Alcune definizioni e risultati di teoria della probabilità e teoria statistica

1. Definizione. Variabili gaussiane (o normali)

Una variabile casuale continua $x \in \mathbb{R}$ si dice **gaussiana** o **normale** se la sua distribuzione di probabilità è della forma:

$$p(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2} \qquad x \in \mathbb{R} \ ,$$

con μ e $\sigma > 0$ costanti reali arbitrarie. La prima di tali costanti è identificabile con il valor medio della variabile casuale:

$$\mu = \mathbb{E}(x) = \int_{-\infty}^{+\infty} p(x) x \, dx = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2} x \, dx$$

mentre σ^2 coincide con la varianza — e quindi σ con la deviazione standard corrispondente:

$$\sigma^2 = \mathbb{E}[(x-\mu)^2] = \int_{-\infty}^{+\infty} p(x)(x-\mu)^2 \, dx = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2} (x-\mu)^2 \, dx \ .$$

La variabile gaussiana si dice **standard** se a media nulla e varianza unitaria; in tal caso la distribuzione di probabilità si riduce all'espressione:

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \ .$$

2. Definizione. Variabili gaussiane (o normali) multivariate

Le variabili casuali

$$(x_1, x_2, \ldots, x_n) = x \in \mathbb{R}^n$$

costituiscono un insieme di n variabili gaussiane multivariate se la loro distribuzione di probabilità congiunta è data dall'espressione

$$p(x) = p(x_1, x_2, \ldots, x_n) = \frac{|A|^{1/2}}{(2\pi)^{n/2}} e^{-\frac{1}{2}(x-\mu)^T A (x-\mu)} \qquad (2.1)$$

essendo A una matrice $n \times n$ reale simmetrica definita positiva, $|A| > 0$ il suo determinante e $\mu \in \mathbb{R}^n$ un vettore arbitrario. Le variabili gaussiane multivariate si dicono **indipendenti** se la matrice A risulta diagonale, concordemente con la definizione generale che individua come indipendenti quelle variabili casuali la cui distribuzione di probabilità congiunta si fattorizza in un prodotto di termini ciascuno dipendente da una sola variabile. La distribuzione (2.1) può intendersi come densità della corrispondente misura di probabilità m della variabile casuale x. Dato un qualsiasi insieme misurabile secondo Lebesgue, $\Omega \subseteq \mathbb{R}^n$, la probabilità di una realizzazione $x \in \Omega$ si scrive

$$m(\Omega) = \int_{\Omega} p(x) \, dx = \int_{\Omega} \frac{|A|^{1/2}}{(2\pi)^{n/2}} e^{-\frac{1}{2}(x-\mu)^T A (x-\mu)} \, dx \ .$$

3. Normalizzazione

La distribuzione di probabilità (2.1) *soddisfa la condizione di normalizzazione*

$$\mathbb{E}(1) = \int_{\mathbb{R}^n} p(x)\, dx = \int_{\mathbb{R}^n} p(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \ldots dx_n = 1 \,. \tag{3.1}$$

Dimostrazione

Nell'integrale

$$\mathbb{E}(1) = \int_{\mathbb{R}^n} \frac{|A|^{1/2}}{(2\pi)^{n/2}}\, e^{-\frac{1}{2}(x-\mu)^T A (x-\mu)}\, dx$$

si esegua il cambiamento di variabili $x = \mu + \xi$, in modo che

$$\mathbb{E}(1) = \int_{\mathbb{R}^n} \frac{|A|^{1/2}}{(2\pi)^{n/2}}\, e^{-\frac{1}{2}\xi^T A \xi}\, d\xi \,.$$

Poiché la matrice A è per ipotesi reale simmetrica e definita positiva, è sempre possibile determinare una matrice ortogonale R ed una matrice diagonale reale ad elementi diagonali positivi Λ tali che

$$\Lambda = R^T A R = R^{-1} A R \,, \quad \Lambda = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \mathbb{O} \\ & & \ddots & \\ & \mathbb{O} & & \lambda_n \end{pmatrix} \,, \quad 0 < \lambda_i \,,\ 1 \le i \le n \,.$$

Si può così introdurre l'ulteriore cambiamento di variabili $\xi = Rz$, ottenendo

$$\mathbb{E}(1) = \int_{\mathbb{R}^n} \frac{|A|^{1/2}}{(2\pi)^{n/2}}\, e^{-\frac{1}{2}z^T R^T A R z} \left|\frac{\partial \xi}{\partial z}\right| dz = \int_{\mathbb{R}^n} \frac{|A|^{1/2}}{(2\pi)^{n/2}}\, e^{-\frac{1}{2}z^T \Lambda z} |R|\, dz =$$

$$= \int_{\mathbb{R}^n} \frac{|A|^{1/2}}{(2\pi)^{n/2}}\, e^{-\frac{1}{2}z^T \Lambda z}\, dz = \frac{|A|^{1/2}}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{-\frac{1}{2}\sum_{i=1}^n \lambda_i z_i^2}\, dz \tag{3.2}$$

in cui

$$\left|\frac{\partial \xi}{\partial z}\right| = \big||R|\big| = 1$$

è il valore assoluto del determinante jacobiano della trasformazione. L'ultimo membro in (3.2) per Fubini assume la forma

$$\mathbb{E}(1) = \frac{|A|^{1/2}}{(2\pi)^{n/2}} \prod_{i=1}^n \int_{\mathbb{R}} e^{-\frac{1}{2}\lambda_i z_i^2}\, dz_i$$

e si integra esplicitamente porgendo

$$\mathbb{E}(1) \;=\; \frac{|A|^{1/2}}{(2\pi)^{n/2}} \prod_{i=1}^{n} \sqrt{\frac{2\pi}{\lambda_i}} \;=\; \frac{|A|^{1/2}}{(2\pi)^{n/2}} \frac{(2\pi)^{n/2}}{\sqrt{\prod_{i=1}^{n} \lambda_i}} \;=\; 1 \;,$$

in cui si è fatto uso della ben nota espressione

$$\int_{\mathbb{R}} e^{-az^2}\, dz \;=\; \sqrt{\pi/a} \qquad \forall\, a \in \mathbb{R}_+ \tag{3.3}$$

e dell'ovvia uguaglianza

$$\prod_{i=1}^{n} \lambda_i \;=\; |\Lambda| \;=\; |R^T A R| \;=\; |R^{-1} A R| \;=\; |R^{-1}||A||R| \;=\; |R|^{-1}|A||R| \;=\; |A| \;. \qquad \square$$

4. Valor medio e matrice di covarianza di variabili gaussiane multivariate

Sia dato un insieme di n variabili gaussiane multivariate con distribuzione di probabilità (2.1). Il valor medio $\mathbb{E}(x_i)$ della i-esima variabile e la covarianza $\mathbb{E}\big[(x_i - \mathbb{E}(x_i))(x_j - \mathbb{E}(x_j))\big]$ fra le variabili x_i ed x_j sono allora espressi da

$$\mathbb{E}(x_i) \;=\; \mu_i \qquad \mathbb{E}\big[(x_i - \mathbb{E}(x_i))(x_j - \mathbb{E}(x_j))\big] \;=\; C_{ij} \;=\; (A^{-1})_{ij} \qquad 1 \le i, j \le n \,, \tag{4.1}$$

dove $C = A^{-1}$ è la matrice di covarianza dell'insieme di variabili. In particolare, le variabili gaussiane multivariate sono indipendenti se e soltanto se scorrelate — vale a dire, se e solo se la matrice di covarianza C è diagonale.

Dimostrazione

Il valor medio della variabile x_i è dato dalla relazione

$$\mathbb{E}(x_i) \;=\; \int_{\mathbb{R}^n} \frac{|A|^{1/2}}{(2\pi)^{n/2}}\, e^{-\frac{1}{2}(x-\mu)^T A (x-\mu)}\, x_i \, dx$$

che con il cambiamento di variabili $x = \mu + \xi$ diviene

$$\mathbb{E}(x_i) \;=\; \int_{\mathbb{R}^n} \frac{|A|^{1/2}}{(2\pi)^{n/2}}\, e^{-\frac{1}{2}\xi^T A \xi}\, (\mu_i + \xi_i)\, d\xi \;=\; \mu_i + 0 \;=\; \mu_i \;.$$

Le covarianze diventano perciò, indicata con C la matrice di covarianza del sistema di variabili gaussiane multivariate,

$$\mathbb{E}\big[(x_i - \mu_i)(x_j - \mu_j)\big] \;=\; C_{ij} \;=\; \int_{\mathbb{R}^n} \frac{|A|^{1/2}}{(2\pi)^{n/2}}\, e^{-\frac{1}{2}(x-\mu)^T A (x-\mu)}\, (x_i - \mu_i)(x_j - \mu_j)\, dx \;=$$

$$= \int_{\mathbb{R}^n} \frac{|A|^{1/2}}{(2\pi)^{n/2}}\, e^{-\frac{1}{2}\xi^T A\xi}\, \xi_i \xi_j \; d\xi \; . \tag{4.2}$$

L'espressione (4.2) si semplifica anche in questo caso per mezzo del cambiamento lineare di variabili $\xi = Rz$, con $R^T = R^{-1}$ e $R^T AR = \Lambda$,

$$\begin{aligned} C_{ij} &= \int_{\mathbb{R}^n} \frac{|A|^{1/2}}{(2\pi)^{n/2}}\, e^{-\frac{1}{2}z^T R^T ARz} \sum_{a=1}^{n} R_{ia} z_a \sum_{b=1}^{n} R_{jb} z_b \, \big||R|\big| \; dz = \\ &= \frac{|A|^{1/2}}{(2\pi)^{n/2}} \sum_{a,b=1}^{n} R_{ia}R_{jb} \int_{\mathbb{R}^n} e^{-\frac{1}{2}z^T\Lambda z} z_a z_b \; dz = \\ &= \frac{|A|^{1/2}}{(2\pi)^{n/2}} \sum_{a,b=1}^{n} R_{ia}R_{jb}\delta_{ab} \Bigg[\prod_{k=1,(k\neq a)}^{n} \int_{\mathbb{R}} e^{-\frac{1}{2}\lambda_k z_k^2} \; dz_k \Bigg] \int_{\mathbb{R}} e^{-\frac{1}{2}\lambda_a z_a^2} z_a^2 \, dz_a \; . \end{aligned} \tag{4.3}$$

Siccome poi dall'equazione (3.3) si deduce

$$\int_{\mathbb{R}} e^{-az^2} z^2 \; dz = -\frac{d}{da} \int_{\mathbb{R}} e^{-az^2} \; dz = \frac{\sqrt{\pi}}{2} a^{-3/2} \qquad \forall\, a \in \mathbb{R}_+$$

e quindi

$$\int_{\mathbb{R}} e^{-\frac{1}{2}\lambda_a z_a^2} z_a^2 \, dz_a = \frac{\sqrt{\pi}}{2} \Big(\frac{\lambda_a}{2}\Big)^{-3/2} = \frac{(2\pi)^{1/2}}{\lambda_a^{3/2}} \; ,$$

sostituendo in (4.3) segue che

$$\begin{aligned} C_{ij} &= \frac{|A|^{1/2}}{(2\pi)^{n/2}} \sum_{a,b=1}^{n} R_{ia}R_{jb}\, \delta_{ab}\, (2\pi)^{(n-1)/2} \sqrt{\frac{\lambda_a}{\prod_{k=1}^{n} \lambda_k}}\, \frac{(2\pi)^{1/2}}{\lambda_a^{3/2}} = \\ &= \frac{|A|^{1/2}}{(2\pi)^{n/2}} \sum_{a,b=1}^{n} R_{ia}R_{jb}\, \delta_{ab}\, (2\pi)^{n/2} \frac{1}{\lambda_a} \frac{1}{\sqrt{|A|}} = \\ &= \sum_{a,b=1}^{n} R_{ia}R_{jb}\, \delta_{ab} \frac{1}{\lambda_a} = \sum_{a,b=1}^{n} R_{ia}R_{jb}\, (\Lambda^{-1})_{ab} = (R\Lambda^{-1}R^T)_{ij} \; . \end{aligned}$$

La matrice di covarianza diventa così

$$C = R\Lambda^{-1}R^T = R\Lambda^{-1}R^{-1} = R\big(R^T AR\big)^{-1} R^{-1} = RR^{-1}A^{-1}RR^{-1} = A^{-1}$$

come si voleva dimostrare. In considerazione del fatto che la matrice $C = A^{-1}$ risulta diagonale se e soltanto se A lo è a propria volta, si conclude che le condizione di indipendenza stocastica delle variabili gaussiane multivariate $x_1, x_2, \ldots, x_n$ e l'assenza di correlazione fra le stesse sono proprietà equivalenti. Si ricorda che in generale l'indipendenza stocastica di due o più variabili casuali è una condizione *più forte* e non equivalente alla mancanza di correlazione, per cui variabili indipendenti sono sempre scorrelate, ma non viceversa. $\square$

5. Combinazioni lineari di variabili gaussiane multivariate

Si consideri un insieme di variabili gaussiane multivariate $x = (x_1, x_2, \ldots, x_n)$ *con distribuzione di probabilità* (2.1). *Data una qualsiasi matrice* $n \times n$ R*, reale e non singolare, l'insieme di variabili casuali definito da* $y = Rx$ *costituisce ancora un insieme di variabili gaussiane multivariate di distribuzione*

$$\hat{p}(y) \;=\; \frac{\big|(R^{-1})^T \, A \, R^{-1}\big|^{1/2}}{(2\pi)^{n/2}} \, e^{-\frac{1}{2}(y - R\mu)^T (R^{-1})^T A R^{-1} (y - R\mu)} \;.$$

Dimostrazione

Dalla distribuzione di probabilità (2.1), con il cambiamento di variabili definito da $y = Rx$ si ricava:

$$\hat{p}(y) \;=\; \frac{|A|^{1/2}}{(2\pi)^{n/2}} \, e^{-\frac{1}{2}(R^{-1}y - \mu)^T A (R^{-1}y - \mu)} \left|\frac{\partial x}{\partial y}\right| =$$

$$= \frac{|A|^{1/2}}{(2\pi)^{n/2}} \, e^{-\frac{1}{2}(y - R\mu)^T (R^{-1})^T A R^{-1} (y - R\mu)} \, \big||R^{-1}|\big|$$

ed essendo

$$|A|^{1/2}\big||R^{-1}|\big| \;=\; \big||R^{-1}|\big|^{1/2}|A|^{1/2}\big||R^{-1}|\big|^{1/2} \;=\; \big||(R^{-1})^T|\big|^{1/2}|A|^{1/2}\big||R^{-1}|\big|^{1/2} \;=$$

$$= \Big(\big||(R^{-1})^T|\big| \, |A| \, \big||R^{-1}|\big|\Big)^{1/2} \;=\; \big(|(R^{-1})^T| \, |A| \, |R^{-1}|\big)^{1/2} \;=\; \big|(R^{-1})^T \, A \, R^{-1}\big|^{1/2}$$

ne deriva che

$$\hat{p}(y) \;=\; \frac{\big|(R^{-1})^T \, A \, R^{-1}\big|^{1/2}}{(2\pi)^{n/2}} \, e^{-\frac{1}{2}(y - R\mu)^T (R^{-1})^T A R^{-1} (y - R\mu)} \;.$$

Il risultato segue immediatamente per il fatto che la matrice $(R^{-1})^T \, A \, R^{-1}$ è reale simmetrica e definita positiva al pari di A, mentre il vettore dei valori medi per la variabile y risulta $R\mu$. □

6. Distribuzione di probabilità di una singola variabile casuale che sia combinazione lineare di variabili gaussiane multivariate

Sia $x = (x_1, x_2, \ldots, x_n)$ *un sistema di variabili gaussiane multivariate con distribuzione di probabilità* (2.1). *Allora per un qualunque vettore* $t \in \mathbb{R}^n \setminus \{0\}$ *fissato, la variabile casuale*

$$z \;=\; \langle t|x\rangle \;=\; \sum_{i=1}^{n} t_i x_i$$

è gaussiana con distribuzione di probabilità

$$p(z) \;=\; \frac{1}{\sqrt{2\pi}\sigma} \, e^{-(z - \langle t|\mu\rangle)^2/2\sigma^2} \;,$$

essendo $\sigma^2 = \langle t|Ct\rangle = \langle t|A^{-1}t\rangle > 0.$

Dimostrazione
Poiché $t \neq 0$, è sempre possibile determinare una matrice R $n \times n$ tale che

$$|R| \neq 0 \qquad \text{e} \qquad R_{1i} = t_i \qquad 1 \leq i \leq n \;.$$

In forza del risultato precedente $z = y_1$ è allora una variabile gaussiana dell'insieme $y = Rx$. Si tratta di determinarne la distribuzione di probabilità. In primo luogo, è chiaro che per definizione il valor medio di z deve essere dato dall'espressione

$$\mathbb{E}(z) = \mathbb{E}\Big(\sum_{i=1}^{n} t_i x_i\Big) = \sum_{i=1}^{n} t_i \mathbb{E}(x_i) = \sum_{i=1}^{n} t_i \mu_i = \langle t|\mu\rangle \;, \tag{6.1}$$

e la varianza da

$$\mathbb{E}\big[(z - \langle t|\mu\rangle)^2\big] = \mathbb{E}\big[\langle t|x-\mu\rangle^2\big] = \sum_{i,j=1}^{n} t_i t_j \mathbb{E}\big[(x_i-\mu_i)(x_j-\mu_j)\big] = \sum_{i,j=1}^{n} t_i t_j C_{ij} \;,$$

mentre la distribuzione di probabilità della $z = y_1$ si ottiene integrando la distribuzione congiunta $\hat{p}(y)$ sulle variabili residue $y_2, \ldots, y_n$

$$p(y_1) = \frac{\big|(R^{-1})^T A\, R^{-1}\big|^{1/2}}{(2\pi)^{n/2}} \int_{\mathbb{R}^{n-1}} e^{-\frac{1}{2}(y-R\mu)^T (R^{-1})^T A R^{-1}(y-R\mu)}\, dy_2\, dy_3 \ldots dy_n \;.$$

Ci si può ricondurre ad una variabile casuale di media nulla ponendo $y_1 = z_1 + \langle t|\mu\rangle = z_1 + (R\mu)_1$, per cui

$$\tilde{p}(z_1) = p(z_1 + \langle t|\mu\rangle) = p(z_1 + (R\mu)_1) = \frac{\big|(R^{-1})^T A\, R^{-1}\big|^{1/2}}{(2\pi)^{n/2}} \cdot$$

$$\cdot \int_{\mathbb{R}^{n-1}} e^{-\frac{1}{2}(y-R\mu)^T (R^{-1})^T A R^{-1}(y-R\mu)}\Big|_{y_1 = z_1 + (R\mu)_1} dy_2\, dy_3 \ldots dy_n \;,$$

relazione che si semplifica apprezzabilmente per mezzo dell'ulteriore cambiamento di variabili $y_i = z_i + (R\mu)_i$, $2 \leq i \leq n$:

$$\tilde{p}(z_1) = \frac{\big|(R^{-1})^T A\, R^{-1}\big|^{1/2}}{(2\pi)^{n/2}} \int_{\mathbb{R}^{n-1}} e^{-\frac{1}{2} z^T (R^{-1})^T A R^{-1} z}\, dz_2\, dz_3 \ldots dz_n \;. \tag{6.2}$$

