementare a Ω_0: $\Omega_a \equiv \overline{H}_0$, così che $(\Omega_0 \cup \Omega_a)$ coincida con l'intero spazio dei parametri S.

In particolare, Ω_0 può estendersi, in alcune delle dimensioni dello spazio dei parametri, da $-\infty$ a $+\infty$; e, in tal caso, il vincolo sulle θ_i cui corrisponde l'ipotesi nulla riguarderà un numero di parametri minore di M.

Scritta la funzione di verosimiglianza,

$$\mathcal{L}(\boldsymbol{x};\boldsymbol{\theta}) = \prod_{i=1}^{N} f(x_i;\boldsymbol{\theta}) \tag{13.8}$$

indichiamo con $\mathcal{L}(\hat{S})$ il suo massimo valore nell'intero spazio dei parametri; e con $\mathcal{L}(\hat{R})$ il massimo valore assunto sempre della (13.8), ma con i parametri vincolati a trovarsi nella regione Ω_0 (quindi limitatamente a quei casi nei quali H_0 è vera). Il rapporto

$$\lambda = \frac{\mathcal{L}(\hat{R})}{\mathcal{L}(\hat{S})} \tag{13.9}$$

deve essere un numero appartenente all'intervallo $[0,1]$; se si fissa un arbitrario valore k $(0 < k < 1)$, esso definisce una generica regione di rigetto, $\mathcal{R}_k$, attraverso la

$$\mathcal{R}_k \equiv \left\{ \lambda = \frac{\mathcal{L}(\hat{R})}{\mathcal{L}(\hat{S})} < k \right\}$$

(ovvero si accetta H_0 quando $\lambda \geq k$ e la si rigetta quando $\lambda < k$). Nel caso si sappia determinare la densità di probabilità di λ condizionata all'assunzione che H_0 sia vera, $g(\lambda|H_0)$, la probabilità di un errore di prima specie è data ovviamente da

$$P_I = \alpha = \Pr(\lambda \in [0,k]|H_0) = \int_0^k g(\lambda|H_0)\,\mathrm{d}\lambda \ .$$

L'importanza del metodo sta nel fatto che si può dimostrare il seguente

TEOREMA: *se l'ipotesi nulla H_0 consiste nell'appartenenza di un insieme di $P \leq M$ dei parametri θ_i ad una determinata regione Ω_0, e se l'ipotesi alternativa H_a consiste nel fatto che essi non vi appartengano ($H_a \equiv \overline{H}_0$), allora $-2\ln\lambda$, ove λ è definito dalla* (13.9)*, ha densità di probabilità che, ammessa vera l'ipotesi nulla, converge in probabilità (all'aumentare di N) alla distribuzione del χ^2 a P gradi di libertà.*

che, dicendoci quale è (almeno nel limite di grandi campioni) la forma di $g(\lambda|H_0)$, ci mette comunque in grado di calcolare la significanza del test.

Illustriamo il metodo con un esempio: disponendo ancora di un campione di N determinazioni indipendenti, provenienti da una popolazione normale di varianza nota, vogliamo applicarlo per discriminare tra l'ipotesi nulla che il valore medio abbia valore 0 ($H_0 \equiv \{\mu = 0\}$) e quella che esso abbia valore differente ($H_a \equiv \{\mu \neq 0\}$).

Il logaritmo della funzione di verosimiglianza è ancora dato dalla (13.1); e già sappiamo, dal paragrafo 11.3, che $\mathcal{L}$ assume il suo massimo valore quando $\mu = \bar{x}$, per cui

$$\ln \mathcal{L}(\hat{S}) = -N \ln(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2}\left(\sum_{i=1}^{N} {x_i}^2 - N\bar{x}^2\right) \quad .$$

Inoltre Ω_0 si riduce ad un unico punto, $\mu = 0$; per cui

$$\ln \mathcal{L}(\hat{R}) = -N \ln(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2}\sum_{i=1}^{N} {x_i}^2 \quad .$$

Dalla (13.9) si ricava

$$\ln\lambda \;=\; \ln \mathcal{L}(\hat{R}) - \ln \mathcal{L}(\hat{S}) \;=\; -\frac{1}{2\sigma^2} N\bar{x}^2$$

e la regione di rigetto è definita dalla $\ln\lambda < \ln k$; ovvero (ricordando che $\ln k < 0$) da

$$\mathcal{R}_k \;\equiv\; \left\{ \bar{x}^2 > -\frac{2\sigma^2 \ln k}{N} \right\}$$

e, posto

$$c = \sigma\sqrt{-\frac{2\ln k}{N}}$$

si accetterà H_0 se $|\bar{x}| \leq c$ (e la si rigetterà se $|\bar{x}| > c$).

In questo caso il teorema precedentemente citato afferma che

$$-2\ln\lambda = \frac{\bar{x}^2}{\dfrac{\sigma^2}{N}}$$

è distribuito asintoticamente come il χ^2 ad un grado di libertà (cosa che del resto già sapevamo, vista l'espressione di $-2\ln\lambda$); per cui, indicando

con $F(t;N)$ la densità di probabilità della distribuzione del χ^2 a N gradi di libertà, avremo

$$P_I \;=\; \alpha \;=\; \int_0^k g(\lambda|H_0)\,\mathrm{d}\lambda \;=\; \int_{-2\ln k}^{+\infty} F(t;1)\,\mathrm{d}t$$

della quale ci possiamo servire per ricavare k se vogliamo che la significanza del test abbia un certo valore: ad esempio un livello di confidenza del 95% corrisponde ad $\alpha = 0.05$ e, dalle tabelle della distribuzione del χ^2, ricaviamo

$$-2\ln k = 3.84 \qquad \text{e quindi} \qquad c = 1.96\,\frac{\sigma}{\sqrt{N}}\ .$$

Anche senza dover ricorrere al teorema sul comportamento asintotico di $-2\ln\lambda$, allo stesso risultato si può pervenire per altra via: in questo caso si conosce infatti esattamente α, che vale

$$P_I \;=\; \alpha \;=\; \Pr\Big(|\bar{x}| > c\,|H_0\Big) \;=\; 2\int_c^{+\infty} N\left(t;0,\frac{\sigma}{\sqrt{N}}\right)\mathrm{d}t$$

e, dalle tabelle della distribuzione normale standardizzata, si ricava che un'area two-tailed del 5% corrisponde ad un valore assoluto dello scarto normalizzato $t_0 = 1.96$; per cui, ancora, si ricaverebbe $|\bar{x}| > 1.96(\sigma/\sqrt{N})$ come test per un livello di confidenza del 95%.

13.5 Applicazione: ipotesi sulle probabilità

Nel paragrafo 11.5 abbiamo preso in considerazione il caso di un evento casuale che si può manifestare in un numero finito M di modalità, aventi ognuna probabilità incognita p_i; la stima di massima verosimiglianza delle p_i è data dal rapporto tra la frequenza assoluta di ogni modalità, n_i, ed il numero totale di prove, N.

Vogliamo ora applicare il metodo del rapporto delle massime verosimiglianze per discriminare, sulla base di un campione di determinazioni indipendenti, l'ipotesi nulla che le probabilità abbiano valori noti a priori e l'ipotesi alternativa complementare, $H_a \equiv \overline{H}_0$:

$$\begin{cases} H_0 \;\equiv\; \{p_i = \pi_i\} & (\forall\, i \in \{1,2,\ldots,M\}) \\ H_a \;\equiv\; \{p_i \neq \pi_i\} & (\exists\, i \in \{1,2,\ldots,M\}) \end{cases}$$

Ricordiamo che la funzione di verosimiglianza, a meno di un fattore moltiplicativo costante, è data da

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

e che, essendo la stima di massima verosimiglianza data da

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

il massimo assoluto di $\mathcal{L}$ è

$$\mathcal{L}(\hat{S}) \;=\; \prod_{i=1}^{M} \left(\frac{n_i}{N}\right)^{n_i} \;=\; \frac{1}{N^N} \prod_{i=1}^{M} {n_i}^{n_i} \quad .$$