Posto per brevità $B = (R^{-1})^T A R^{-1}$, la (6.2) diventa

$$\tilde{p}(z_1) = \frac{|B|^{1/2}}{(2\pi)^{n/2}} \int_{\mathbb{R}^{n-1}} e^{-\frac{1}{2}\sum_{i,j=1}^{n} B_{ij} z_i z_j}\, dz_2\, dz_3 \ldots dz_n$$

e si può riscrivere nella forma equivalente

$$\tilde{p}(z_1) = \frac{|B|^{1/2}}{(2\pi)^{n/2}} \int\limits_{\mathbb{R}^{n-1}} e^{-\frac{1}{2}\left[B_{11}z_1^2 + 2\sum_{i=2}^n B_{1i}z_iz_1 + \sum_{i,j=2}^n B_{ij}z_iz_j\right]} dz_2 \ldots dz_n \tag{6.3}$$

allo scopo di evidenziare i termini contenenti la variabile z_1. Nel seguito della dimostrazione conviene inoltre introdurre le notazioni sotto riportate

$$z' = (z_2 \ldots z_n)^T \qquad B' = (B)_{ij=2}^n \qquad h_i = z_1 B_{1i}\ , \quad 2 \le i \le n\ ,$$

sicché risulta

$$2\sum_{i=2}^n B_{1i}z_iz_1 + \sum_{i,j=2}^n B_{ij}z_iz_j = 2h^Tz' + z'^TB'z'\ . \tag{6.4}$$

Conviene introdurre il cambiamento di variabile $z' = w' + a'$, con $a' \in \mathbb{R}^{n-1}$ vettore costante da determinare opportunamente come verrà descritto nel seguito. Questa sostituzione nella (6.4) porge:

$$\begin{aligned} 2h^Tz' + z'^TB'z' &= 2h^Ta' + 2h^Tw' + (w'+a')^T(B'w' + B'a') = \\ &= 2h^Ta' + 2h^Tw' + w'^TB'w' + w'^TB'a' + a'^TB'w' + a'^TB'a' = \\ &= w'^TB'w' + 2\big(h^T + a'^TB'\big)w' + 2h^Ta' + a'^TB'a' = \\ &= w'^TB'w' + 2\big(h + B'a'\big)^T w' + 2h^Ta' + a'^TB'a' \end{aligned}$$

e se si fissa a' in modo che $h + B'a' = 0$, si pone cioè $a' = B'^{-1}h$(^{o}), l'espressione (6.4) viene ricondotta alla forma più semplice

$$2h^Tz' + z'^TB'z' = w'^TB'w' + 2h^Ta' + a'^TB'a'\ .$$

Per effetto della precedente sostituzione, l'espressione (6.3) per la distribuzione di probabilità di z_1 diventa pertanto

$$\tilde{p}(z_1) = \frac{|B|^{1/2}}{(2\pi)^{n/2}} e^{-\frac{1}{2}B_{11}z_1^2} \int\limits_{\mathbb{R}^{n-1}} e^{-\frac{1}{2}\left[w'^TB'w' + 2h^Ta' + a'^TB'a'\right]} dw_2' \ldots dw_n'$$

ed eseguendo l'integrale multiplo, che è ora di una forma nota, già trattata in precedenza al punto 2,

$$\tilde{p}(z_1) = \frac{|B|^{1/2}}{(2\pi)^{n/2}} e^{-h^Ta' - \frac{1}{2}a'^TB'a'} \cdot \frac{(2\pi)^{(n-1)/2}}{|B'|^{1/2}} e^{-\frac{1}{2}B_{11}z_1^2} =$$

(^{o}) L'invertibilità di B' segue immediatamente da quella di B, per ipotesi definita positiva. Per ogni $z \in \mathbb{R}^n \setminus \{0\}$ vale infatti $z^TBz > 0$, ovvero $B_{11}z_1^2 + 2\sum_{i=2}^n B_{1i}z_iz_1 + z'^TB'z' > 0$. Se per assurdo B' fosse non invertibile ed esistesse quindi un $\overline{z}' \neq 0$, di componenti $\overline{z}_2, \ldots, \overline{z}_n$, tale che $B'\overline{z}' = 0$, allora $\overline{z}^T = (z_1\ \overline{z}'^T) \neq 0$ dovrebbe soddisfare $\overline{z}^TB\overline{z} = B_{11}z_1^2 + 2\sum_{i=2}^n B_{1i}\overline{z}_iz_1 > 0$, $\forall\, z_1 \in \mathbb{R}$, risultato assurdo in quanto la diseguaglianza non è verificata per $z_1 = 0$.

$$= \frac{|B|^{1/2}}{\sqrt{2\pi}\,|B'|^{1/2}}\; e^{-h^T a' - \frac{1}{2}a'^T B' a'}\; e^{-\frac{1}{2}B_{11}z_1^2} \,. \tag{6.5}$$

D'altra parte, $h = -B'a'$ implica $h^T a' = -a'^T B' a'$ e quindi

$$-h^T a' - \frac{1}{2}a'^T B' a' \;=\; a'^T B' a' - \frac{1}{2}a'^T B' a' \;=\; \frac{1}{2}a'^T B' a' \;=\; \frac{1}{2}h^T B'^{-1} h \,.$$

Sostituendo nella (6.5) si ottiene perciò:

$$\tilde{p}(z_1) \;=\; \frac{|B|^{1/2}}{\sqrt{2\pi}\,|B'|^{1/2}}\; e^{\frac{1}{2}h^T B'^{-1} h}\; e^{-\frac{1}{2}B_{11}z_1^2}$$

dove

$$h^T B'^{-1} h \;=\; \sum_{i,j=2}^{n} h_i h_j (B'^{-1})_{ij} \;=\; \sum_{i,j=2}^{n} B_{1i}(B'^{-1})_{ij} B_{j1} z_1^2$$

e di conseguenza

$$h^T B'^{-1} h - B_{11}z_1^2 \;=\; \Big[-B_{11} + \sum_{i,j=2}^{n} B_{1i}(B'^{-1})_{ij} B_{j1}\Big]\, z_1^2 \,,$$

in modo che

$$\tilde{p}(z_1) \;=\; \frac{|B|^{1/2}}{\sqrt{2\pi}\,|B'|^{1/2}}\; e^{-\frac{1}{2}\big[B_{11} - \sum_{i,j=2}^{n} B_{1i}(B'^{-1})_{ij}B_{j1}\big] z_1^2} \,.$$

La distribuzione di probabilità ottenuta può semplificarsi ulteriormente osservando che per il teorema dell'aggiunto classico — o cofattore —(^{o}) vale l'identità

$$(B^{-1})_{11} \;=\; \frac{|B'|}{|B|}(-1)^{1+1} \;=\; \frac{|B'|}{|B|}$$

la quale porge così

$$\tilde{p}(z_1) \;=\; \frac{1}{\sqrt{2\pi\,(B^{-1})_{11}}}\; e^{-\frac{1}{2}\big[B_{11} - \sum_{i,j=2}^{n} B_{1i}(B'^{-1})_{ij}B_{j1}\big] z_1^2} \,. \tag{6.6}$$

Quest'ultima espressione non ha ancora un aspetto molto familiare e l'argomento dell'esponenziale richiede ancora qualche manipolazione algebrica. Indicato con $\mathrm{cof}_{ij}(B)$ il cofattore della matrice B relativo all'elemento B_{ij}, dall'esame delle matrici B e B' è facile ricavare la relazione di ricorrenza sotto riportata

$$\sum_{j=2}^{n} \mathrm{cof}_{ij}(B')\, B_{j1} \;=\; -\mathrm{cof}_{1i}(B) \,, \quad 2 \le i \le n \,, \qquad \begin{pmatrix} B_{11} & & & & \\ & B_{22} & B_{23} & \dots & B_{2n} \\ & B_{32} & B_{33} & \dots & B_{3n} \\ & \vdots & \vdots & \ddots & \vdots \\ & B_{n2} & B_{n3} & \dots & B_{nn} \end{pmatrix}$$

(^{o}) la regola usualmente applicata per la determinazione analitica della matrice inversa unitamente alla più ovvia identità $\mathrm{cof}_{11}(B) = |B'|$. Grazie a queste relazioni, ricordando che $|B'|(B'^{-1})_{ij} = \mathrm{cof}_{ji}(B')$, $2 \leq i,j \leq n$, e che $B'^T = B'$, è dato scrivere:

$$B_{11} - \sum_{i,j=2}^{n} B_{1i}(B'^{-1})_{ij}B_{j1} = \frac{1}{|B'|}\Big[B_{11}|B'| - \sum_{i,j=2}^{n} B_{1i}|B'|(B'^{-1})_{ij}B_{j1}\Big] =$$

$$= \frac{1}{|B'|}\Big[B_{11}|B'| - \sum_{i,j=2}^{n} B_{1i}\,\mathrm{cof}_{ji}(B')B_{j1}\Big] = \frac{1}{|B'|}\Big[B_{11}|B'| - \sum_{i,j=2}^{n} B_{1i}\,\mathrm{cof}_{ij}(B')B_{j1}\Big] =$$

$$= \frac{1}{|B'|}\Big[B_{11}|B'| - \sum_{i=2}^{n} B_{1i}\sum_{j=2}^{n}\mathrm{cof}_{ij}(B')B_{j1}\Big] = \frac{1}{|B'|}\Big[B_{11}\mathrm{cof}_{11}(B) - \sum_{i=2}^{n} B_{1i}\big[-\mathrm{cof}_{1i}(B)\big]\Big] =$$

$$= \frac{1}{|B'|}\Big[B_{11}\mathrm{cof}_{11}(B) + \sum_{i=2}^{n} B_{1i}\,\mathrm{cof}_{1i}(B)\Big],$$

in cui l'espressione finale entro parentesi quadre non è altro che lo sviluppo di Laplace del determinante di B rispetto alla sua prima riga. Pertanto

$$B_{11} - \sum_{i,j=2}^{n} B_{1i}(B'^{-1})_{ij}B_{j1} = \frac{1}{|B'|}\,|B| = \frac{1}{(B^{-1})_{11}}$$

e la (6.6) si riduce a

$$\tilde{p}(z_1) = \frac{1}{\sqrt{2\pi\,(B^{-1})_{11}}}\; e^{-z_1^2/2(B^{-1})_{11}}$$

in modo che

$$p(z) = p(y_1) = \tilde{p}(y_1 - \langle t|\mu\rangle) = \tilde{p}(z - \langle t|\mu\rangle) = \frac{1}{\sqrt{2\pi\,(B^{-1})_{11}}}\; e^{-(z - \langle t|\mu\rangle)/2(B^{-1})_{11}}\,.$$

Per concludere non resta che dimostrare la relazione $(B^{-1})_{11} = \langle t|Ct\rangle$. In effetti, nelle precedenti ipotesi risulta $B = (R^{-1})^T A R^{-1}$ con $R_{1i} = t_i$, $1 \leq i \leq n$, per cui $B^{-1} = RA^{-1}R^T$ e

$$(B^{-1})_{11} = \sum_{i,j=1}^{n} R_{1i}(A^{-1})_{ij}(R^T)_{j1} = \sum_{i,j=1}^{n} R_{1i}(A^{-1})_{ij}R_{1j} = \sum_{i,j=1}^{n} t_i(A^{-1})_{ij}t_j = \sum_{i,j=1}^{n} t_i C_{ij} t_j$$

che si identifica con la varianza σ^2 di z, come deve essere. $\square$

7. Definizione. Variabile di $\mathcal{X}^2$ ad n gradi di libertà

La variabile casuale

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

si dice una variabile di chi-quadrato — $\mathcal{X}^2$ — ad n gradi di libertà se le $(x_1, x_2, \ldots, x_n)$ costituiscono un sistema di n variabili gaussiane stocasticamente indipendenti e standard — cioè a dire, di media nulla e varianza unitaria, caratterizzate dunque da una distribuzione di probabilità della forma (2.1), con $\mu = 0$ ed $A = \mathbb{I}$.

8. Distribuzione di probabilità di una variabile di $\mathcal{X}^2$ ad n gradi di libertà

La distribuzione di probabilità di una variabile di $\mathcal{X}^2$ a n gradi di libertà è data da

$$p_n(\mathcal{X}^2) = \frac{1}{\Gamma(n/2)2^{n/2}} e^{-\mathcal{X}^2/2} (\mathcal{X}^2)^{\frac{n}{2}-1}$$

in cui si denota con $\Gamma(\alpha)$ la funzione gamma di Eulero calcolata in $\alpha \in \mathbb{R}^+$

$$\Gamma(\alpha) = \int_0^{+\infty} e^{-x} x^{\alpha-1}\, dx$$

Dimostrazione

Si parte dalla condizione di normalizzazione della distribuzione di probabilità per un sistema di n variabili gaussiane indipendenti e standard $x_1, \ldots, x_n$

$$1 = \int_{\mathbb{R}^n} e^{-\frac{1}{2}\sum_{i=1}^n x_i^2} dx_1 \ldots dx_n$$

in cui si introduce il cambiamento di variabili

$$\begin{aligned}
x_1 &= \mathcal{X} \cos\phi_1 & \qquad \mathcal{X} &= \Big[\sum_{i=1}^n x_i^2\Big]^{1/2} \in \mathbb{R}^+ \\
x_2 &= \mathcal{X} \sin\phi_1 \cos\phi_2 & & \\
x_3 &= \mathcal{X} \sin\phi_1 \sin\phi_2 \cos\phi_3 & \phi_1 &\in [0,\pi] \\
x_4 &= \mathcal{X} \sin\phi_1 \sin\phi_2 \sin\phi_3 \cos\phi_4 & \phi_2 &\in [0,\pi] \\
&\ldots\ldots\ldots\ldots\ldots\ldots & &\ldots\ldots \\
x_n &= \mathcal{X} \sin\phi_1 \sin\phi_2 \sin\phi_3 \ldots \sin\phi_{n-1} & \phi_{n-1} &\in [0,2\pi]
\end{aligned}$$

essendo $(\mathcal{X}, \phi_1, \ldots, \phi_{n-1})$ un set di coordinate polari n-dimensionali. Anche senza ricorrere al calcolo esplicito, è comunque evidente che il determinante jacobiano della trasformazione di coordinate assume la forma

$$\mathcal{X}^{n-1}\Omega_n(\phi_1, \ldots, \phi_{n-1})$$

in modo che la condizione di normalizzazione diventa

$$1 = \int_{\mathbb{R}^+} d\mathcal{X} \int_0^{\pi} d\phi_1 \ldots \int_0^{2\pi} d\phi_{n-1} e^{-\frac{1}{2}\mathcal{X}^2} \mathcal{X}^{n-1} \Omega_n(\phi_1, \ldots, \phi_{n-1}) =$$

$$= \int_0^{\pi} d\phi_1 \ldots \int_0^{2\pi} d\phi_{n-1} \Omega_n(\phi_1, \ldots, \phi_{n-1}) \int_{\mathbb{R}^+} e^{-\frac{1}{2}\mathcal{X}^2} \mathcal{X}^{n-1} d\mathcal{X}$$

L'ulteriore cambiamento di variabile

$$\mathcal{X} = (\mathcal{X}^2)^{1/2} \qquad\qquad d\mathcal{X} = \frac{1}{2}(\mathcal{X}^2)^{-1/2}d(\mathcal{X}^2)$$

porge infine

$$\begin{aligned} 1 &= \frac{1}{2}\int_0^{\pi} d\phi_1 \ldots \int_0^{2\pi} d\phi_{n-1}\Omega_n(\phi_1,\ldots,\phi_{n-1}) \int_{\mathbb{R}^+} e^{-\frac{1}{2}\mathcal{X}^2}(\mathcal{X}^2)^{\frac{n}{2}-1}d(\mathcal{X}^2) = \\ &= N_n \int_{\mathbb{R}^+} e^{-\frac{1}{2}\mathcal{X}^2}(\mathcal{X}^2)^{\frac{n}{2}-1}d(\mathcal{X}^2) \end{aligned}$$

con N_n appropriato fattore di normalizzazione. In effetti deve aversi

$$\frac{1}{N_n} = \int_{\mathbb{R}^+} e^{-\frac{1}{2}\mathcal{X}^2}(\mathcal{X}^2)^{\frac{n}{2}-1}d(\mathcal{X}^2)$$

e con la sostituzione $\mathcal{X}^2/2 = \xi$ l'integrale a secondo membro si riduce a

$$2^{\frac{n}{2}}\int_0^{+\infty} e^{-\xi}\xi^{\frac{n}{2}-1}\,d\xi = 2^{n/2}\Gamma(n/2)$$

per cui il fattore di normalizzazione risulta

$$N_n = \frac{1}{2^{n/2}\Gamma(n/2)}$$

e la distribuzione della variabile $\mathcal{X}^2$

$$\frac{1}{2^{n/2}\Gamma(n/2)}e^{-\frac{1}{2}\mathcal{X}^2}(\mathcal{X}^2)^{\frac{n}{2}-1}d(\mathcal{X}^2)$$

con condizione di normalizzazione

$$\int_0^{+\infty}\frac{1}{2^{n/2}\Gamma(n/2)}e^{-\frac{1}{2}\mathcal{X}^2}(\mathcal{X}^2)^{\frac{n}{2}-1}d(\mathcal{X}^2) = 1 \qquad \square$$

9. Media e varianza di una variabile di $\mathcal{X}^2$ a n gradi di libertà

La media e la varianza di una variabile di $\mathcal{X}^2$ a n gradi di libertà sono date dalle relazioni

$$\mu = n \qquad \sigma^2 = 2n$$

Dimostrazione

Per definizione, il valor medio della variabile $\mathcal{X}^2$ a n gradi di libertà si scrive

$$\mu = \mathbb{E}(\mathcal{X}^2) = \int_0^{+\infty} \frac{1}{2^{\frac{n}{2}}\Gamma\Big(\dfrac{n}{2}\Big)} e^{-\frac{1}{2}\mathcal{X}^2} (\mathcal{X}^2)^{\frac{n}{2}-1} \mathcal{X}^2 \, d(\mathcal{X}^2) = \int_0^{+\infty} \frac{1}{2^{\frac{n}{2}}\Gamma\Big(\dfrac{n}{2}\Big)} e^{-\frac{1}{2}u} u^{\frac{n}{2}} \, du$$

e con il cambiamento di variabile $u = 2\xi$ diventa

$$\mu = \frac{2}{\Gamma\Big(\dfrac{n}{2}\Big)} \int_0^{+\infty} e^{-\xi} \xi^{\frac{n}{2}+1-1} d\xi = \frac{2\,\Gamma\Big(\dfrac{n}{2}+1\Big)}{\Gamma\Big(\dfrac{n}{2}\Big)} = 2\,\frac{n}{2} = n$$

in cui si è fatto uso dell'ovvia identità

$$\Gamma(a+1) = \int_0^{+\infty} e^{-\xi}\xi^a d\xi = \Big[-e^{-\xi}\xi^a\Big]_0^{+\infty} + \int_0^{+\infty} e^{-\xi} a \xi^{a-1} d\xi = a\,\Gamma(a) \qquad \forall\, a \in \mathbb{R}^+ .$$