Inoltre, nell'unico punto dello spazio dei parametri che corrisponde ad H_0,

$$\mathcal{L}(\hat{R}) = \prod_{i=1}^{M} {\pi_i}^{n_i}$$

per cui

$$\lambda \;=\; \frac{\mathcal{L}(\hat{R})}{\mathcal{L}(\hat{S})} \;=\; N^N \prod_{i=1}^{M} \left(\frac{\pi_i}{n_i}\right)^{n_i}$$

dalla quale si può, come sappiamo, derivare una generica regione di rigetto attraverso la consueta $\mathcal{R}_k \equiv \{\lambda < k\}$.

$$-2\,\ln\lambda \;=\; -2\left[N\,\ln N + \sum_{i=1}^{M} n_i\,(\ln\pi_i - \ln n_i)\right]$$

è inoltre asintoticamente distribuita come il χ^2 a $M-1$ gradi di libertà (c'è un vincolo: che le n_i abbiano somma N), e questo può servire a scegliere un k opportuno (nota la dimensione del campione) una volta fissata α.

Il criterio di verifica dell'ipotesi dato in precedenza consisteva nel calcolo del valore della variabile casuale

$$X = \sum_{i=1}^{M} \frac{(n_i - N\pi_i)^2}{N\pi_i}$$

e nel suo successivo confronto con la distribuzione del χ^2 a $M-1$ gradi di libertà; lo studio del rapporto delle massime verosimiglianze porta dunque ad un criterio *differente* e, senza sapere nulla della probabilità di commettere errori di seconda specie, non è possibile dire quale dei due risulti migliore (a parità di significanza).

13.6 Applicazione: valore medio di una popolazione normale

Ancora un esempio: sia una popolazione normale $N(x;\mu,\sigma)$ dalla quale vengano ottenuti N valori indipendenti x_i, ma questa volta *la varianza σ sia ignota*; vogliamo discriminare, sulla base del campione, tra l'ipotesi nulla che il valore medio della popolazione abbia un valore prefissato e l'ipotesi alternativa complementare,

$$\begin{cases} H_0 \equiv \{\mu = \mu_0\} \\ H_a \equiv \{\mu \neq \mu_0\} \end{cases}$$

Il logaritmo della funzione di verosimiglianza è

$$\ln \mathcal{L}(\boldsymbol{x};\mu,\sigma) = -N \ln\sigma - \frac{N}{2}\ln(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\mu)^2 \tag{13.10}$$

ed essendo le stime di massima verosimiglianza date, come avevamo trovato nel paragrafo 11.5, da

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

ne deriva, sostituendo nella (13.10), che

$$\ln \mathcal{L}(\hat{S}) = -\frac{N}{2}\ln\left[\sum_{i=1}^{N}(x_i-\bar{x})^2\right] + \frac{N}{2}\ln N - \frac{N}{2}\ln(2\pi) - \frac{N}{2} \ .$$

D'altra parte, ammessa vera H_0, abbiamo che

$$\ln \mathcal{L}(\boldsymbol{x}|H_0) = -N\ln\sigma - \frac{N}{2}\ln(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\mu_0)^2$$

e, derivando rispetto a σ,

$$\frac{\mathrm{d}}{\mathrm{d}\sigma}\ln \mathcal{L}(\boldsymbol{x}|H_0) = -\frac{N}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{N}(x_i-\mu_0)^2 \ .$$

Annullando la derivata prima, si trova che l'unico estremante di $\mathcal{L}(\boldsymbol{x}|H_0)$ si ha per

$$\sigma_0 = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu_0)^2$$

mentre la derivata seconda vale

$$\frac{\mathrm{d}^2}{\mathrm{d}\sigma^2} \ln \mathcal{L}(\boldsymbol{x}|H_0) = \frac{N}{\sigma^2} - \frac{3}{\sigma^4} \sum_{i=1}^{N} (x_i - \mu_0)^2$$

e, calcolata per $\sigma = \sigma_0$,

$$\left.\frac{\mathrm{d}^2(\ln \mathcal{L})}{\mathrm{d}\sigma^2}\right|_{\sigma=\sigma_0} = -\frac{2N^2}{\sum_i (x_i - \mu_0)^2} < 0$$

per cui l'estremante è effettivamente un massimo. Sostituendo,

$$\ln \mathcal{L}(\hat{R}) = -\frac{N}{2} \ln \left[\sum_{i=1}^{N} (x_i - \mu_0)^2 \right] + \frac{N}{2} \ln N - \frac{N}{2} \ln(2\pi) - \frac{N}{2}$$

$$\ln \lambda = \ln \mathcal{L}(\hat{R}) - \ln \mathcal{L}(\hat{S}) = -\frac{N}{2} \left\{ \ln \left[\sum_{i=1}^{N} (x_i - \mu_0)^2 \right] - \ln \left[\sum_{i=1}^{N} (x_i - \bar{x})^2 \right] \right\}$$

ed infine

$$\begin{aligned} \ln \lambda &= -\frac{N}{2} \ln \left[\frac{\sum_i (x_i - \mu_0)^2}{\sum_i (x_i - \bar{x})^2} \right] \\ &= -\frac{N}{2} \ln \left[1 + \frac{N(\bar{x} - \mu_0)^2}{\sum_i (x_i - \bar{x})^2} \right] \\ &= -\frac{N}{2} \ln \left(1 + \frac{t^2}{N-1} \right) \end{aligned}$$

tenendo conto dapprima del fatto che $\sum_i (x_i - \mu_0)^2 = \sum_i (x_i - \bar{x})^2 + N(\bar{x} - \mu_0)^2$, e definendo poi una nuova variabile casuale

$$t = (\bar{x} - \mu_0) \sqrt{\frac{N(N-1)}{\sum_i (x_i - \bar{x})^2}} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{N}}} .$$

Un qualunque metodo per il rigetto di H_0 definito confrontando λ con un prefissato valore k si traduce, in sostanza, in un corrispondente confronto da eseguire per t:

$$\mathcal{R}_k \equiv \{\ln \lambda < \ln k\}$$

che porta alla

$$-\frac{N}{2}\ln\left(1+\frac{t^2}{N-1}\right) \;<\; \ln k$$

ed alla condizione

$$t^2 \;>\; (N-1)\left(k^{-\frac{2}{N}}-1\right) \;\;;$$

ovvero si rigetta l'ipotesi nulla se $|t|$ è maggiore di un certo t_0 (derivabile dall'equazione precedente), e la si accetta altrimenti.

Ma t (vedi anche l'equazione (12.17)) segue la distribuzione di Student a $N-1$ gradi di libertà, e quindi accettare o rigettare H_0 sotto queste ipotesi si riduce ad un test relativo a quella distribuzione: come già si era concluso nel capitolo 12. Il livello di significanza α è legato a t_0 dalla

$$\frac{\alpha}{2} = \int_{t_0}^{+\infty} F(t;N-1)\,\mathrm{d}t$$

(indicando con $F(t;N)$ la funzione di frequenza di Student a N gradi di libertà), tenendo conto che abbiamo a che fare con un two-tailed test ($\mathcal{R}_k \equiv \{|t| > t_0\}$).

Insomma non c'è differenza, in questo caso, tra quanto esposto nel capitolo precedente e la teoria generale discussa in quello presente: nel senso che i due criteri di verifica dell'ipotesi portano per questo problema allo stesso metodo di decisione (ma, come abbiamo visto nel paragrafo precedente, non è sempre così).

Appendice A

Cenni di calcolo combinatorio

Il *calcolo combinatorio* è una branca della matematica orientata alla discussione ed allo sviluppo di formule che permettano di ottenere il numero di casi distinti che si possono presentare in un esperimento, od il numero di elementi che compongono un insieme, senza ricorrere alla loro enumerazione esplicita.

Il calcolo combinatorio trova importanti applicazioni nella teoria della probabilità e nella statistica: alcune formule, specificatamente quelle per le permutazioni e le combinazioni, vengono usate nel corso del testo; qui se ne dà una breve giustificazione.