Per il calcolo della varianza si procedere in modo analogo, richiamando la definizione

$$\begin{aligned}
\sigma^2 &= \mathbb{E}[(\mathcal{X}^2 - \mu)^2] = \mathbb{E}[(\mathcal{X}^2 - n)^2] = \int_0^{+\infty} \frac{1}{2^{\frac{n}{2}}\Gamma\Big(\dfrac{n}{2}\Big)} e^{-\frac{1}{2}\mathcal{X}^2} (\mathcal{X}^2)^{\frac{n}{2}-1} (\mathcal{X}^2 - n)^2 \, d(\mathcal{X}^2) = \\
&= \int_0^{+\infty} \frac{1}{2^{\frac{n}{2}}\Gamma\Big(\dfrac{n}{2}\Big)} e^{-\frac{1}{2}u} u^{\frac{n}{2}-1} (u-n)^2 \, du = \frac{1}{2^{\frac{n}{2}}\Gamma\Big(\dfrac{n}{2}\Big)} \int_0^{+\infty} e^{-\xi} 2^{\frac{n}{2}-1} \xi^{\frac{n}{2}-1} (2\xi - n)^2 2 \, d\xi = \\
&= \frac{1}{\Gamma\Big(\dfrac{n}{2}\Big)} \int_0^{+\infty} e^{-\xi} \Big(4\xi^{\frac{n}{2}+1} - 4n\xi^{\frac{n}{2}} + n^2 \xi^{\frac{n}{2}-1}\Big) \, d\xi = \\
&= \frac{1}{\Gamma\Big(\dfrac{n}{2}\Big)} \Big[4\Gamma\Big(\frac{n}{2}+2\Big) - 4n\Gamma\Big(\frac{n}{2}+1\Big) + n^2\Gamma\Big(\frac{n}{2}\Big)\Big] = \\
&= \frac{1}{\Gamma\Big(\dfrac{n}{2}\Big)} \Gamma\Big(\frac{n}{2}\Big) \Big[4\Big(\frac{n}{2}+1\Big)\frac{n}{2} - 4n\frac{n}{2} + n^2\Big] = 2n
\end{aligned}$$

Ciò completa la prova del risultato. $\square$

10. Teorema di caratterizzazione delle forme quadratiche di variabili normali standard indipendenti che presentano una distribuzione di chi quadrato

Sia $x^T Ax$ una forma quadratica semidefinita positiva delle variabili normali standard indipendenti $(x_1, x_2, \ldots, x_n) = x$. Allora la variabile casuale $Q = x^T Ax$ risulta una variabile di $\mathcal{X}^2$ se e soltanto se la matrice A è idempotente:

$$A^2 = A \, .$$

In tal caso, indicato con p il rango della matrice A, il numero di gradi di libertà di $x^T Ax$ è pari a p. Si ha inoltre che p coincide con la traccia $\mathrm{tr}(A)$ della matrice A.

Dimostrazione

(i) **Se A è idempotente, di rango $p < n$, allora $Q = x^T Ax$ risulta una variabile di chi quadrato a p gradi di libertà**

Dato un generico autovalore λ di A, se $v \in \mathbb{R}^n \setminus \{0\}$ è un autovettore del relativo autospazio vale la relazione

$$Av = \lambda v \, , \quad v \in \mathbb{R}^n \setminus \{0\}$$

e quindi, in forza della condizione di idempotenza,

$$\lambda v = Av = A^2 v = \lambda^2 v$$

da cui si deduce l'equazione

$$\lambda(1 - \lambda)v = 0$$

che per $v \neq 0$ implica $\lambda(1 - \lambda) = 0$. **I soli autovalori possibili di A sono dunque $\lambda = 0$ e $\lambda = 1$.** Si osservi, in particolare, che la matrice simmetrica A è necessariamente semidefinita (o definita) positiva.
Il rango di una qualsiasi matrice reale e simmetrica coincide con il numero dei suoi autovalori non nulli, nella fattispecie con la molteplicità algebrica dell'autovalore 1.

D'altra parte, la traccia di A coincide con la somma degli autovalori, e quindi in questo caso con la molteplicità di 1. Se ne deduce l'identità di rango e traccia:

$$\mathrm{range}(A) = \mathrm{tr}(A) = p = \text{molteplicità algebrica dell'autovalore 1}$$

Dall'ipotesi che la matrice A sia reale e simmetrica segue altresì che esiste una matrice ortogonale C che la diagonalizza, verificando la relazione

$$A = C^T DC \qquad \Longleftrightarrow \qquad CAC^T = D$$

in cui $C^T = C^{-1}$ e D è la matrice diagonale avente come elementi diagonali gli autovalori di A in un qualche ordine, per esempio:

$$D = \begin{pmatrix} 1 & 0 & 0 & 0 & \ldots & 0 \\ 0 & \ddots & 0 & 0 & \ldots & 0 \\ 0 & 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 0 \end{pmatrix}$$

Con il cambiamento di variabili

$$z = Cx \quad , \qquad z = (z_1, \ldots, z_n) \in \mathbb{R}^n$$

la variabile casuale $Q = x^T Ax$ si riduce ad una somma di quadrati delle z:

$$Q = x^T Ax = (C^{-1}z)^T AC^{-1}z = (C^T z)^T AC^T z = z^T CAC^T z = z^T Dz = \sum_{i=1}^{p} z_i^2$$

Ma in virtù del teorema 5 **le nuove variabili z costituiscono un set di variabili casuali gaussiane**, la cui matrice di struttura si calcola immediatamente ricordando che le x sono gaussiane indipendenti e standard (la matrice di struttura di x è l'identità $\mathbb{I}$):

$$[C^{-1}]^T \mathbb{I} C^{-1} = [C^T]^T \mathbb{I} C^T = CC^T = \mathbb{I}$$

per cui **anche le z definiscono un sistema di variabili gaussiane indipendenti e standard**. Se ne conclude che la variabile $Q = x^T Ax$ può intendersi come una somma di quadrati di p variabili normali indipendenti e standard, e che pertanto essa segue una distribuzione di $\mathcal{X}^2$ a p gradi di libertà, essendo $p = \text{range}(A) = \text{tr}(A)$.

(ii) **Se $Q = x^T Ax$ è una variabile di chi quadrato a p gradi di libertà, allora A è idempotente, con rango e traccia uguali a p**
Come già osservato al punto precedente, la simmetria di A consente di introdurre un conveniente cambiamento di variabili, mediante una trasformazione lineare ortogonale $z = Cx$, che:

- consente di ridurre la matrice A alla forma diagonale, con autovalori reali e non negativi $\lambda_1, \ldots, \lambda_n$;
- sostituisce alle variabili gaussiane indipendenti e standard x, delle nuove variabili gaussiane z a loro volta indipendenti e standard;
- permette di identificare la variabile casuale $x^T Ax$ con una somma pesata di quadrati delle z

$$Q = x^T Ax = \sum_{i=1}^{n} \lambda_i z_i^2 \tag{10.1}$$

Nel seguito si assumerà di avere introdotto tale cambiamento di variabili e di operare quindi con la variabile di $\mathcal{X}^2$ già scritta nella forma (10.1).

Per $\lambda > 0$ si calcola la **funzione generatrice dei momenti di** $Q = \sum_{i=1}^n \lambda_i z_i^2$

$$G(\lambda) = \int_{\mathbb{R}^n} e^{-\lambda \sum_{i=1}^n \lambda_i z_i^2} \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\sum_{i=1}^n z_i^2} dz_1 \ldots dz_n =$$

$$= \prod_{i=1}^{n} \frac{1}{(2\pi)^{1/2}} \int_{\mathbb{R}} e^{-\frac{1}{2}(1+2\lambda_i\lambda)z_i^2} dz_i = \prod_{i=1}^{n} \frac{1}{\sqrt{1+2\lambda_i\lambda}}$$

che per una distribuzione di $\mathcal{X}^2$ a p gradi di libertà deve peraltro coincidere con la funzione

$$\begin{aligned}G(\lambda) &= \int_{\mathbb{R}^+} e^{-\lambda Q} \frac{1}{2^{p/2}\Gamma(p/2)} Q^{\frac{p}{2}-1} e^{-Q/2} dQ = \\ &= \int_{\mathbb{R}^+} \frac{1}{2^{p/2}\Gamma(p/2)} e^{-\frac{1}{2}(1+2\lambda)Q} Q^{\frac{p}{2}-1} dQ = \\ &= \frac{1}{(1+2\lambda)^{p/2}} \int_{\mathbb{R}^2} \frac{1}{2^{p/2}\Gamma(p/2)} e^{-\xi/2} \xi^{\frac{p}{2}-1}\, d\xi = \frac{1}{(1+2\lambda)^{p/2}}\end{aligned}$$

essendosi fatto uso del cambiamento di variabile $\xi = (1+2\lambda)Q$. Deve quindi aversi:

$$\prod_{j=1}^{n} \frac{1}{\sqrt{1+2\lambda_j\lambda}} = \frac{1}{(1+2\lambda)^{p/2}} \qquad \forall\, \lambda > 0$$

ossia

$$\prod_{j=1}^{n} (1+2\lambda_j\lambda) = (1+2\lambda)^p \qquad \forall\, \lambda > 0$$

che equivale a

$$\prod_{j=1}^{n} \Big(\frac{1}{\lambda} + 2\lambda_j\Big) = \frac{1}{\lambda^{n-p}} \Big(\frac{1}{\lambda} + 2\Big)^p \qquad \forall\, \lambda > 0$$

ed infine all'identità

$$\prod_{j=1}^{n} (z + 2\lambda_j) = z^{n-p}(z+2)^p \qquad \forall\, z > 0$$

Per l'identità dei polinomi monici a primo e secondo membro, il teorema fondamentale dell'algebra implica che degli n coefficienti non negativi $\lambda_1, \ldots, \lambda_n$ ce ne debbano essere p uguali ad 1 e $n-p$ uguali a 0. Ne segue allora che la matrice A ha rango p e che, potendosi scrivere nella forma

$$A = C^T \begin{pmatrix} \lambda_1 & & \mathbb{O} \\ & \ddots & \\ \mathbb{O} & & \lambda_n \end{pmatrix} C\,,$$

è idempotente

$$\begin{aligned}A^2 &= C^T \begin{pmatrix} \lambda_1 & & \mathbb{O} \\ & \ddots & \\ \mathbb{O} & & \lambda_n \end{pmatrix} CC^T \begin{pmatrix} \lambda_1 & & \mathbb{O} \\ & \ddots & \\ \mathbb{O} & & \lambda_n \end{pmatrix} C = \\ &= C^T \begin{pmatrix} \lambda_1 & & \mathbb{O} \\ & \ddots & \\ \mathbb{O} & & \lambda_n \end{pmatrix} \begin{pmatrix} \lambda_1 & & \mathbb{O} \\ & \ddots & \\ \mathbb{O} & & \lambda_n \end{pmatrix} C = C^T \begin{pmatrix} \lambda_1^2 & & \mathbb{O} \\ & \ddots & \\ \mathbb{O} & & \lambda_n^2 \end{pmatrix} C = \\ &= C^T \begin{pmatrix} \lambda_1 & & \mathbb{O} \\ & \ddots & \\ \mathbb{O} & & \lambda_n \end{pmatrix} C = A \qquad \square\end{aligned}$$

11. La somma dei quadrati degli scarti rispetto alla media $\overline{x}$ di n variabili gaussiane indipendenti e di varianza unitaria costituisce una variabile di $\mathcal{X}^2$ a $n-1$ gradi di libertà

Si considerino le variabili gaussiane $x_1, x_2, \ldots, x_n$ stocasticamente indipendenti, con media nulla e varianza unitaria. Indicato con $\overline{x}$ il valore medio:

$$\overline{x} \;=\; \frac{1}{n}\sum_{i=1}^{n} x_i \;,$$

la somma dei quadrati degli scarti delle variabili rispetto a tale valore medio:

$$(n-1)s^2 \;=\; \sum_{i=1}^{n}(x_i-\overline{x})^2$$

segue una distribuzione di $\mathcal{X}^2$ ad $n-1$ gradi di libertà.

Dimostrazione
Si verifica in primo luogo che la variabile casuale $(n-1)s^2$ è una forma quadratica (ovviamente semidefinita positiva) delle variabili gaussiane, indipendenti e standard x_i. Vale infatti:

$$\begin{aligned}\sum_{i=1}^{n}(x_i-\overline{x})^2 \;&=\; \sum_{i=1}^{n} x_i^2 \;-\; n\overline{x}^2 \;=\; \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\Big(\sum_{i=1}^{n} x_i\Big)^2 \;= \\ &=\; \sum_{i,j=1}^{n}\Big(x_i x_j \delta_{ij} \;-\; \frac{1}{n}x_i x_j\Big) \;=\; \sum_{i,j=1}^{n} x_i\Big(\delta_{ij}-\frac{1}{n}\Big)x_j \;=\; \sum_{i,j=1}^{n} x_i A_{ij} x_j\end{aligned}$$

in cui si è posto $A_{ij} = \delta_{ij} - \dfrac{1}{n}$. La matrice rappresentativa della forma quadratica

$$A \;=\; \mathbb{I} - \frac{1}{n}Q^{(n)}$$

con $Q^{(n)}_{ij} = 1$, $1 \le i,j \le n$, è reale, simmetrica, semidefinita definita positiva.

In forza della diagonalizzabilità di A rispetto ad una base ortonormale di autovettori, l'identificazione di $(n-1)s^2 \;=\; x^T A x$ come variabile di $\mathcal{X}^2$ si riduce semplicemente a provare che gli autovalori della matrice sono soltanto 0 e 1, determinandone le rispettive molteplicità.

Autovalori di A
Dall'identità

$$|A-\lambda\mathbb{I}| = \Big|\mathbb{I}-\frac{1}{n}Q^{(n)}-\lambda\mathbb{I}\Big| = \Big|-\mathbb{I}+\frac{1}{n}Q^{(n)}+\lambda\mathbb{I}\Big| = \Big|\frac{1}{n}Q^{(n)}-(1-\lambda)\mathbb{I}\Big| = \frac{1}{n}\Big|Q^{(n)}-n(1-\lambda)\mathbb{I}\Big|$$

si deduce l'equivalenza

$$\lambda \text{ è autovalore di } A \qquad \Longleftrightarrow \qquad \mu = n(1-\lambda) \text{ è autovalore di } Q$$

ovvero

$$\mu \text{ è autovalore di } Q \qquad \Longleftrightarrow \qquad \lambda = 1 - \frac{\mu}{n} \text{ è autovalore di } A$$

Gli autovalori di $Q^{(n)}$ possono essere determinati osservando che per il relativo polinomio caratteristico, il determinante

$$\left|Q^{(n)} - \mu\mathbb{I}\right| = \begin{vmatrix} 1-\mu & 1 & \cdots & 1 \\ 1 & 1-\mu & \cdots & \vdots \\ \vdots & \vdots & \ddots & 1 \\ 1 & \cdots & 1 & 1-\mu \end{vmatrix}$$

vale la relazione differenziale di ricorrenza

$$\frac{d}{d\mu}\left|Q^{(n)} - \mu\mathbb{I}\right| = \begin{vmatrix} -1 & 1 & \cdots & 1 \\ 0 & 1-\mu & \cdots & \vdots \\ \vdots & \vdots & \ddots & 1 \\ 0 & \cdots & 1 & 1-\mu \end{vmatrix} + \begin{vmatrix} 1-\mu & 0 & \cdots & 1 \\ 1 & -1 & \cdots & \vdots \\ \vdots & 0 & \ddots & 1 \\ 1 & 0 & 1 & 1-\mu \end{vmatrix} + \cdots$$

$$\cdots + \begin{vmatrix} 1-\mu & 1 & \cdots & 0 \\ 1 & 1-\mu & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 1 & \cdots & 1 & -1 \end{vmatrix} = -n\left|Q^{(n-1)} - \mu\mathbb{I}\right|$$

Posto allora $\Delta_n(\mu) = \left|Q^{(n)} - \mu\mathbb{I}\right|$, polinomio caratteristico di $Q^{(n)}$, si ha:

$$\frac{d}{d\mu}\Delta_n(\mu) = -n\,\Delta_{(n-1)}(\mu) \qquad \forall\,\mu \in \mathbb{R}$$

e quindi, integrando in μ con punto iniziale $\mu = 0$,

$$\Delta_n(\mu) = \Delta_n(0) - n\int_0^\mu \Delta_{n-1}(\xi)\,d\xi$$

dove $\Delta_n(0) = |Q^{(n)}| = 0$ — $Q^{(n)}$ è singolare, avendo tutte le righe linearmente dipendenti. Perciò

$$\Delta_n(\mu) = -n\int_0^\mu \Delta_{n-1}(\xi)\,d\xi$$

ed è allora facile dimostrare per induzione che

$$\Delta_n(\mu) = (-1)^n\left[\mu^n - n\mu^{n-1}\right]. \tag{11.1}$$

Infatti, posto la (11.1) soddisfatta per un n qualsivoglia, si ottiene

$$
\begin{aligned}
\Delta_{n+1}(\mu) &= -(n+1)\int_0^\mu (-1)^n\big(\xi^n - n\xi^{n-1}\big)\,d\xi = \\
&= (-1)^{n+1}(n+1)\Big[\frac{\mu^{n+1}}{n+1} - \mu^n\Big] = (-1)^{n+1}\big[\mu^{n+1} - (n+1)\mu^n\big]
\end{aligned}
$$

il che prova l'asserto.
Gli autovalori di $Q^{(n)}$ sono pertanto tutte e sole le soluzioni dell'equazione caratteristica

$$|Q^{(n)} - \mu\mathbb{I}| = 0 \qquad \Longleftrightarrow \qquad \mu^{n-1}(\mu - n) = 0$$

ossia:

$$
\begin{aligned}
&\mu = 0 \ \text{ con molteplicità algebrica } \ n-1 \\
&\mu = n \ \text{ con molteplicità algebrica } \ 1
\end{aligned}
$$

Di qui si ricavano gli autovalori di A

$$
\begin{aligned}
&\lambda = 1 \ \text{ con molteplicità algebrica } \ n-1 \\
&\lambda = 0 \ \text{ con molteplicità algebrica } \ 1
\end{aligned}
$$

In conclusione, la forma quadratica $(n-1)s^2 = x^T Ax$ può essere ricondotta ad una somma di $n-1$ quadrati di variabili gaussiane indipendenti e standard, in modo che $(n-1)s^2$ segue una distribuzione di $\mathcal{X}^2$ a $n-1$ gradi di libertà. $\square$

Allo stesso risultato si può pervenire per mezzo del precedente teorema 10, verificando che la matrice rappresentativa A è idempotente e con traccia $n-1$. Si ha infatti, per ogni i,j, $1 \leq i,j \leq n$:

$$
\begin{aligned}
A^2_{ij} &= \sum_{k=1}^n A_{ik}A_{kj} = \sum_{k=1}^n \Big(\delta_{ik} - \frac{1}{n}\Big)\Big(\delta_{kj} - \frac{1}{n}\Big) = \\
&= \sum_{k=1}^n \Big[\delta_{ik}\delta_{kj} - \delta_{ik}\frac{1}{n} - \delta_{kj}\frac{1}{n} + \frac{1}{n^2}\Big] = \delta_{ij} - \frac{1}{n} - \frac{1}{n} + n\,\frac{1}{n^2} = \delta_{ij} - \frac{1}{n} = A_{ij}
\end{aligned}
$$

e quindi $A^2 = A$, in modo che la matrice rappresentativa A risulta idempotente e la forma quadratica $x^T Ax$ è una variabile di $\mathcal{X}^2$. Il numero di gradi di libertà di questa è dato dalla traccia di A, ossia

$$\mathrm{tr}(A) = \sum_{i=1}^n A_{ii} = \sum_{i=1}^n \Big(1 - \frac{1}{n}\Big) = n\Big(1 - \frac{1}{n}\Big) = n-1$$

come già dimostrato. $\square$

Osservazione
È utile sottolineare come il risultato precedente rimanga valido nell'ipotesi che le variabili gaussiane indipendenti $x_1, x_2, \ldots, x_n$ abbiano varianza unitaria ed eguale media $m \neq 0$. La dimostrazione si ripete inalterata, salvo per il fatto di dover considerare le variabili casuali a media nulla $x_i - m$ in luogo delle x_i.

Analogamente, partendo da un sistema di variabili gaussiane indipendenti $x_1, x_2, \ldots, x_n$ con eguale media m e varianza $\sigma^2 > 0$ è dato definire la variabile di $\mathcal{X}^2$

$$\frac{(n-1)s^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

a $n-1$ gradi di libertà. Immediata è la dimostrazione — si tratta di considerare le variabili normali $(x_i - m)/\sigma$ in sostituzione delle x_i.

12. Teorema di Craig

Sia $x^T = (x_1 \ldots x_n)$ un vettore di n variabili gaussiane indipendenti e standard. Date due matrici reali simmetriche A e B si considerino le forme quadratiche

$$Q_1 = x^T A x \qquad Q_2 = x^T B x\ .$$

Allora le variabili casuali Q_1 e Q_2 sono stocasticamente indipendenti se e solo se $AB = 0$.

Dimostrazione
(i) **La condizione è sufficiente**
Se $AB = 0$, allora $0 = (AB)^T = B^T A^T = BA$. Perciò $AB = BA = 0$; le matrici reali e simmetriche A e B commutano.

Di conseguenza, esiste una base di $\mathbb{R}^n$ di autovettori ortonormali **comuni** di A e B e si può quindi determinare una matrice reale ortogonale C tale che

$$C^T A C = D_1 = \begin{pmatrix} \lambda_1 & & \mathbb{O} \\ & \ddots & \\ \mathbb{O} & & \lambda_n \end{pmatrix} \qquad \text{e} \qquad C^T B C = D_2 = \begin{pmatrix} \mu_1 & & \mathbb{O} \\ & \ddots & \\ \mathbb{O} & & \mu_n \end{pmatrix}.$$

Introdotto allora il cambiamento di coordinate $x = Cz$, le variabili casuali Q_1 e Q_2 si riducono a somme pesate di quadrati delle variabili z_i:

$$Q_1 = z^T C^T A C z = z^T D_1 z = \sum_{i=1}^{n} \lambda_i z_i^2$$
$$Q_2 = z^T C^T B C z = z^T D_2 z = \sum_{i=1}^{n} \mu_i z_i^2$$

con $z = C^T x = (z_1 \ldots z_n)^T$ sistema di n variabili gaussiane indipendenti e standard al pari delle $(x_1 \ldots x_n)$, in quanto C^T è ortogonale.

Inoltre

$$AB = CD_1C^TCD_2C^T = CD_1D_2C^T = C \begin{pmatrix} \lambda_1\mu_1 & & \mathbb{O} \\ & \ddots & \\ \mathbb{O} & & \lambda_n\mu_n \end{pmatrix} C^T$$

ed ovviamente $AB = 0$ se e soltanto se $D_1D_2 = 0$, per cui gli autovalori delle matrici A e B devono soddisfare le condizioni

$$\lambda_i\mu_i = 0 \qquad \forall\, i = 1, \ldots, n\, .$$

Assunti allora

$$\lambda_{i_1}, \ldots, \lambda_{i_\ell} \neq 0 \qquad \text{e} \qquad \lambda_i = 0 \quad \forall\, i \in \{1, 2, \ldots, n\} \setminus \{i_1, \ldots, i_\ell\}\, ;$$

$$\mu_{j_1}, \ldots, \mu_{j_m} \neq 0 \qquad \text{e} \qquad \mu_j = 0 \quad \forall\, j \in \{1, 2, \ldots, n\} \setminus \{j_1, \ldots, j_m\}\, ;$$

deve valere la condizione

$$\{i_1, \ldots, i_\ell\} \cap \{j_1, \ldots, j_m\} = \emptyset\, .$$

Perciò comunque si fissino $\overline{Q}_1, \overline{Q}_2 \in \mathbb{R}$, la probabilità congiunta $p[Q_1 \leq \overline{Q}_1, Q_2 \leq \overline{Q}_2]$ che si abbia contemporaneamente $Q_1 \leq \overline{Q}_1$ e $Q_2 \leq \overline{Q}_2$ deve essere data dall'espressione:

$$\int\limits_{\left\{\sum_{a=1}^{\ell} \lambda_{i_a} z_{i_a}^2 \leq \overline{Q}_1\right\} \cap \left\{\sum_{b=1}^{m} \mu_{j_b} z_{j_b}^2 \leq \overline{Q}_2\right\}} \frac{1}{(2\pi)^{n/2}}\, e^{-\frac{1}{2}\sum_{i=1}^{n} z_i^2}\, dz_1 \ldots dz_n =$$

$$= \frac{1}{(2\pi)^{(m+\ell)/2}} \int\limits_{\left\{\sum_{a=1}^{\ell} \lambda_{i_a} z_{i_a}^2 \leq \overline{Q}_1\right\}} e^{-\frac{1}{2}\sum_{a=1}^{\ell} z_{i_a}^2} \prod_{a=1}^{\ell} dz_{i_a} \cdot$$

$$\cdot \int\limits_{\left\{\sum_{b=1}^{m} \mu_{j_b} z_{j_b}^2 \leq \overline{Q}_2\right\}} e^{-\frac{1}{2}\sum_{b=1}^{m} z_{j_b}^2} \prod_{b=1}^{m} dz_{j_b} = p[Q_1 \leq \overline{Q}_1] \cdot p[Q_2 \leq \overline{Q}_2]$$

ed è quindi riconducibile al semplice prodotto delle distribuzioni cumulative di probabilità delle singole variabili casuali Q_1 e Q_2. Si tratta pertanto di variabili stocasticamente indipendenti.

(ii) **La condizione è necessaria**
Se le variabili $Q_1 = x^TAx$ e $Q_2 = x^TBx$ sono stocasticamente indipendenti, la loro distribuzione di probabilità congiunta si scrive come prodotto di due distribuzioni di probabilità, la prima relativa alla sola x^TAx e la seconda relativa a x^TBx soltanto. La funzione generatrice dei momenti della distribuzione di probabilità congiunta $p(Q_1, Q_2)$ è allora

data dal prodotto delle funzioni generatrici delle distribuzioni di probabilità per le singole variabili, $p_1(Q_1)$ e $p_2(Q_2)$:

$$G(\lambda,\mu) \,=\, \int_{\mathbb{R}^2} p(Q_1,Q_2)\, e^{-\lambda Q_1 - \mu Q_2}\, dQ_1 dQ_2 \,=\, \int_{\mathbb{R}} p_1(Q_1) e^{-\lambda Q_1} dQ_1 \int_{\mathbb{R}} p_2(Q_2) e^{-\mu Q_2} dQ_2\,.$$

In effetti, dalla definizione di Q_1 e Q_2 si ottiene:

$$G(\lambda,\mu) \,=\, \int_{\mathbb{R}^n} \Big(\frac{1}{2\pi}\Big)^n e^{-\frac{1}{2}\sum_{i=1}^n x_i^2 - \lambda x^T A x - \mu x^T B x} dx_1 \ldots dx_n \,=$$

$$= \int_{\mathbb{R}^n} \Big(\frac{1}{2\pi}\Big)^n e^{-\frac{1}{2} x^T (\mathbb{I} + 2\lambda A + 2\mu B) x} dx_1 \ldots dx_n \,=\, \big[\det(\mathbb{I} + 2\lambda A + 2\mu B)\big]^{-1/2}$$

$\forall\, \lambda,\mu \in \mathbb{R}$ abbastanza piccoli, mentre in modo analogo:

$$\int_{\mathbb{R}} p_1(Q_1) e^{-\lambda Q_1} dQ_1 \,=\, \int_{\mathbb{R}^n} \Big(\frac{1}{2\pi}\Big)^n e^{-\frac{1}{2} x^T (\mathbb{I} + 2\lambda A) x} dx_1 \ldots dx_n \,=\, \big[\det(\mathbb{I} + 2\lambda A)\big]^{-1/2}$$

$$\int_{\mathbb{R}} p_2(Q_2) e^{-\mu Q_2} dQ_2 \,=\, \int_{\mathbb{R}^n} \Big(\frac{1}{2\pi}\Big)^n e^{-\frac{1}{2} x^T (\mathbb{I} + 2\mu B) x} dx_1 \ldots dx_n \,=\, \big[\det(\mathbb{I} + 2\mu B)\big]^{-1/2}$$

per cui la condizione di indipendenza stocastica di Q_1 e Q_2 implica che:

$$\big[\det(\mathbb{I} + 2\lambda A + 2\mu B)\big]^{-1/2} \,=\, \big[\det(\mathbb{I} + 2\lambda A)\big]^{-1/2} \big[\det(\mathbb{I} + 2\mu B)\big]^{-1/2} \qquad \forall\, \lambda,\mu \sim 0\,.$$

Quest'ultima condizione equivale a:

$$\det(\mathbb{I} + \lambda A) \det(\mathbb{I} + \mu B) \,=\, \det(\mathbb{I} + \lambda A + \mu B) \qquad \forall\, \lambda,\mu \sim 0$$

ossia:

$$\det(\mathbb{I} + \mu B) \,=\, \frac{\det(\mathbb{I} + \lambda A + \mu B)}{\det(\mathbb{I} + \lambda A)} \,=\, \det[(\mathbb{I} + \lambda A)^{-1}] \det(\mathbb{I} + \lambda A + \mu B) \,=$$

$$= \det[\mathbb{I} + \mu(\mathbb{I} + \lambda A)^{-1} B]$$

e quindi posto $y = -1/\mu$ si ha che $\forall\, y \in \mathbb{R}$ di valore assoluto abbastanza grande deve valere

$$\det(-y\mathbb{I} + B) \,=\, \det[-y\mathbb{I} + (\mathbb{I} + \lambda A)^{-1} B]\,.$$

Nella relazione precedente il primo ed il secondo membro sono polinomi di grado n nella variabile y, per cui l'uguaglianza deve in realtà estendersi ad ogni $y \in \mathbb{R}$, in virtù del teorema di identità dei polinomi (due polinomi che coincidono su un intervallo reale coincidono su tutto $\mathbb{R}$). Detti polinomi sono i polinomi caratteristici delle matrici B e $(\mathbb{I} + \lambda A)^{-1} B$, le quali devono presentare dunque lo stesso polinomio caratteristico $\forall\, \lambda \sim 0$. Da ciò segue che $\forall\, \lambda \sim 0$ le matrici B e $(\mathbb{I} + \lambda A)^{-1} B$ hanno gli stessi autovalori. L'arbitrarietà di λ consente di concludere che $AB = 0$. $\square$

13. Teorema di Fisher-Cochran

Sia $(x_1, x_2, \ldots, x_n) = x$ *un insieme di variabili normali indipendenti e standard. Siano* $Q_1, Q_2, \ldots, Q_k$ *forme quadratiche semidefinite positive delle stesse variabili* $(x_1, x_2, \ldots, x_n)$

$$Q_r \;=\; x^T A_r x \;=\; \sum_{i,h=1}^{n} (A_r)_{ih} x_i x_h \qquad r = 1, 2, \ldots, k$$

con matrici rappresentative $A_1, A_2, \ldots, A_k$*, di ranghi rispettivi* $n_1, n_2, \ldots, n_k$*. Valga inoltre*

$$\sum_{i=1}^{n} x_i^2 \;=\; \sum_{j=1}^{k} Q_j \;.$$

Allora le $Q_1, Q_2, \ldots, Q_k$ *sono variabili di* $\mathcal{X}^2$ *stocasticamente indipendenti se e soltanto se*

$$\sum_{j=1}^{k} n_j \;=\; n$$

e in tal caso Q_r *ha* n_r *gradi di libertà,* $\forall\, r = 1, \ldots k$*.*

Dimostrazione

(i) **La condizione è necessaria**

Siano $Q_i = x^T A_i x$, $i = 1, \ldots, k$ variabili di $\mathcal{X}^2$ indipendenti fra loro. Allora il teorema di caratterizzazione delle forme quadratiche $\mathcal{X}^2$ e quello di Craig implicano che le matrici rappresentative delle Q_i — reali e simmetriche — soddisfino la condizione

$$A_i A_j \;=\; \delta_{ij} A_j$$

mentre l'ipotesi che sia $\sum_{i=1}^{n} x_i^2 \;=\; \sum_{j=1}^{k} Q_j$ impone che si abbia altresì

$$\sum_{i=1}^{k} A_i \;=\; \mathbb{I} \;.$$

Gli insiemi

$$V_i \;=\; \mathrm{range}(A_i) \;=\; \big\{A_i x \;; \quad x \in \mathbb{R}^n\big\} \qquad i = 1, \ldots, k$$

sono sottospazi vettoriali ad intersezione banale

$$V_i \cap V_j \;=\; \{0\} \qquad \forall\, i, j = 1, \ldots, k \,,\; i \neq j \,.$$

Se infatti $z \in V_i \cap V_j$ per $i \neq j$, devono esistere $x, y \in \mathbb{R}^n$ opportuni tali che

$$z \;=\; A_i x \;=\; A_j y$$

e quindi

$$z = A_i x = A_i^2 x = A_i A_i x = A_i A_j y = 0y = 0\,.$$

Per ogni $x \in \mathbb{R}^n$ vale allora

$$x = \mathbb{I}x = \sum_{i=1}^{k} A_i x \qquad \text{con} \qquad A_i x \in V_i \quad \forall\, i = 1, \ldots, k\,.$$

La decomposizione così ottenuta è unica, in quanto

$$x = \sum_{i=1}^{k} v_i \;, \quad v_i \in V_i$$

implica

$$A_j x = A_j \sum_{i=1}^{k} v_i = \sum_{i=1}^{k} A_j v_i = \sum_{i=1}^{k} \delta_{ji} v_i = v_j \qquad \forall\, j = 1, \ldots, k\,.$$

Se ne conclude che $\mathbb{R}^n$ si decompone in somma diretta dei V_i

$$\mathbb{R}^n = \bigoplus_{i=1}^{k} V_i$$

e che di conseguenza la dimensione dello spazio è la somma delle dimensioni dei sottospazi componenti

$$n = \sum_{i=1}^{k} \dim(V_i) = \sum_{i=1}^{k} \operatorname{rank}(A_i) = \sum_{i=1}^{k} n_i\,.$$

(ii) **La condizione è sufficiente**
Se viceversa $\operatorname{rank}(A_i) = n_i\ \forall\, i = 1, \ldots, n$ e $\sum_{i=1}^{k} n_i = n$ allora lo spazio $\mathbb{R}^n$ è la somma diretta dei sottospazi $V_i = \operatorname{range}(A_i)$

$$\mathbb{R}^n = \bigoplus_{i=1}^{k} V_i\,. \tag{13.1}$$

Per ogni $x \in \mathbb{R}^n$ vale infatti

$$x = \mathbb{I}x = \left(\sum_{i=1}^{k} A_i\right) x = \sum_{i=1}^{k} A_i x$$

con

$$A_i x \in \{A_i y\,;\ y \in \mathbb{R}^n\} = \operatorname{range}(A_i) = V_i$$

e

$$\dim V_i = \dim \operatorname{range}(A_i) = \operatorname{rank}(A_i) = n_i \qquad \forall\, i = 1, \ldots, n\,.$$

Qualsiasi sistema $\{e_1, e_2, \ldots, e_n\}$ ottenuto collezionando i vettori di base dei sottospazi $V_1, V_2, \ldots, V_k$ si compone di n vettori

$$\sum_{i=1}^{k} \dim V_i \;=\; \sum_{i=1}^{k} \dim \operatorname{range}(A_i) \;=\; \sum_{i=1}^{k} \operatorname{rank}(A_i) \;=\; \sum_{i=1}^{k} n_i \;=\; n$$

e **ogni** vettore $x \in \mathbb{R}^n$ si scrive come combinazione lineare dei vettori e_i, $i = 1, \ldots n$: ne segue che $\{e_1, \ldots, e_n\}$ deve essere un sistema linearmente indipendente e costituire dunque una base di $\mathbb{R}^n$. In caso contrario ogni vettore di $\mathbb{R}^n$ si potrebbe esprimere come combinazione lineare di un sistema linearmente indipendente estratto da $\{e_1, \ldots, e_n\}$ e quindi di un numero di vettori minore di $n = \dim \mathbb{R}^n$, in contrasto con il lemma di Steiniz. Di qui la (13.1).