A.1 Il lemma fondamentale del calcolo combinatorio

LEMMA FONDAMENTALE DEL CALCOLO COMBINATORIO: *dati due insiemi I_1 ed I_2, composti da N_1 ed N_2 elementi distinti rispettivamente, l'insieme $I = I_1 \otimes I_2$ di tutte le coppie ordinate che si possono costruire associando un elemento di I_1 con un elemento di I_2 è composto da $N_1 \cdot N_2$ elementi.*

Questo lemma si può immediatamente generalizzare (per induzione completa) a K insiemi $I_1, \ldots, I_K$ composti da $N_1, \ldots, N_K$ elementi distinti rispettivamente: l'insieme $I = I_1 \otimes I_2 \otimes \cdots \otimes I_K$, costituito da tutte le possibili associazioni ordinate di K elementi ognuno dei quali provenga da un differente insieme I_j, con $j = 1, \ldots, K$, è composto da $N_1 \cdot N_2 \cdots N_K$ elementi.

A.2 Fattoriale di un numero intero

Si definisce come *fattoriale* di un numero intero positivo N, e si indica con il simbolo $N!$, il prodotto dei primi N numeri interi:

$$N! = 1 \cdot 2 \cdot 3 \cdots N \ ;$$

per motivi che appariranno chiari più avanti[1], si definisce poi il fattoriale di zero come $0! = 1$.

A.3 Disposizioni

Se N e K sono due numeri interi positivi tali che sia $K \leq N$, si definisce come numero delle disposizioni di N oggetti di classe K (che si indica con il simbolo D_K^N) il numero dei gruppi *distinti* di K oggetti che è possibile formare a partire dagli N originali; definendo come *distinti* due gruppi se essi differiscono o per qualche elemento o per l'ordine.

Come esempio, le disposizioni di classe 2 che si possono formare con le 21 lettere dell'alfabeto italiano sono le seguenti:

$$\left\{ \begin{array}{cccccccc} & AB & AC & AD & \cdots & AV & AZ \\ BA & & BC & BD & \cdots & BV & BZ \\ & & & & \cdots & & \\ ZA & ZB & ZC & ZD & \cdots & ZV & \end{array} \right.$$

Il valore di D_K^N si può facilmente trovare sfruttando il lemma fondamentale del calcolo combinatorio: il primo elemento di una disposizione si può infatti scegliere in N modi distinti, il secondo in $N-1$, e così via. Di conseguenza D_K^N è il prodotto di K numeri interi decrescenti a partire da N:

$$D_K^N \ = \ N \cdot (N-1) \cdot (N-2) \cdots (N-K+1) \ = \ \frac{N!}{(N-K)!} \qquad \text{(A.1)}$$

(nel caso dell'esempio fatto, le disposizioni sono $D_2^{21} = 21 \cdot 20 = 420$; nella tabella in cui sono state elencate vi sono 21 righe di 20 elementi ciascuna).

L'espressione (A.1) è verificata anche se $K = N$, però purché (come prima detto) si ponga $0! = 1$.

[1] La "definizione" $0! = 1$ non è così arbitraria come può sembrare: in realtà si comincia definendo una certa funzione di variabile complessa $\Gamma(z)$ che, quando l'argomento z è un numero intero positivo, coincide con il suo fattoriale; e per la quale si vede che $\Gamma(0) = 1$.

A.4 Permutazioni

Se N è un numero intero positivo, si definisce come numero delle permutazioni di N oggetti, e si indica con P_N, il numero di maniere distinte in cui si possono ordinare gli N oggetti stessi. Evidentemente risulta

$$P_N \equiv D_N^N = N! \ .$$

A.5 Permutazioni con ripetizione

Se gli N oggetti che si hanno a disposizione sono tali da poter essere divisi in M gruppi (composti da $N_1, N_2, \ldots, N_M$ oggetti rispettivamente; ovviamente $N_1 + N_2 + \cdots + N_M = N$), tali che gli oggetti in ognuno di questi gruppi siano *indistinguibili* tra loro, il numero di permutazioni che con essi si possono realizzare è inferiore a P_N; più precisamente, visto che gli oggetti di ogni gruppo si possono scambiare tra loro in qualsiasi modo senza per questo dare luogo a una sequenza distinta, il numero di *permutazioni con ripetizione* è dato da