Per ogni $x \in \mathbb{R}^n$ e $j = 1, \ldots, n$ deve aversi pertanto

$$A_j x \in V_j \qquad \text{e} \qquad A_j x \;=\; \mathbb{I} A_j x \;=\; \sum_{i=1}^{k} A_i A_j x \;=\; A_j^2 x + \sum_{i \neq j} A_i A_j x$$

con $A_i A_j x \in V_i \; \forall\, i = 1, \ldots, k$. Poiché la decomposizione di $A_j x$ deve essere unica si deduce che

$$A_i A_j x \;=\; 0 \qquad \forall\, i \neq j \qquad \text{e} \qquad A_j^2 x \;=\; A_j x \qquad \forall\, j$$

e per l'arbitrarietà di x si ha infine

$$A_i A_j \;=\; \delta_{ij} A_j \qquad \forall\, i, j = 1, \ldots, n\,.$$

Per il teorema di Craig la condizione

$$A_i A_j \;=\; 0 \qquad \forall\, i \neq j$$

sulle matrici simmetriche A_i(^{o}) implica che le variabili casuali

$$Q_i \;=\; x^T A_i x \qquad Q_j \;=\; x^T A_j x \qquad \forall\, i \neq j$$

siano tutte stocasticamente indipendenti, mentre la proprietà di idempotenza

$$A_j^2 \;=\; A_j \qquad \forall\, j$$

assicura che le

$$Q_j \;=\; x^T A_j x \qquad \forall\, j = 1, \ldots, k$$

siano tutte variabili di $\mathcal{X}^2$. Essendo poi range$(A_i) = n_i$, si conclude immediatamente, diagonalizzando la matrice A_j con un cambiamento di base ortonormale, che $Q_j = x^T A_j x$ costituisce una variabile di $\mathcal{X}^2$ a n_j gradi di libertà. $\square$

(^{o}) N.B. La simmetria delle matrici A_i interviene soltanto a questo punto della dimostrazione. Il ragionamento fin qui seguito prova che le matrici in questione individuano necessariamente degli operatori di proiezione, ma non ne impone la simmetria — ciò richiederebbe la mutua ortogonalità dei sottospazi V_i, informazione di cui non si è fatto uso. Se infatti $\mathbb{R}^n = \bigoplus_{i=1}^{k} V_i$, gli operatori di proiezione $P_i : \mathbb{R}^n \to V_i$ sono tutti simmetrici se e soltanto se i sottospazi V_i risultano tutti mutuamente ortogonali.

14. Primo caso notevole a due variabili

Siano

$$Q_1 = x^T A_1 x \qquad Q_2 = x^T A_2 x$$

due forme quadratiche semidefinite positive delle variabili gaussiane indipendenti e standard $(x_1, x_2, \ldots, x_n) = x$*. Le variabili casuali* Q_1 *e* Q_2 *soddisfino la condizione*

$$Q_1 + Q_2 = \sum_{i=1}^{n} x_i^2$$

Allora se Q_1 *è una variabile di* $\mathcal{X}^2$ *a* $p < n$ *gradi di libertà, la* Q_2 *è una variabile di* $\mathcal{X}^2$ *a* $n - p < n$ *gradi di libertà, e le due variabili sono stocasticamente indipendenti*

Dimostrazione
Segue immediatamente dai teoremi di Craig e di caratterizzazione delle variabili $\mathcal{X}^2$, osservando che $A_1 + A_2 = \mathbb{I}$. Si ha infatti che $A_2 = \mathbb{I} - A_1$ e l'essere Q_1 una variabile di $\mathcal{X}^2$ implica che si abbia

$$A_2^2 = (\mathbb{I} - A_1)^2 = \mathbb{I} - 2A_1 + A_1^2 = \mathbb{I} - 2A_1 + A_1 = \mathbb{I} - A_1 = A_2$$

per cui anche Q_2 è una variabile di $\mathcal{X}^2$. L'indipendenza stocastica di Q_1 e Q_2 segue poi dall'ovvia relazione e dal teorema di Craig

$$A_1 A_2 = A_1(\mathbb{I} - A_1) = A_1 - A_1^2 = A_1 - A_1 = 0 .$$

Il teorema di Fisher-Cocran implica poi che

$$\text{rank}(A_1) + \text{rank}(A_2) = n$$

che cioè la somma del numero di gradi di libertà di Q_1 e Q_2 sia n. Di conseguenza, Q_2 è una variabile di $\mathcal{X}^2$ a

$$\text{rank}(A_2) = n - \text{rank}(A_1) = n - p$$

gradi di libertà([1]). $\square$

15. Secondo caso notevole a due variabili

Siano Q*,* Q_1 *e* Q_2 *tre forme quadratiche semidefinite positive delle variabili gaussiane* $(x_1, x_2, \ldots, x_n) = x$*, indipendenti e standard. Le variabili casuali* Q*,* Q_1 *e* Q_2 *soddisfino la condizione*

$$Q_1 + Q_2 = Q$$

([1]) Alla stessa conclusione si perviene osservando che $A_1 A_2 = 0$ implica la commutatività di A_1 e A_2 e dunque l'esistenza di una base di $\mathbb{R}^n$ costituita da autovettori comuni di A_1 e A_2. Di conseguenza $\mathbb{R}^n = \ker(A_1 - \mathbb{I}) \bigoplus \ker(A_2 - \mathbb{I})$ e $n = \dim\ker(A_1 - \mathbb{I}) + \dim\ker(A_2 - \mathbb{I})$, ossia $n = \text{rank}(A_1) + \text{rank}(A_2) = p + \text{rank}(A_2)$.

Allora se Q e Q_1 sono variabili di $\mathcal{X}^2$ rispettivamente a n e $p < n$ gradi di libertà, la Q_2 è una variabile di $\mathcal{X}^2$ a $n - p < n$ gradi di libertà, e risulta stocasticamente indipendente da Q_1

Dimostrazione
Basta osservare che la variabile di $\mathcal{X}^2$ a n gradi di libertà Q deve potersi scrivere nella forma

$$Q = x^T A x$$

con A matrice reale simmetrica, semidefinita positiva e tale da soddisfare le condizioni

$$A^2 = A \qquad \text{rank}(A) = n$$

Tale matrice deve quindi avere 1 come solo autovalore, e di molteplicità algebrica n, sicché

$$A = \mathbb{I} \qquad \text{e} \qquad Q = x^T \mathbb{I} x = x^T x = \sum_{i=1}^{n} x_i^2$$

ed il teorema si riduce al risultato precedente. □

16. Terzo caso notevole a due variabili

Siano Q, Q_1 e Q_2 tre forme quadratiche semidefinite positive delle variabili gaussiane $(x_1, x_2, \ldots, x_n) = x$, indipendenti e standard. Le variabili Q, Q_1 e Q_2 siano variabili di $\mathcal{X}^2$ con s, p e q gradi di libertà rispettivamente, soddisfacenti la condizione

$$Q = Q_1 + Q_2$$

Allora Q_1 e Q_2 sono stocasticamente indipendenti e $p + q = s$.

Dimostrazione
Si indichino con A, B e C le matrici rappresentative, reali simmetriche e semidefinite positive, delle variabili casuali Q, Q_1 e Q_2 rispettivamente. L'essere Q, Q_1 e Q_2 variabili di $\mathcal{X}^2$ implica che si abbia

$$A^2 = A \qquad B^2 = B \qquad C^2 = C$$

per cui

$$B + C = (B + C)^2 = B^2 + C^2 + BC + CB = B + C + BC + CB$$

e quindi le matrici B e C anticommutano

$$BC + CB = 0$$

Poiché B è reale e simmetrica, essa ammette una base di autovettori ortonormali; la condizione di idempotenza B^2 e il carattere semidefinito positivo comportano poi che gli autovalori relativi debbano essere soltanto 0 e 1. Conviene esaminare separatamente i due casi.

(*i*) Sia $Bv = 0$, ovvero $v \in \ker(B)$. Allora, moltiplicando a destra per la matrice C si ottiene

$$CBv = C0 = 0$$

e nel contempo

$$BCv = -CBv = 0$$

Per ogni v autovettore di B relativo all'autovalore 0 vale pertanto

$$CBv = BCv = 0$$

(*ii*) Sia all'opposto $Bv = v$, ossia $v \in \ker(B - \mathbb{I})$. In modo analogo a quanto già visto nel caso precedente, si deduce che

$$-BCv = CBv = Cv$$

sicché

$$B(Cv) = -Cv$$

La relazione ottenuta implica che o $Cv = 0$ oppure Cv costituisce un autovettore di B con autovalore -1, circostanza quest'ultima esclusa dall'essere B semidefinita positiva. Deve perciò ricorrere la prima condizione $Cv = 0$, dalla quale segue

$$BCv = B0 = 0 \qquad \text{e} \qquad CBv = -BCv = 0$$

In definitiva, per ogni autovettore v di B si ha

$$BCv = CBv = 0$$

e poiché la stessa relazione caratterizza anche la base di autovettori (ortonormali) di B, si conclude che

$$BCx = CBx = 0 \qquad \forall\, x \in \mathbb{R}^n$$

cioè che $BC = CB = 0$. Il teorema di Craig assicura l'indipendenza stocastica delle variabili casuali Q_1 e Q_2. La stessa condizione $BC = CB = 0$ comporta che le matrici reali e simmetriche B e C commutano, ammettendo perciò una base di $\mathbb{R}^n$ di autovettori comuni. Indicato con y uno di tali autovettori comuni, dovrà aversi

$$By = \beta y \qquad Cy = \gamma y$$

in modo che gli stessi vettori saranno anche autovettori di A

$$Ay = (B + C)y = \beta y + \gamma y = (\beta + \gamma)y$$

con gli autovalori relativi sempre uguali a zero o uno

$$\beta \in \{0,1\} \qquad \gamma \in \{0,1\} \qquad \beta+\gamma \in \{0,1\}$$

Ogni autovettore di B relativo all'autovalore 1 risulta autovalore di C relativo all'autovalore 0 e viceversa, ogni autovalore di C relativo a 1 costituisce anche un autovettore di B relativo a 0; gli autovettori di B e C associati a 0 sono altresì autovettori di A relativi a 0. Infatti:

$$\begin{array}{lll} y = By & \Longrightarrow & Cy = CBy = 0y = 0 \\ y = Cy & \Longrightarrow & By = BCy = 0y = 0 \\ By = 0 \qquad Cy = 0 & \Longrightarrow & Ay = (B+C)y = By + Cy = 0 \end{array}$$

e queste relazioni applicate agli autovettori di base implicano

$$\begin{aligned} \ker(B-\mathbb{I}) &= \ker(C) \\ \ker(C-\mathbb{I}) &= \ker(B) \\ \ker(A) &= \ker(B) \cap \ker(C) \\ \ker(A-\mathbb{I}) &= \ker(B-\mathbb{I}) \cup \ker(C-\mathbb{I}) \\ \emptyset &= \ker(B-\mathbb{I}) \cap \ker(C-\mathbb{I}) \end{aligned}$$

Conseguentemente

$$\begin{aligned} s = \operatorname{rank}(A) = \dim\ker(A-\mathbb{I}) &= \dim\ker(B-\mathbb{I}) + \dim\ker(C-\mathbb{I}) = \\ &= \operatorname{rank}(B) + \operatorname{rank}(C) = p+q \end{aligned}$$

a completamento della dimostrazione. $\square$

17. Definizione. Variabile t di Student a n gradi di libertà

La variabile t di Student a n gradi di libertà è definita dal quoziente

$$t = \sqrt{n}\frac{\xi}{\sqrt{\mathcal{X}^2}} \in \mathbb{R}$$

essendo ξ una variabile gaussiana standard, $\mathcal{X}^2$ una variabile di chi-quadrato a n gradi di libertà e ξ, $\mathcal{X}^2$ stocasticamente indipendenti fra loro.

18. Distribuzione di probabilità di una variabile di Student a n gradi di libertà

La distribuzione di probabilità di una variabile di Student a n gradi di libertà è data da

$$p_n(t) = \frac{\Gamma\Big(\frac{n+1}{2}\Big)}{\sqrt{\pi n}\,\Gamma\Big(\frac{n}{2}\Big)} \frac{1}{\Big(1+\frac{t^2}{n}\Big)^{\frac{n+1}{2}}} = \frac{1}{\sqrt{n}\;\beta\Big(\frac{n}{2},\frac{1}{2}\Big)} \frac{1}{\Big(1+\frac{t^2}{n}\Big)^{\frac{n+1}{2}}}$$

in cui si indica con $\beta(a,b)$ la funzione beta di Eulero calcolata in $(a,b) \in \mathbb{R}^{+2}$

$$\beta(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx$$

Dimostrazione
La distribuzione da determinare è quella di una variabile casuale della forma

$$t = \sqrt{n}\frac{\xi}{\sqrt{\mathcal{X}^2}}$$

in cui ξ e $\mathcal{X}^2$ sono variabili indipendenti, rispettivamente gaussiana standard e di chi quadrato a n gradi di libertà. La condizione di normalizzazione per la distribuzione di probabilità congiunta si scrive

$$1 = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\xi^2/2}d\xi \quad \frac{1}{2^{n/2}\,\Gamma(n/2)}\int_0^{+\infty} e^{-\mathcal{X}^2/2}(\mathcal{X}^2)^{\frac{n}{2}-1}d(\mathcal{X}^2)$$

e per il teorema di Fubini si riduce all'espressione equivalente

$$1 = \frac{1}{\sqrt{2\pi}\;2^{n/2}\Gamma(n/2)}\int_{\mathbb{R}\times\mathbb{R}^+} d\xi\,d(\mathcal{X}^2)\; e^{-\frac{\xi^2+\mathcal{X}^2}{2}}(\mathcal{X}^2)^{\frac{n}{2}-1}\,.$$

A $\mathcal{X}^2 > 0$ fissato, nell'integrale doppio a secondo membro si introduca il cambiamento di variabile

$$\xi = t\,\frac{\sqrt{\mathcal{X}^2}}{\sqrt{n}}$$

ottenendo così

$$1 = \frac{1}{\sqrt{2\pi}\;2^{n/2}\Gamma(n/2)}\int_{\mathbb{R}\times\mathbb{R}^+} dt\;d(\mathcal{X}^2)\;\frac{\sqrt{\mathcal{X}^2}}{\sqrt{n}}e^{-\frac{\mathcal{X}^2}{2}\left(\frac{t^2}{n}+1\right)}(\mathcal{X}^2)^{\frac{n}{2}-1} =$$

$$= \frac{1}{\sqrt{2\pi}\;2^{n/2}\Gamma(n/2)}\frac{1}{\sqrt{n}}\int_{\mathbb{R}} dt\int_0^{+\infty} d(\mathcal{X}^2)\;\; e^{-\frac{\mathcal{X}^2}{2}\left(1+\frac{t^2}{n}\right)}(\mathcal{X}^2)^{\frac{n-1}{2}}\,.$$

L'ulteriore cambiamento di variabile, a t fissato,

$$z = \mathcal{X}^2\Big(1+\frac{t^2}{n}\Big)$$

nell'integrale più interno a secondo membro, porge infine

$$1 = \frac{1}{\sqrt{2\pi}\;2^{n/2}\Gamma(n/2)\;\sqrt{n}}\int_{\mathbb{R}} dt\int_0^{+\infty} dz\;e^{-z/2}z^{\frac{n-1}{2}}\;\frac{1}{\Big(1+\dfrac{t^2}{n}\Big)^{\frac{n+1}{2}}}$$

ossia, posto $z = 2u$ e scambiati per Fubini i due integrali iterati

$$1 = \frac{1}{\sqrt{2\pi n}\; 2^{n/2}\Gamma(n/2)} 2^{\frac{n+1}{2}} \int_0^{+\infty} du\, e^{-u} u^{\frac{n+1}{2}-1} \int_{\mathbb{R}} dt\; \frac{1}{\Big(1+\dfrac{t^2}{n}\Big)^{\frac{n+1}{2}}}$$

e quindi

$$1 = \frac{1}{\sqrt{\pi n}\;\Gamma(n/2)} \Gamma\Big(\frac{n+1}{2}\Big) \int_{\mathbb{R}} \frac{1}{\Big(1+\dfrac{t^2}{n}\Big)^{\frac{n+1}{2}}}\, dt\,.$$

Di qui si deduce la distribuzione di probabilità desiderata

$$p_n(t) = \frac{1}{\sqrt{\pi n}\;\Gamma(n/2)} \Gamma\Big(\frac{n+1}{2}\Big) \frac{1}{\Big(1+\dfrac{t^2}{n}\Big)^{\frac{n+1}{2}}} \qquad t \in \mathbb{R}\,.$$

La stessa espressione può riscritta in termini della beta euleriana

$$p_n(t) = \frac{1}{\sqrt{n}\;\beta\Big(\dfrac{n}{2},\dfrac{1}{2}\Big)} \frac{1}{\Big(1+\dfrac{t^2}{n}\Big)^{\frac{n+1}{2}}}$$

facendo uso dell'identità

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} = \beta(a,b) \qquad \forall\,(a,b) \in {\mathbb{R}^+}^2 \tag{18.1}$$

e della relazione $\Gamma(1/2) = \sqrt{\pi}$, dalle quali segue che

$$\frac{1}{\sqrt{\pi}\;\Gamma(n/2)} \Gamma\Big(\frac{n+1}{2}\Big) = \frac{1}{\Gamma(1/2)\;\Gamma(n/2)} \Gamma\Big(\frac{n+1}{2}\Big) = \frac{1}{\beta(1/2,n/2)}\,. \qquad \square$$

Osservazione

Può essere utile ricordare l'origine della relazione fondamentale (18.1) che lega fra loro le funzioni gamma e beta di Eulero. Assegnate a piacere due costanti reali positive a e b, vale infatti

$$\Gamma(a)\Gamma(b) = \int_0^{+\infty} e^{-x} x^{a-1}\, dx \int_0^{+\infty} e^{-y} y^{b-1}\, dy$$

e con il doppio cambiamento di variabile $x = u^2$ e $y = v^2$, nel primo e nel secondo integrale rispettivamente, si ottiene

$$\Gamma(a)\Gamma(b) = 2\int_0^{+\infty} du\; u^{2a-1} e^{-u^2}\; 2\int_0^{+\infty} dv\; v^{2b-1} e^{-v^2}\,.$$

Il teorema di Fubini permette di scrivere un unico integrale doppio in luogo dell'integrale iterato

$$\Gamma(a)\Gamma(b) = 4\int_{\mathbb{R}^+\times\mathbb{R}^+} dudv\; e^{-(u^2+v^2)} u^{2a-1} v^{2b-1}\,.$$

Basta ora introdurre un sistema di coordinate polari piane (ρ, θ), definito da

$$\begin{cases} u = \rho\cos\phi \\ v = \rho\sin\phi \end{cases} \qquad \rho \in \mathbb{R}^+ \qquad \phi \in [0, \pi/2]$$

nel quadrante $\mathbb{R}^+ \times \mathbb{R}^+$, per ricavare la relazione

$$\Gamma(a)\Gamma(b) = 4\int_{\mathbb{R}^+\times[0,\pi/2]} d\rho d\phi \; \rho \; e^{-\rho^2}\rho^{2a+2b-2}(\cos\phi)^{2a-1}(\sin\phi)^{2b-1} \, .$$

Questa si separa, sempre per Fubini, in un integrale iterato e quindi in un semplice prodotto di integrali unidimensionali

$$\begin{aligned}\Gamma(a)\Gamma(b) &= 4\int_{\mathbb{R}^+} e^{-\rho^2}\rho^{2a+2b-1}d\rho \int_0^{\pi/2}(\cos\phi)^{2a-2}(\sin\phi)^{2b-2}\cos\phi\sin\phi\, d\phi = \\ &= \int_{\mathbb{R}^+} e^{-\rho^2}\rho^{2(a+b-1)}2\rho\; d\rho \int_0^{\pi/2}(\cos^2\phi)^{a-1}(\sin^2\phi)^{b-1}2\sin\phi\cos\phi\, d\phi\end{aligned}$$

che con le trasformazioni $\zeta = \rho^2$ e $\eta = \sin^2\phi$ si riducono a

$$\Gamma(a)\Gamma(b) = \int_0^{+\infty} e^{-\zeta}\zeta^{a+b-1}d\zeta \int_0^1 (1-\eta)^{a-1}\eta^{b-1}\, d\eta = \Gamma(a+b)\,\beta(a,b)$$

provando il risultato. In particolare, per $a = b = 1/2$ la relazione precedente diviene

$$\begin{aligned}[\Gamma(1/2)]^2 &= \Gamma(1)\,\beta(1/2,1/2) = 1\int_0^1 (1-\eta)^{-1/2}\eta^{-1/2}\, d\eta = \\ &= \int_0^{\pi/2}(1-\sin^2\phi)^{-1/2}(\sin^2\phi)^{-1/2}2\sin\phi\cos\phi\, d\phi = \int_0^{\pi/2} 2\; d\phi = \pi\end{aligned}$$

in modo che $\Gamma(1/2) = \sqrt{\pi}$. Questo risultato è utile nella teoria delle variabili normali, in quanto

$$\Gamma(1/2) = \int_0^{+\infty} e^{-\zeta}\zeta^{-1/2}d\zeta = 2\int_0^{+\infty} e^{-z^2}dz$$

e consente quindi di determinare il corretto fattore di normalizzazione di una qualsivoglia variabile gaussiana.

19. Indipendenza stocastica della varianza stimata e della media stimata di un sistema di variabili gaussiane indipendenti e identicamente distribuite

Questo risultato costituisce una applicazione notevole del teorema di Craig. Si considerano le variabili casuali

$$Q_1 = \sum_{i=1}^n (x_i - \overline{x})^2 = x^T Ax \qquad Q_2 = \overline{x}^2 = x^T Bx \; ,$$

forme quadratiche semidefinite positive delle variabili gaussiane indipendenti a media nulla e varianza unitaria $x^T = (x_1 \ldots x_n)$. Risulta allora

$$Q_1 = \sum_{i=1}^n (x_i - \overline{x})^2 = \sum_{i=1}^n x_i^2 - n\overline{x}^2 = \sum_{i=1}^n x_i^2 - \frac{1}{n}\sum_{i,j=1}^n x_i x_j = \sum_{i,j=1}^n \Big(\delta_{ij} - \frac{1}{n}\Big) x_i x_j$$

ed analogamente

$$Q_2 = \frac{1}{n^2}\sum_{i,j=1}^n x_i x_j \; ,$$

per cui le corrispondenti matrici rappresentative sono

$$A_{ij} = \delta_{ij} - \frac{1}{n} \qquad \text{e} \qquad B_{ij} = \frac{1}{n^2} \quad , \qquad 1 \le i,j \le n \, .$$

Se ne deduce che $\forall\, i,j = 1,\ldots,n$ vale

$$(AB)_{ij} = \sum_{k=1}^n A_{ik}B_{kj} = \sum_{k=1}^n \Big(\delta_{ik} - \frac{1}{n}\Big)\frac{1}{n^2} = \frac{1}{n^2}\Big[\sum_{k=1}^n \delta_{ik} - \sum_{k=1}^n \frac{1}{n}\Big] = \frac{1}{n^2}\Big(1 - \frac{n}{n}\Big) = 0$$

per cui $AB = 0$ ed in virtù del teorema di Craig si può concludere che le variabili Q_1 e Q_2 sono stocasticamente indipendenti

$$p(Q_1 \le \overline{Q}_1, Q_2 \le \overline{Q}_2) = p(Q_1 \le \overline{Q}_1) \cdot p(Q_2 \le \overline{Q}_2) \qquad \forall\, \overline{Q}_1, \overline{Q}_2 \in \mathbb{R}^+ \, .$$

È allora ovvio che stocasticamente indipendenti risultano anche le variabili Q_1 ed $\overline{x}$, come pure la stima della varianza e la stima della media

$$s^2 = \frac{1}{n-1}Q_1 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \overline{x})^2 \qquad \text{e} \qquad \overline{x} \, .$$

Una importante conseguenza di quanto provato è che la variabile casuale

$$t = \sqrt{n-1}\,\overline{x} \Big/ \sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}$$

risulta distribuita secondo una distribuzione di Student a $n-1$ gradi di libertà, in quanto $\overline{x}$ è gaussiana con media nulla e varianza unitaria, mentre $\sum_{i=1}^n (x_i - \overline{x})^2$ è una variabile di $\mathcal{X}^2$ a $n-1$ gradi di libertà. $\square$

Osservazione
Lo stesso argomento consente di verificare immediatamente che nel caso le variabili gaussiane indipendenti e identicamente distribuite $x^T = (x_1, \ldots, x_n)$ abbiano varianza $\sigma^2 > 0$ e media μ, la variabile casuale

$$t = \sqrt{n-1}\,(\overline{x} - \mu) \Big/ \sqrt{\sum_{i=1}^n (x_i - \overline{x})^2} = \frac{\overline{x} - \mu}{s}$$

segue anch'essa una distribuzione di Student a $n-1$ gradi di libertà.

20. Variabile di Student per il test di confronto delle medie di due campioni gaussiani indipendenti di eguale varianza

Siano $y = (y_1, \ldots, y_p)^T$ e $z = (z_1, \ldots, z_q)^T$ due sistemi di variabili gaussiane indipendenti identicamente distribuite, di media μ e varianza σ^2. Indicate allora con $\overline{y}$, $\overline{z}$, s_y^2, s_z^2 medie e varianze stimate sui singoli campioni

$$\overline{y} = \frac{1}{p}\sum_{i=1}^{p} y_i \qquad \overline{z} = \frac{1}{q}\sum_{i=1}^{q} z_i \qquad s_y^2 = \frac{1}{p-1}\sum_{i=1}^{p}(y_i-\overline{y})^2 \qquad s_z^2 = \frac{1}{q-1}\sum_{i=1}^{q}(z_i-\overline{z})^2$$

e con s^2 la varianza stimata sul campione complessivo

$$s^2 = \frac{1}{p+q-2}\big[(p-1)s_y^2 + (q-1)s_z^2\big] = \frac{1}{p+q-2}\left[\sum_{i=1}^{p}(y_i-\overline{y})^2 + \sum_{i=1}^{q}(z_i-\overline{z})^2\right],$$

la variabile casuale

$$t = \sqrt{\frac{pq}{p+q}}\,\frac{\overline{y}-\overline{z}}{s}$$

segue una distribuzione di Student a $p+q-2$ gradi di libertà.

Dimostrazione

La prova del risultato si articola in tre parti. Si tratta di verificare che:

(i) la variabile casuale

$$\xi = \sqrt{\frac{pq}{p+q}}\,\frac{\overline{y}-\overline{z}}{\sigma}$$

è una variabile normale — ossia gaussiana di media nulla e varianza unitaria;

(ii) la variabile casuale

$$\mathcal{X}^2 = \frac{1}{\sigma^2}(p+q-2)s^2 = \frac{1}{\sigma^2}\left[\sum_{i=1}^{p}(y_i-\overline{y})^2 + \sum_{i=1}^{q}(z_i-\overline{z})^2\right]$$

segue una distribuzione di $\mathcal{X}^2$ a $p+q-2$ gradi di libertà;

(iii) le variabili ξ e $\mathcal{X}^2$ sono stocasticamente indipendenti.

Ciò fatto la dimostrazione è completa in quanto, per definizione una variabile di Student a $\nu = p+q-2$ gradi di libertà è data dall'espressione

$$\begin{aligned}\sqrt{\nu}\,\frac{\xi}{\mathcal{X}^2} &= \sqrt{p+q-2}\,\frac{\xi}{\sqrt{\mathcal{X}^2}} = \\ &= \sqrt{p+q-2}\sqrt{\frac{pq}{p+q}}\,\frac{\overline{y}-\overline{z}}{\sigma}\,\frac{1}{\dfrac{1}{\sigma}\sqrt{\displaystyle\sum_{i=1}^{p}(y_i-\overline{y})^2 + \sum_{i=1}^{q}(z_i-\overline{z})^2}} = \sqrt{\frac{pq}{p+q}}\,\frac{\overline{y}-\overline{z}}{s}\end{aligned}$$

che coincide con la variabile casuale t precedentemente introdotta.

In quanto combinazioni lineari di variabili gaussiane di media μ e varianza σ^2, le medie stimate $\overline{y}$ e $\overline{z}$ sono variabili gaussiane pure di media μ e con varianze rispettive

$$\mathbb{E}[(\overline{y}-\mu)^2] \,=\, p\,\frac{\sigma^2}{p^2} \,=\, \frac{\sigma^2}{p} \qquad \text{e} \qquad \mathbb{E}[(\overline{z}-\mu)^2] \,=\, q\,\frac{\sigma^2}{q^2} \,=\, \frac{\sigma^2}{q}$$

in modo che anche la differenza delle medie stimate, $\overline{y}-\overline{z}$, è una variabile gaussiana, di media nulla

$$\mathbb{E}(\overline{y}-\overline{z}) \,=\, \mu-\mu \,=\, 0$$

e varianza

$$\mathbb{E}[(\overline{y}-\overline{z})^2] \,=\, \frac{\sigma^2}{p}+\frac{\sigma^2}{q} \,=\, \Big(\frac{1}{p}+\frac{1}{q}\Big)\sigma^2\,.$$

Ne segue che la differenza normalizzata delle medie

$$\frac{\overline{y}-\overline{z}}{\sigma\sqrt{\dfrac{1}{p}+\dfrac{1}{q}}} \,=\, \sqrt{\frac{pq}{p+q}}\,\frac{\overline{y}-\overline{z}}{\sigma} \,=\, \xi$$

definisce a sua volta una variabile gaussiana di valor medio nullo e varianza unitaria; l'asserto (i) è provato.

Per dimostrare il punto (ii) si scrive preliminarmente la $\mathcal{X}^2$ come forma quadratica delle variabili gaussiane y e z, convenientemente normalizzate

$$\begin{aligned}
\mathcal{X}^2 \,&=\, \sum_{i=1}^{p}\Big(\frac{y_i-\overline{y}}{\sigma}\Big)^2+\sum_{i=1}^{q}\Big(\frac{z_i-\overline{z}}{\sigma}\Big)^2 \,=\\
&=\, \sum_{i,j=1}^{p}\Big(\delta_{ij}-\frac{1}{p}\Big)\frac{y_i-\mu}{\sigma}\frac{y_j-\mu}{\sigma}+\sum_{i,j=1}^{q}\Big(\delta_{ij}-\frac{1}{q}\Big)\frac{z_i-\mu}{\sigma}\frac{z_j-\mu}{\sigma} \,=\\
&=\, \Big(\frac{y-\mu}{\sigma}\Big)^T\Big(\mathbb{I}-\frac{1}{p}\mathbb{S}^{pp}\Big)\,\frac{y-\mu}{\sigma}+\Big(\frac{z-\mu}{\sigma}\Big)^T\Big(\mathbb{I}-\frac{1}{q}\mathbb{S}^{qq}\Big)\,\frac{z-\mu}{\sigma}\\
&=\, \Big(\frac{x-\mu}{\sigma}\Big)^T\mathbb{G}\,\frac{x-\mu}{\sigma}
\end{aligned}$$

espressione nella quale si sono introdotti il vettore delle variabili gaussiane dell'intero campione $x^T = (y^T\,z^T)$, la matrice reale simmetrica e semidefinita positiva $\mathbb{G}$ definita da

$$\mathbb{G} \,=\, \begin{pmatrix} \mathbb{I}-\dfrac{1}{p}\mathbb{S}^{pp} & \mathbb{O} \\ \mathbb{O} & \mathbb{I}-\dfrac{1}{q}\mathbb{S}^{qq} \end{pmatrix}$$

e $\forall\, m,n\in\mathbb{N}$ la matrice $\mathbb{S}^{mn}$ di elementi tutti eguali ad 1

$$\mathbb{S}^{mn}_{ij} \,=\, 1 \qquad \forall\, i=1,\ldots,m \quad j=1,\ldots,n\,.$$

Data l'indipendenza stocastica delle variabili normali $\dfrac{x-\mu}{\sigma}$, la forma quadratica ottenuta costituirà una variabile di $\mathcal{X}^2$ se e soltanto se

$$\mathbb{G}^2 = \mathbb{G}$$

ed, in tal caso, il relativo numero di gradi di libertà sarà dato dalla traccia $\mathrm{tr}\mathbb{G}$ della matrice $\mathbb{G}$. In effetti, si è già verificato al punto 11 che le variabili casuali

$$\sum_{i=1}^{p}(y_i-\overline{y})^2 \qquad \text{e} \qquad \sum_{i=1}^{q}(z_i-\overline{z})^2$$

seguono distribuzioni di $\mathcal{X}^2$ rispettivamente a $p-1$ e $q-1$ gradi di libertà ovvero, equivalentemente, che le loro matrici rappresentative

$$\mathbb{I}-\frac{1}{p}\mathbb{S}^{pp} \qquad \text{e} \qquad \mathbb{I}-\frac{1}{q}\mathbb{S}^{qq}$$

sono idempotenti

$$\Big(\mathbb{I}-\frac{1}{p}\mathbb{S}^{pp}\Big)^2 = \mathbb{I}-\frac{1}{p}\mathbb{S}^{pp} \qquad\qquad \Big(\mathbb{I}-\frac{1}{q}\mathbb{S}^{qq}\Big)^2 = \mathbb{I}-\frac{1}{q}\mathbb{S}^{qq}$$

con tracce rispettive

$$\mathrm{tr}\Big(\mathbb{I}-\frac{1}{p}\mathbb{S}^{pp}\Big) = p-1 \qquad \text{e} \qquad \mathrm{tr}\Big(\mathbb{I}-\frac{1}{q}\mathbb{S}^{qq}\Big) = q-1\,.$$

Ne deriva che

$$\begin{aligned}
\mathbb{G}^2 &= \begin{pmatrix} \mathbb{I}-\frac{1}{p}\mathbb{S}^{pp} & \mathbb{O} \\ \mathbb{O} & \mathbb{I}-\frac{1}{q}\mathbb{S}^{qq} \end{pmatrix}^2 = \begin{pmatrix} \Big(\mathbb{I}-\frac{1}{p}\mathbb{S}^{pp}\Big)^2 & \mathbb{O} \\ \mathbb{O} & \Big(\mathbb{I}-\frac{1}{q}\mathbb{S}^{qq}\Big)^2 \end{pmatrix} = \\
&= \begin{pmatrix} \mathbb{I}-\frac{1}{p}\mathbb{S}^{pp} & \mathbb{O} \\ \mathbb{O} & \mathbb{I}-\frac{1}{q}\mathbb{S}^{qq} \end{pmatrix} = \mathbb{G}
\end{aligned}$$

e

$$\mathrm{tr}\begin{pmatrix} \mathbb{I}-\frac{1}{p}\mathbb{S}^{pp} & \mathbb{O} \\ \mathbb{O} & \mathbb{I}-\frac{1}{q}\mathbb{S}^{qq} \end{pmatrix} = \mathrm{tr}\Big(\mathbb{I}-\frac{1}{p}\mathbb{S}^{pp}\Big)+\mathrm{tr}\Big(\mathbb{I}-\frac{1}{q}\mathbb{S}^{qq}\Big) = p+q-2$$

per cui $\mathcal{X}^2$ è effettivamente una variabile di $\mathcal{X}^2$ a $p+q-2$ gradi di libertà, come si voleva dimostrare.

Verificare l'indipendenza stocastica delle variabili ξ e $\mathcal{X}^2$ — asserto (iii) — equivale a provare l'indipendenza di ξ^2 e $\mathcal{X}^2$. Il vantaggio è dato dal fatto che ξ^2 costituisce una forma quadratica semidefinita positiva delle variabili normali indipendenti $(x-\mu)^T/\sigma$, al pari della $\mathcal{X}^2$, in modo che l'indipendenza stocastica può essere caratterizzata mediante il teorema di Craig.