$$\frac{N!}{N_1! \cdot N_2! \cdots N_M!} \ . \tag{A.2}$$

A.6 Combinazioni

Se N e K sono due numeri interi positivi tali che sia $K \leq N$, si definisce come numero delle combinazioni di classe K di N oggetti il numero dei sottoinsiemi *distinti* composti da K oggetti che è possibile formare a partire dagli N originali; definendo come *distinti* due sottoinsiemi se essi differiscono per qualche elemento. Il numero delle combinazioni di classe K di N oggetti si indica con uno dei due simboli

$$C_K^N \qquad \text{o} \qquad \binom{N}{K}$$

(l'ultimo dei quali si chiama *coefficiente binomiale*).

Consideriamo l'insieme composto da tutte le disposizioni di classe K di N oggetti, e pensiamo di raggruppare i suoi elementi in sottoinsiemi in modo

che ciascuno di essi contenga tutte e sole quelle disposizioni che differiscano esclusivamente per l'ordine ma siano composte dagli stessi oggetti; ovviamente il numero di questi sottoinsiemi è C_K^N: ed ognuno di essi contiene un numero di elementi che è P_K.

Da qui ricaviamo

$$C_K^N \equiv \binom{N}{K} = \frac{D_K^N}{P_K} = \frac{N \cdot (N-1) \cdots (N-K+1)}{K \cdot (K-1) \cdots 1} = \frac{N!}{K!\,(N-K)!} \tag{A.3}$$

O, in altre parole, il numero di combinazioni di classe K di N oggetti è uguale al rapporto tra il prodotto di K numeri interi decrescenti a partire da N ed il prodotto di K numeri interi crescenti a partire dall'unità.

Si dimostrano poi facilmente, a partire dalla definizione, due importanti proprietà dei coefficienti binomiali:

$$\binom{N}{K} = \binom{N}{N-K}$$

e

$$\binom{N+1}{K} = \binom{N}{K-1} + \binom{N}{K} \ .$$

È da osservare che, così come sono stati ricavati (dalla definizione delle possibili combinazioni di N oggetti), i coefficienti binomiali hanno senso solo se N e K sono numeri interi; ed inoltre se risulta sia $N > 0$ che $0 \leq K \leq N$. La definizione (A.3) può comunque essere estesa a valori interi qualunque, ed anche a valori reali di N — ma questo esula dal nostro interesse.

A.7 Partizioni ordinate

Consideriamo un insieme di N oggetti; vogliamo calcolare il numero di maniere in cui essi possono essere divisi in M gruppi che siano composti da $N_1, N_2, \ldots, N_M$ oggetti rispettivamente (essendo $N_1 + N_2 + \cdots + N_M = N$).

Gli N_1 oggetti che compongono il primo gruppo possono essere scelti in $C_{N_1}^N$ modi differenti; quelli del secondo gruppo in $C_{N_2}^{N-N_1}$ modi; e così via. Per il lemma fondamentale del calcolo combinatorio, il numero delle *partizioni*

ordinate deve essere uguale a

$$\binom{N}{N_1}\binom{N-N_1}{N_2}\binom{N-N_1-N_2}{N_3}\cdots\binom{N-N_1-\cdots-N_{M-1}}{N_M} =$$

$$= \frac{N!}{N_1!\,(N-N_1)!}\cdot\frac{(N-N_1)!}{N_2!\,(N-N_1-N_2)!}\cdots\frac{(N-N_1-\cdots-N_{M-1})!}{N_M!\,(N-N_1-\cdots-N_M)!} =$$

$$= \frac{N!}{N_1!\,N_2!\cdots N_M!}$$

(sfruttando il fatto che tutti i numeratori dei termini dal secondo in poi si semplificano con uno dei fattori del denominatore del termine precedente; inoltre, nell'ultimo termine, $N-N_1-\cdots-N_M \equiv 0$). Si può notare che l'ultimo termine della prima espressione, essendo $N-N_1-\cdots-N_{M-1} = N_M$, vale sempre uno; cosa non sorprendente visto che, quando i primi $M-1$ gruppi sono stati scelti, anche l'ultimo risulta univocamente determinato.