A questo scopo la variabile ξ^2 si scrive

$$\xi^2 \;=\; \frac{pq}{p+q}\Big(\frac{\overline{y}-\overline{z}}{\sigma}\Big)^2 \;=\; \frac{pq}{p+q}\Big[\frac{\overline{y}-\mu-(\overline{z}-\mu)}{\sigma}\Big]^2 \;=\; \frac{pq}{p+q}\Big[\frac{1}{p}\sum_{i=1}^{p}\frac{y_i-\mu}{\sigma}-\frac{1}{q}\sum_{j=1}^{q}\frac{z_j-\mu}{\sigma}\Big]^2$$

ossia

$$\xi^2 \;=\; \frac{pq}{p+q}\Big[\frac{1}{p^2}\Big(\sum_{i=1}^{p}\frac{y_i-\mu}{\sigma}\Big)^2+\frac{1}{q^2}\Big(\sum_{j=1}^{q}\frac{z_j-\mu}{\sigma}\Big)^2-2\frac{1}{pq}\sum_{i=1}^{p}\frac{y_i-\mu}{\sigma}\sum_{j=1}^{q}\frac{z_j-\mu}{\sigma}\Big]$$

che può anche esprimersi nella forma matriciale

$$\xi^2 \;=\; \Big(\frac{x-\mu}{\sigma}\Big)^T\mathbb{H}\,\frac{x-\mu}{\sigma}$$

in termini della matrice reale simmetrica e semidefinita positiva

$$\mathbb{H} \;=\; \frac{pq}{p+q}\begin{pmatrix} \dfrac{1}{p^2}\mathbb{S}^{pp} & -\dfrac{1}{pq}\mathbb{S}^{pq} \\ -\dfrac{1}{qp}\mathbb{S}^{qp} & \dfrac{1}{q^2}\mathbb{S}^{qq}\end{pmatrix}.$$

Per il teorema di Craig provare l'indipendenza stocastica di ξ^2 e $\mathcal{X}^2$ equivale a dimostrare l'annullarsi del prodotto matriciale $\mathbb{G}\,\mathbb{H}$, come puntualmente si verifica

$$\begin{aligned}\mathbb{G}\,\mathbb{H} \;&=\; \frac{pq}{p+q}\begin{pmatrix}\mathbb{I}-\dfrac{1}{p}\mathbb{S}^{pp} & \mathbb{O}\\ \mathbb{O} & \mathbb{I}-\dfrac{1}{q}\mathbb{S}^{qq}\end{pmatrix}\begin{pmatrix} \dfrac{1}{p^2}\mathbb{S}^{pp} & -\dfrac{1}{pq}\mathbb{S}^{pq} \\ -\dfrac{1}{qp}\mathbb{S}^{qp} & \dfrac{1}{q^2}\mathbb{S}^{qq}\end{pmatrix} \;=\\ &=\; \frac{pq}{p+q}\begin{pmatrix}\dfrac{1}{p^2}\Big(\mathbb{S}^{pp}-\dfrac{1}{p}\mathbb{S}^{pp}\mathbb{S}^{pp}\Big) & -\dfrac{1}{pq}\Big(\mathbb{S}^{pq}-\dfrac{1}{p}\mathbb{S}^{pp}\mathbb{S}^{pq}\Big)\\ -\dfrac{1}{qp}\Big(\mathbb{S}^{qp}-\dfrac{1}{q}\mathbb{S}^{qq}\mathbb{S}^{qp}\Big) & \dfrac{1}{q^2}\Big(\mathbb{S}^{qq}-\dfrac{1}{q}\mathbb{S}^{qq}\mathbb{S}^{qq}\Big)\end{pmatrix} \;=\; 0\end{aligned}$$

essendo

$$\begin{aligned}\frac{1}{p}\mathbb{S}^{pp}\mathbb{S}^{pp} \;&=\; \mathbb{S}^{pp} & \qquad \frac{1}{p}\mathbb{S}^{pp}\mathbb{S}^{pq} \;&=\; \mathbb{S}^{pq}\\ \frac{1}{q}\mathbb{S}^{qq}\mathbb{S}^{qp} \;&=\; \mathbb{S}^{qp} & \qquad \frac{1}{q}\mathbb{S}^{qq}\mathbb{S}^{qq} \;&=\; \mathbb{S}^{qq}\,.\end{aligned}$$

La dimostrazione è completa. □

21. Definizione. Variabile F di Fisher a (n_1, n_2) gradi di libertà

La variabile F di Fisher a (n_1, n_2) gradi di libertà è definita dal quoziente

$$F \;=\; \frac{n_2}{n_1}\frac{\mathcal{X}_1^2}{\mathcal{X}_2^2} \;=\; \frac{\mathcal{X}_1^2/n_1}{\mathcal{X}_2^2/n_2} \geq 0$$

in cui i simboli $\mathcal{X}_1^2$ e $\mathcal{X}_2^2$ designano variabili di chi quadrato stocasticamente indipendenti a n_1 ed n_2 gradi di libertà rispettivamente.

Le variabili $\mathcal{X}_1^2/n_1$ e $\mathcal{X}_2^2/n_2$ vengono anche denominate variabili di chi quadrato **ridotte**. Pertanto una variabile di Fisher può essere intesa come il quoziente di due variabili di chi quadrato ridotte ed indipendenti.

22. Distribuzione di probabilità di una variabile F di Fisher a (n_1, n_2) gradi di libertà

La distribuzione di probabilità di una variabile F di Fisher a (n_1, n_2) gradi di libertà si scrive

$$p_{(n_1,n_2)}(F) = \frac{\Gamma\Big(\dfrac{n_1+n_2}{2}\Big)}{\Gamma\Big(\dfrac{n_1}{2}\Big)\,\Gamma\Big(\dfrac{n_2}{2}\Big)}\Big(\frac{n_1}{n_2}\Big)^{\frac{n_1}{2}}\frac{F^{\frac{n_1}{2}-1}}{\Big(1+\dfrac{n_1}{n_2}F\Big)^{\frac{n_1+n_2}{2}}}$$

essendo al solito Γ la funzione gamma euleriana.

Dimostrazione

Conviene partire dalla condizione di normalizzazione per le due variabili di chi quadrato

$$1 = \int_0^{+\infty} d(\mathcal{X}_1^2)\frac{1}{2^{\frac{n_1}{2}}\Gamma\Big(\dfrac{n_1}{2}\Big)}e^{-\mathcal{X}_1^2/2}(\mathcal{X}_1^2)^{\frac{n_1}{2}-1}\int_0^{+\infty} d(\mathcal{X}_2^2)\frac{1}{2^{\frac{n_2}{2}}\Gamma\Big(\dfrac{n_2}{2}\Big)}e^{-\mathcal{X}_2^2/2}(\mathcal{X}_2^2)^{\frac{n_2}{2}-1}$$

dalla quale, per Fubini, si deduce la relazione

$$2^{\frac{n_1+n_2}{2}}\Gamma\Big(\frac{n_1}{2}\Big)\Big(\frac{n_2}{2}\Big) = \int_{\mathbb{R}^+\times\mathbb{R}^+} d(\mathcal{X}_1^2)d(\mathcal{X}_2^2)\; e^{-\frac{1}{2}(\mathcal{X}_1^2+\mathcal{X}_2^2)}(\mathcal{X}_1^2)^{\frac{n_1}{2}-1}(\mathcal{X}_2^2)^{\frac{n_2}{2}-1}\,.$$

L'integrale doppio a secondo membro viene riscritto, sempre per Fubini, come integrale iterato

$$2^{\frac{n_1+n_2}{2}}\Gamma\Big(\frac{n_1}{2}\Big)\Big(\frac{n_2}{2}\Big) = \int_{\mathbb{R}^+} d(\mathcal{X}_2^2)\int_{\mathbb{R}^+} d(\mathcal{X}_1^2)e^{-\frac{1}{2}(\mathcal{X}_1^2+\mathcal{X}_2^2)}(\mathcal{X}_1^2)^{\frac{n_1}{2}-1}(\mathcal{X}_2^2)^{\frac{n_2}{2}-1}$$

ed in quello più interno si introduce, a $\mathcal{X}_2^2 > 0$ fissato, il cambiamento di variabile $\mathcal{X}_1^2 \to F$ definito da

$$\mathcal{X}_1^2 = \frac{n_1}{n_2}\mathcal{X}_2^2\; F$$

che porge così

$$\begin{aligned}
&2^{\frac{n_1+n_2}{2}}\Gamma\Big(\frac{n_1}{2}\Big)\Big(\frac{n_2}{2}\Big) = \\
&= \int_{\mathbb{R}^+} d(\mathcal{X}_2^2)\int_{\mathbb{R}^+} dF\;\frac{n_1}{n_2}\mathcal{X}_2^2 e^{-\frac{1}{2}\left(1+\frac{n_1}{n_2}F\right)\mathcal{X}_2^2}\Big(\frac{n_1}{n_2}\Big)^{\frac{n_1}{2}-1}F^{\frac{n_1}{2}-1}(\mathcal{X}_2^2)^{\frac{n_1+n_2}{2}-2} = \\
&= \int_{\mathbb{R}^+} dF\int_{\mathbb{R}^+} d(\mathcal{X}_2^2)\;\Big(\frac{n_1}{n_2}\Big)^{\frac{n_1}{2}}(\mathcal{X}_2^2)^{\frac{n_1+n_2}{2}-1}F^{\frac{n_1}{2}-1}e^{-\frac{1}{2}\left(1+\frac{n_1}{n_2}F\right)\mathcal{X}_2^2} = \\
&= \int_{\mathbb{R}^+} dF\;\Big(\frac{n_1}{n_2}\Big)^{\frac{n_1}{2}}F^{\frac{n_1}{2}-1}\int_{\mathbb{R}^+} d(\mathcal{X}_2^2)\;(\mathcal{X}_2^2)^{\frac{n_1+n_2}{2}-1}e^{-\frac{1}{2}\left(1+\frac{n_1}{n_2}F\right)\mathcal{X}_2^2}
\end{aligned}$$

in cui lo scambio nell'ordine delle integrazioni è lecito in virtù del teorema di Fubini. Se poi $\forall\, F \in \mathbb{R}^+$ fissato si pone

$$z = \Big(1+\frac{n_1}{n_2}F\Big)\mathcal{X}_2^2$$

nell'integrale più interno a secondo membro, la relazione precedente diventa

$$\Gamma\Big(\frac{n_1}{2}\Big)\Gamma\Big(\frac{n_2}{2}\Big) = \int_{\mathbb{R}+} dF \, \Big(\frac{n_1}{n_2}\Big)^{\frac{n_1}{2}} F^{\frac{n_1}{2}-1} \int_{\mathbb{R}+} \frac{dz}{2} \, \Big(\frac{z}{2}\Big)^{\frac{n_1+n_2}{2}-1} e^{-\frac{z}{2}} \frac{1}{\Big(1+\dfrac{n_1}{n_2}F\Big)^{\frac{n_1+n_2}{2}}}$$

e da essa si ricava

$$\Gamma\Big(\frac{n_1}{2}\Big)\Gamma\Big(\frac{n_2}{2}\Big) = \int_{\mathbb{R}+} dF \, \Big(\frac{n_1}{n_2}\Big)^{\frac{n_1}{2}} F^{\frac{n_1}{2}-1} \frac{1}{\Big(1+\dfrac{n_1}{n_2}F\Big)^{\frac{n_1+n_2}{2}}} \int_{\mathbb{R}+} du \; u^{\frac{n_1+n_2}{2}-1} e^{-u} \, .$$

Riconoscendo nell'integrale in u la definizione della funzione gamma euleriana calcolata in $(n_1+n_2)/2$, si perviene infine all'espressione

$$\Gamma\Big(\frac{n_1}{2}\Big)\Gamma\Big(\frac{n_2}{2}\Big) = \int_{\mathbb{R}+} dF \, \Big(\frac{n_1}{n_2}\Big)^{\frac{n_1}{2}} F^{\frac{n_1}{2}-1} \frac{1}{\Big(1+\dfrac{n_1}{n_2}F\Big)^{\frac{n_1+n_2}{2}}} \Gamma\Big(\frac{n_1+n_2}{2}\Big)$$

che fornisce la densità di probabilità cercata

$$p_{(n_1,n_2)}(F) = \frac{\Gamma\Big(\dfrac{n_1+n_2}{2}\Big)}{\Gamma\Big(\dfrac{n_1}{2}\Big)\,\Gamma\Big(\dfrac{n_2}{2}\Big)} \Big(\frac{n_1}{n_2}\Big)^{\frac{n_1}{2}} \frac{F^{\frac{n_1}{2}-1}}{\Big(1+\dfrac{n_1}{n_2}F\Big)^{\frac{n_1+n_2}{2}}} \, . \qquad \square$$

23. Proprietà di reciprocità delle variabili di Fisher

Se F è una variabile di Fisher a $(n_1,n_2) \in \mathbb{N}^2$ gradi di libertà, la variabile reciproca

$$F' = \frac{1}{F}$$

segue una distribuzione di Fisher a (n_2,n_1) gradi di libertà.

Conseguenza immediata del teorema è che per tutte le applicazioni risulta sufficiente calcolare e/o tabulare soltanto le distribuzioni di Fisher ad (n_1,n_2) gradi di libertà con $n_1 \le n_2$.

Dimostrazione
Sia F una variabile di Fisher a (n_1,n_2) gradi di libertà. Se si pone $F = 1/F'$ nell'integrale di normalizzazione della distribuzione di F si ottiene

$$1 = \int_0^{+\infty} dF \, \frac{\Gamma\Big(\dfrac{n_1+n_2}{2}\Big)}{\Gamma\Big(\dfrac{n_1}{2}\Big)\,\Gamma\Big(\dfrac{n_2}{2}\Big)} \Big(\frac{n_1}{n_2}\Big)^{\frac{n_1}{2}} \frac{F^{\frac{n_1}{2}-1}}{\Big(1+\dfrac{n_1}{n_2}F\Big)^{\frac{n_1+n_2}{2}}} =$$

$$= \int_0^{+\infty} dF' \, \frac{1}{F'^2} \frac{\Gamma\Big(\dfrac{n_1+n_2}{2}\Big)}{\Gamma\Big(\dfrac{n_1}{2}\Big)\,\Gamma\Big(\dfrac{n_2}{2}\Big)} \Big(\frac{n_1}{n_2}\Big)^{\frac{n_1}{2}} \frac{(F')^{-\frac{n_1}{2}+1}}{\Big(1+\dfrac{n_1}{n_2}\dfrac{1}{F'}\Big)^{\frac{n_1+n_2}{2}}} =$$

$$= \int_0^{+\infty} dF' \, \frac{\Gamma\Big(\dfrac{n_1+n_2}{2}\Big)}{\Gamma\Big(\dfrac{n_1}{2}\Big)\,\Gamma\Big(\dfrac{n_2}{2}\Big)} \Big(\frac{n_1}{n_2}\Big)^{\frac{n_1}{2}} \frac{(F')^{-\frac{n_1}{2}-1}}{\Big(\dfrac{n_1}{n_2}\dfrac{1}{F'}\Big)^{\frac{n_1+n_2}{2}} \Big(1+\dfrac{n_2}{n_1}F'\Big)^{\frac{n_1+n_2}{2}}} =$$

$$= \int_0^{+\infty} dF' \, \frac{\Gamma\Big(\dfrac{n_1+n_2}{2}\Big)}{\Gamma\Big(\dfrac{n_1}{2}\Big)\,\Gamma\Big(\dfrac{n_2}{2}\Big)} \Big(\frac{n_2}{n_1}\Big)^{\frac{n_2}{2}} \frac{F'^{\frac{n_2}{2}-1}}{\Big(1+\dfrac{n_2}{n_1}F'\Big)^{\frac{n_1+n_2}{2}}}$$

per cui la distribuzione di probabilità di $F' = 1/F$

$$p(F') = \frac{\Gamma\Big(\dfrac{n_1+n_2}{2}\Big)}{\Gamma\Big(\dfrac{n_1}{2}\Big)\,\Gamma\Big(\dfrac{n_2}{2}\Big)} \Big(\frac{n_2}{n_1}\Big)^{\frac{n_2}{2}} \frac{F'^{\frac{n_2}{2}-1}}{\Big(1+\dfrac{n_2}{n_1}F'\Big)^{\frac{n_1+n_2}{2}}}$$

si riconosce essere una distribuzione di Fisher a (n_2, n_1) gradi di libertà. $\square$

24. Consistenza della stima $\overline{x}$ della media μ di una variabile casuale.

Sia dato un campione $x_1, \ldots, x_n$ di n dati estratti da una popolazione statistica, che si assumerà descritta da una variabile casuale x con distribuzione di probabilità $p(x)$, di media μ e varianza σ^2 finite. Allora la media del campione

$$\overline{x} = \overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$$

fornisce una stima consistente del valor medio μ della variabile casuale, intendendosi con ciò che:

(i) $\mathbb{E}(\overline{x}_n) = \mu \;\forall\, n \in \mathbb{N}$;

(ii) per $n \to +\infty$ la variabile casuale $\overline{x} = \overline{x}_n$ converge in probabilità al valore μ

$$\forall\, \varepsilon > 0 \qquad \lim_{n\to+\infty} p[\mu-\varepsilon \le \overline{x}_n \le \mu+\varepsilon] = 1$$

dove $p[\mu-\varepsilon \le \overline{x}_n \le \mu+\varepsilon]$ indica la probabilità che $\overline{x}_n$ assuma un qualsiasi valore compreso fra $\mu-\varepsilon$ e $\mu+\varepsilon$.

Dimostrazione

La verifica della prima parte è immediata e segue dalla linearità della media

$$\mathbb{E}(\overline{x}) = \mathbb{E}(\overline{x}_n) = \mathbb{E}\Big[\frac{1}{n}\sum_{i=1}^{n} x_i\Big] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(x_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{n}n\mu = \mu$$

Per la convergenza in probabilità di $\overline{x}_n$ a μ è sufficiente dimostrare che la variabile casuale $\overline{x}_n$ ha varianza finita e tendente a zero per $n \to +\infty$. Data l'indipendenza stocastica delle variabili casuali x_i, vale infatti:

$$\begin{aligned} \mathrm{var}(\overline{x}_n) \,=\, \mathbb{E}[(\overline{x}_n-\mu)^2] \,=\, \mathbb{E}\Big[\frac{1}{n^2}\Big[\sum_{i=1}^n (x_i-\mu)\Big]^2\Big] \,=\, \frac{1}{n^2}\sum_{ij=1}^n \mathbb{E}\big[(x_i-\mu)(x_j-\mu)\big] \,=\, \\ =\, \frac{1}{n^2}\sum_{ij=1}^n \delta_{ij}\sigma^2 \,=\, \frac{1}{n^2}\sum_{i=1}^n \sigma^2 \,=\, \frac{1}{n^2}\,n\sigma^2 \,=\, \frac{\sigma^2}{n} \end{aligned}$$

per cui comunque si assegni un $\varepsilon > 0$, dal teorema di Tchebyshev segue che

$$p[\mu-\varepsilon \le \overline{x}_n \le \mu+\varepsilon] \,=\, p\Big[\mu - \frac{\varepsilon\sqrt{n}}{\sigma}\frac{\sigma}{\sqrt{n}} \le \overline{x}_n \le \mu + \frac{\varepsilon\sqrt{n}}{\sigma}\frac{\sigma}{\sqrt{n}}\Big] \ge 1 - \frac{1}{\Big(\frac{\varepsilon\sqrt{n}}{\sigma}\Big)^2} \,=\, 1 - \frac{\sigma^2}{\varepsilon^2}\frac{1}{n}$$

Dovendo poi essere

$$1 - \frac{\sigma^2}{\varepsilon^2}\frac{1}{n} \le p[\mu-\varepsilon \le \overline{x}_n \le \mu+\varepsilon] \le 1$$

e risultando

$$\lim_{n\to+\infty} 1 - \frac{\sigma^2}{\varepsilon^2}\frac{1}{n} \,=\, 1$$

se ne conclude che

$$\lim_{n\to+\infty} p[\mu-\varepsilon \le \overline{x}_n \le \mu+\varepsilon] \,=\, 1$$

in modo che l'asserto di convergenza in probabilità è dimostrato. $\square$

25. Consistenza della stima s^2 della varianza σ^2 di una variabile casuale.

Sia dato un campione $x_1,\ldots,x_n$ di n dati estratti da una popolazione statistica, che si assumerà descritta da una variabile casuale x con distribuzione di probabilità $p(x)$, di media μ, varianza σ^2 e momento quarto $\mathbb{E}[(x-\mu)^4]$ finiti. Allora la varianza stimata sul campione

$$s^2 \,=\, s_n^2 \,=\, \frac{1}{n-1}\sum_{i=1}^n (x_i-\overline{x})^2$$

fornisce una stima consistente della varianza σ^2 della variabile casuale, intendendosi con ciò che:

(i) $\mathbb{E}(s_n^2) \,=\, \sigma^2\ \forall\, n \in \mathbb{N}$;

(ii) nel limite $n \to +\infty$ la variabile casuale $s^2 = s_n^2$ converge in probabilità al valore σ^2

$$\forall\,\varepsilon > 0 \qquad \lim_{n\to+\infty} p[\sigma^2-\varepsilon \le s_n^2 \le \sigma^2+\varepsilon] \,=\, 1$$

dove $p[\sigma^2-\varepsilon \le s_n^2 \le \sigma^2+\varepsilon]$ indica la probabilità che s_n^2 assuma un qualsiasi valore compreso fra $\sigma^2-\varepsilon$ e $\sigma^2+\varepsilon$.