Insomma il numero delle partizioni ordinate è uguale al numero delle permutazioni con ripetizione di N oggetti raggruppabili in M insiemi, composti rispettivamente da $N_1, N_2, \ldots, N_M$ oggetti indistinguibili tra loro, dato dalla formula (A.2)

Appendice B

L'errore della varianza

Può a volte essere utile valutare l'errore della stima della varianza ricavata da un campione di dati sperimentali. Facendo un esempio concreto, supponiamo di disporre di un ampio insieme di valutazioni della stessa grandezza fisica: $N \cdot M$ misure ripetute $x_1, x_2, \ldots, x_{N \cdot M}$. Dividiamo questi valori in M sottoinsiemi costituiti da N dati ciascuno, e per ognuno di questi M sottocampioni calcoliamo la media aritmetica dei dati; otterremo così M medie parziali, che indicheremo con i simboli $\bar{x}_1, \ldots, \bar{x}_M$.

Lo scopo di queste operazioni può essere quello di verificare che le medie di questi sottocampioni sono distribuite su un intervallo di valori più ristretto di quello su cui si distribuisce l'insieme dei dati originali: in sostanza, per verificare che le medie di N dati hanno errore quadratico medio inferiore a quello dei dati di partenza.

L'errore delle medie dei sottocampioni può essere stimato sperimentalmente calcolandone la varianza:

$$\sigma_{\bar{x}}{}^2 = \frac{1}{M-1} \sum_{i=1}^{M} \left(\bar{x}_i - \langle \bar{x} \rangle \right)^2 \qquad \textit{(sperimentale)}$$

intendendo con $\langle \bar{x} \rangle$ la media delle M medie parziali, che coinciderà necessariamente con la media complessiva dell'intero campione di $N \cdot M$ dati.

Questo valore può essere poi confrontato con quello previsto dalla teoria per la varianza della media di un gruppo di dati, allo scopo di verificare in pratica l'adeguatezza della teoria stessa; tale previsione teorica è come sappiamo data dal rapporto tra la varianza di ognuno dei dati che

contribuiscono alla media ed il numero dei dati stessi:

$$\sigma_{\bar{x}}{}^2 = \frac{\sigma^2}{N} \qquad\qquad \textit{(teorico)} \ .$$

Come stima di σ si può usare l'errore quadratico medio dell'insieme di tutti gli $N \cdot M$ dati; ma, naturalmente, perché il confronto tra questi due numeri abbia un significato, *occorre conoscere gli errori* da cui sia la valutazione sperimentale che la previsione teorica di $\sigma_{\bar{x}}$ sono affette.

Consideriamo (come già fatto precedentemente) una popolazione a media zero per semplificare i calcoli:

$$E(x) \equiv x^* = 0 \ ;$$

i risultati si potranno in seguito facilmente estendere ad una popolazione qualsiasi, tenendo presente il teorema di pagina 52 ed i ragionamenti conseguenti. La varianza di una qualsiasi variabile casuale x, indicata di seguito come $\mathrm{Var}(x)$, si può scrivere come

$$\mathrm{Var}(x) = E(x^2) - [E(x)]^2$$

e, usando questa formula per calcolare la varianza della varianza di un campione di N misure s^2, avremo

$$\mathrm{Var}(s^2) = E(s^4) - \left[E(s^2)\right]^2 \ .$$

Ora

$$\begin{aligned} s^4 &= \left[\frac{\sum_i x_i{}^2}{N} - \left(\frac{\sum_i x_i}{N}\right)^2\right]^2 \\ &= \frac{1}{N^2}\left(\sum_i x_i{}^2\right)^2 - \frac{2}{N^3}\left(\sum_i x_i{}^2\right)\left(\sum_i x_i\right)^2 + \frac{1}{N^4}\left(\sum_i x_i\right)^4 \ . \end{aligned}$$

Sviluppiamo uno per volta i tre termini a secondo membro; per il primo risulta

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