Dimostrazione

Al solito, molto semplice è la verifica della prima proprietà

$$
\begin{aligned}
\mathbb{E}(s^2) \;=\; \mathbb{E}(s_n^2) \;&=\; \frac{1}{n-1}\mathbb{E}\Big[\sum_{i=1}^{n}(x_i-\overline{x})^2\Big] \;= \\
&=\; \frac{1}{n-1}\mathbb{E}\Big[\sum_{i=1}^{n}(x_i-\mu)^2 - 2\sum_{i=1}^{n}(x_i-\mu)(\overline{x}-\mu) + \sum_{i=1}^{n}(\overline{x}-\mu)^2\Big] \;= \\
&=\; \frac{1}{n-1}\mathbb{E}\Big[\sum_{i=1}^{n}(x_i-\mu)^2 - 2n(\overline{x}-\mu)^2 + n(\overline{x}-\mu)^2\Big] \;= \\
&=\; \frac{1}{n-1}\mathbb{E}\Big[\sum_{i=1}^{n}(x_i-\mu)^2 - n(\overline{x}-\mu)^2\Big] \;= \\
&=\; \frac{1}{n-1}\Big[\sum_{i=1}^{n}\mathbb{E}[(x_i-\mu)^2] - n\mathbb{E}[(\overline{x}-\mu)^2]\Big] \;= \\
&=\; \frac{1}{n-1}\Big[\sum_{i=1}^{n}\sigma^2 - n\frac{\sigma^2}{n}\Big] \;=\; \frac{1}{n-1}\big[n\sigma^2-\sigma^2\big] \;=\; \sigma^2 \,.
\end{aligned}
$$

La convergenza in probabilità di s_n^2 a σ^2 viene provata anche in questo caso dimostrando che la varianza di s_n^2 è finita $\forall\, n \in \mathbb{N}$ e infinitesima per $n \to +\infty$. Vale infatti

$$
\begin{aligned}
&\mathrm{var}(s_n^2) \;=\; \mathbb{E}\bigg[\Big(\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2 - \sigma^2\Big)^2\bigg] \;= \\
&= \mathbb{E}\bigg[\Big(\frac{1}{n-1}\Big(\sum_{i=1}^{n}(x_i-\mu)^2 - n(\overline{x}-\mu)^2\Big) - \sigma^2\Big)^2\bigg] \;= \\
&= \mathbb{E}\bigg[\frac{1}{(n-1)^2}\Big(\sum_{i=1}^{n}(x_i-\mu)^2 - n(\overline{x}-\mu)^2\Big)^2 - \frac{2\sigma^2}{n-1}\Big(\sum_{i=1}^{n}(x_i-\mu)^2 - n(\overline{x}-\mu)^2\Big) + \sigma^4\bigg] = \\
&= \frac{1}{(n-1)^2}\mathbb{E}\bigg[\Big(\sum_{i=1}^{n}(x_i-\mu)^2 - n(\overline{x}-\mu)^2\Big)^2\bigg] - \frac{2\sigma^2}{n-1}\Big(n\sigma^2 - n\,\frac{\sigma^2}{n}\Big) + \sigma^4 \;= \\
&= \frac{1}{(n-1)^2}\mathbb{E}\bigg[\Big(\sum_{i=1}^{n}(x_i-\mu)^2 - n(\overline{x}-\mu)^2\Big)^2\bigg] - \sigma^4 \,.
\end{aligned}
$$

L'aspettazione residua può essere calcolata nel modo seguente

$$
\begin{aligned}
&\mathbb{E}\bigg[\Big(\sum_{i=1}^{n}(x_i-\mu)^2 - n(\overline{x}-\mu)^2\Big)^2\bigg] \;=\; \mathbb{E}\bigg[\Big(\sum_{ij=1}^{n}\Big(\delta_{ij}-\frac{1}{n}\Big)(x_i-\mu)(x_j-\mu)\Big)^2\bigg] \;= \\
&=\mathbb{E}\bigg[\sum_{ijkl=1}^{n}\Big(\delta_{ij}-\frac{1}{n}\Big)\Big(\delta_{kl}-\frac{1}{n}\Big)(x_i-\mu)(x_j-\mu)(x_k-\mu)(x_l-\mu)\bigg] \;= \\
&= \sum_{ijkl=1}^{n}\Big(\delta_{ij}-\frac{1}{n}\Big)\Big(\delta_{kl}-\frac{1}{n}\Big)\mathbb{E}[(x_i-\mu)(x_j-\mu)(x_k-\mu)(x_l-\mu)]
\end{aligned}
$$

ricordando che per l'indipendenza stocastica delle variabili x_i i soli momenti centrali quarti non banalmente nulli sono quelli per i quali gli indici (i,j,k,l) risultano o tutti coincidenti o coincidenti due a due.
Nella sommatoria precedente si hanno così:

– n termini con $i=j=k=l$, per cui risulta

$$\Big(\delta_{ij}-\frac{1}{n}\Big)\Big(\delta_{kl}-\frac{1}{n}\Big)\mathbb{E}[(x_i-\mu)(x_j-\mu)(x_k-\mu)(x_l-\mu)] \;=\; \Big(1-\frac{1}{n}\Big)^2\mathbb{E}[(x_i-\mu)^4]\ ;$$

– $n(n-1)$ termini con $i=j$, $j\neq k$, $k=l$, per i quali si ottiene

$$\Big(\delta_{ij}-\frac{1}{n}\Big)\Big(\delta_{kl}-\frac{1}{n}\Big)\mathbb{E}[(x_i-\mu)(x_j-\mu)(x_k-\mu)(x_l-\mu)] \;=\; \Big(1-\frac{1}{n}\Big)^2\sigma^4$$

– $n(n-1)$ termini del tipo $i=k$, $j=l$, $i\neq j$, per cui vale

$$\Big(\delta_{ij}-\frac{1}{n}\Big)\Big(\delta_{kl}-\frac{1}{n}\Big)\mathbb{E}[(x_i-\mu)(x_j-\mu)(x_k-\mu)(x_l-\mu)] \;=\; \frac{1}{n^2}\sigma^4$$

– ed infine $n(n-1)$ termini caratterizzati da $i=l$, $j=k$, $i\neq j$, con ancora

$$\Big(\delta_{ij}-\frac{1}{n}\Big)\Big(\delta_{kl}-\frac{1}{n}\Big)\mathbb{E}[(x_i-\mu)(x_j-\mu)(x_k-\mu)(x_l-\mu)] \;=\; \frac{1}{n^2}\sigma^4\,.$$

Il risultato cui si perviene è pertanto

$$\mathbb{E}\bigg[\Big(\sum_{i=1}^{n}(x_i-\mu)^2-n(\overline{x}-\mu)^2\Big)^2\bigg] \;=$$

$$\begin{aligned}
&= n\Big(1-\frac{1}{n}\Big)^2\mathbb{E}[(x_i-\mu)^4]+n(n-1)\Big(1-\frac{1}{n}\Big)^2\sigma^4+n(n-1)\frac{1}{n^2}\sigma^4+n(n-1)\frac{1}{n^2}\sigma^4 \;=\\
&= \frac{(n-1)^2}{n}\mathbb{E}[(x_i-\mu)^4]+n(n-1)\Big(1-\frac{2}{n}+\frac{1}{n^2}+\frac{2}{n^2}\Big)\sigma^4 \;=\\
&= \frac{(n-1)^2}{n}\mathbb{E}[(x_i-\mu)^4]+n(n-1)\Big(1-\frac{2}{n}+\frac{3}{n^2}\Big)\sigma^4
\end{aligned}$$

in modo che

$$\begin{aligned}
\mathrm{var}(s_n^2) &= \frac{1}{n}\mathbb{E}[(x_i-\mu)^4]+\frac{n}{n-1}\Big(1-\frac{2}{n}+\frac{3}{n^2}\Big)\sigma^4-\sigma^4 \;=\\
&= \frac{1}{n}\mathbb{E}[(x_i-\mu)^4]+\frac{n^2-2n+3}{n(n-1)}\sigma^4-\sigma^4 \;=\\
&= \frac{1}{n}\mathbb{E}[(x_i-\mu)^4]+\frac{n^2-2n+3-n^2+n}{n(n-1)}\sigma^4 \;=\\
&= \frac{1}{n}\mathbb{E}[(x_i-\mu)^4]+\frac{3-n}{n(n-1)}\sigma^4
\end{aligned}$$

successione convergente a zero per $n \to +\infty$, grazie all'ipotesi $\mathbb{E}[(x_i - \mu)^4] < \infty$. Dal teorema di Tchebyshev si conclude allora che per ogni $\varepsilon > 0$ assegnato deve aversi

$$p[\sigma^2 - \varepsilon \leq s_n^2 \leq \sigma^2 + \varepsilon] = p\left[\sigma^2 - \frac{\varepsilon}{\sqrt{\operatorname{var}(s_n^2)}}\sqrt{\operatorname{var}(s_n^2)} \leq s_n^2 \leq \sigma^2 + \frac{\varepsilon}{\sqrt{\operatorname{var}(s_n^2)}}\sqrt{\operatorname{var}(s_n^2)}\right]$$

$$\geq 1 - \frac{1}{\left(\frac{\varepsilon}{\sqrt{\operatorname{var}(s_n^2)}}\right)^2} = 1 - \frac{1}{\varepsilon^2}\operatorname{var}(s_n^2) \xrightarrow[n\to+\infty]{} 1$$

e dunque

$$\lim_{n\to+\infty} p[\sigma^2 - \varepsilon \leq s_n^2 \leq \sigma^2 + \varepsilon] = 1$$

come richiesto. $\square$

26. Prodotto tensoriale di spazi vettoriali reali di dimensione finita

Prodotto tensoriale di $\mathbb{R}^a$ e $\mathbb{R}^b$

Dati gli spazi vettoriali $\mathbb{R}^a$ e $\mathbb{R}^b$, si definisce prodotto tensoriale di $\mathbb{R}^a$ e $\mathbb{R}^b$, e lo si indica con $\mathbb{R}^a \otimes \mathbb{R}^b$, lo spazio vettoriale di elementi

$$x_{ij} \in \mathbb{R} \qquad i = 1, \ldots, a \quad j = 1, \ldots, b$$

munito dell'operazione di somma che $\forall\, x, y \in \mathbb{R}^a \otimes \mathbb{R}^b$ è definita da

$$(x + y)_{ij} = x_{ij} + y_{ij} \qquad i = 1, \ldots, a \quad j = 1, \ldots, b$$

e dell'operazione di prodotto per uno scalare che $\forall\, \lambda \in \mathbb{R}$ e $x \in \mathbb{R}^a \otimes \mathbb{R}^b$ si esprime come

$$(\lambda x)_{ij} = \lambda x_{ij} \qquad i = 1, \ldots, a \quad j = 1, \ldots, b\,.$$

Prodotto tensoriale di vettori di $\mathbb{R}^a$ e $\mathbb{R}^b$

Indicate con $e_1, \ldots, e_a$ e $f_1, \ldots, f_b$ rispettivamente una base di $\mathbb{R}^a$ e una base di $\mathbb{R}^b$, per ogni $x \in \mathbb{R}^a$ e per ogni $y \in \mathbb{R}^b$ vale ovviamente

$$x = \sum_{i=1}^{a} x_i e_i \qquad \sum_{j=1}^{b} y_j f_j$$

e si può allora introdurre la nozione di prodotto tensoriale $x \otimes y$ fra i due vettori x e y, definito da

$$(x \otimes y)_{ij} \stackrel{\text{def}}{=} x_i y_j \qquad i = 1, \ldots, a \quad j = 1, \ldots, b\,.$$

Base e dimensione di $\mathbb{R}^a \otimes \mathbb{R}^b$

I prodotti tensoriali fra i vettori di base

$$e_i \otimes f_j \qquad i = 1, \ldots, a \quad j = 1, \ldots, b$$

corrispondono ai tensori di $\mathbb{R}^a \otimes \mathbb{R}^b$ le cui ab componenti sono tutte nulle salvo una

$$(e_i \otimes f_j)_{hk} = \delta_{ih}\delta_{jk} \qquad i = 1, \ldots, a \quad j = 1, \ldots, b$$

ed è immediato verificare che l'insieme di tali prodotti tensoriali

$$\{e_i \otimes f_j \ , \quad i = 1, \ldots, a\,, \ j = 1, \ldots, b\}$$

costituisce una base di $\mathbb{R}^a \otimes \mathbb{R}^b$, sicché

$$\dim(\mathbb{R}^a \otimes \mathbb{R}^b) = \dim(\mathbb{R}^a)\dim(\mathbb{R}^b) = ab\,.$$

Notazione
Per ogni $x \in \mathbb{R}^a \otimes \mathbb{R}^b$ si scrive formalmente la sommatoria simbolica

$$x = \sum_{i=1}^{a}\sum_{j=1}^{b} x_{ij} e_i \otimes f_j$$

e le operazioni dello spazio vettoriale diventano

$$x + y = \sum_{i=1}^{a}\sum_{j=1}^{b} x_{ij} e_i \otimes f_j + \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij} e_i \otimes f_j \stackrel{\text{def}}{=} \sum_{i=1}^{a}\sum_{j=1}^{b} (x_{ij} + y_{ij}) e_i \otimes f_j$$

$$\lambda x = \lambda\Big(\sum_{i=1}^{a}\sum_{j=1}^{b} x_{ij} e_i \otimes f_j\Big) \stackrel{\text{def}}{=} \Big(\sum_{i=1}^{a}\sum_{j=1}^{b} \lambda x_{ij} e_i \otimes f_j\Big)\,,$$

mentre il prodotto tensoriale di due vettori assume la forma

$$x \otimes y = \sum_{i=1}^{a}\sum_{j=1}^{b} x_i y_j e_i \otimes f_j\,.$$

Prodotto scalare in $\mathbb{R}^a \otimes \mathbb{R}^b$
È anche possibile dotare lo spazio prodotto tensoriale di una struttura euclidea naturale, per mezzo del prodotto scalare

$$\langle x|y\rangle = \Big\langle \sum_{i=1}^{a}\sum_{j=1}^{b} x_{ij} e_i \otimes f_j \Big| \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij} e_i \otimes f_j \Big\rangle = \sum_{i=1}^{a}\sum_{j=1}^{b} x_{ij} y_{ij}$$

rispetto alla quale la base $\{e_i \otimes f_j\,,\ 1 \le i \le a\,, 1 \le j \le b\}$ risulta per definizione ortonormale

$$\langle e_i \otimes f_j | e_h \otimes f_k\rangle = \sum_{p=1}^{a}\sum_{q=1}^{b} (e_i \otimes f_j)_{pq}(e_h \otimes f_k)_{pq} = \sum_{p=1}^{a}\sum_{q=1}^{b} \delta_{ip}\delta_{jq}\delta_{hp}\delta_{kq} = \delta_{ih}\delta_{jk}\,.$$

Operatori lineari su $\mathbb{R}^a \otimes \mathbb{R}^b$
La definizione è la stessa che si introduce per un generico spazio vettoriale. Per operatore lineare su $\mathbb{R}^a \otimes \mathbb{R}^b$ si intende perciò una applicazione Q di $\mathbb{R}^a \otimes \mathbb{R}^b$ in sè

$$Q \;:\; \mathbb{R}^a \otimes \mathbb{R}^b \longrightarrow \mathbb{R}^a \otimes \mathbb{R}^b$$

per la quale risulta

$$Q(\alpha x + \beta y) \;=\; \alpha Q x + \beta Q y \qquad \forall\, \alpha, \beta \in \mathbb{R} \quad \forall\, x, y \in \mathbb{R}^a \otimes \mathbb{R}^b\,.$$

Rispetto alla base $\{e_i \otimes f_j,\, 1 \le i \le a,\, 1 \le j \le b\}$ l'azione dell'operatore viene descritta mediante una appropriata **matrice di rappresentazione** $[Q]$

$$Q(e_i \otimes f_j) \;=\; \sum_{h=1}^{a} \sum_{k=1}^{b} [Q]_{hk,ij} e_h \otimes f_k \qquad 1 \le i \le a \quad 1 \le j \le b$$

per cui risulta

$$Q(x) \;=\; Q\Big(\sum_{i=1}^{a}\sum_{j=1}^{b} x_{ij} e_i \otimes f_j\Big) \;=\; \sum_{i=1}^{a}\sum_{j=1}^{b} x_{ij} Q(e_i \otimes f_j) \;=\; \sum_{h=1}^{a}\sum_{k=1}^{b}\sum_{i=1}^{a}\sum_{j=1}^{b} [Q]_{hk,ij} x_{ij} e_h \otimes f_k$$

e conseguentemente

$$Q(x)_{hk} \;=\; \sum_{i=1}^{a}\sum_{j=1}^{b} [Q]_{hk,ij} x_{ij} \qquad 1 \le h \le a \quad 1 \le k \le b\,.$$

Combinazione lineare di operatori lineari su $\mathbb{R}^a \otimes \mathbb{R}^b$
Dati due operatori lineari A e B in $\mathbb{R}^a \otimes \mathbb{R}^b$, e due coefficienti reali α, β, la combinazione lineare $\alpha A + \beta B$ è l'operatore lineare su $\mathbb{R}^a \otimes \mathbb{R}^b$ definito da

$$(\alpha A + \beta B)x \;=\; \alpha A x + \beta B x \qquad \forall\, x \in \mathbb{R}^a \otimes \mathbb{R}^b\,.$$

Tale definizione coincide con quella data per gli operatori lineari su spazi vettoriali arbitrari. È immediato verificare che la matrice rappresentativa di $\alpha A + \beta B$ rispetto alla base degli $e_i \otimes f_j$ è data dalla corrispondente combinazione lineare delle matrici rappresentative di A e B

$$[\alpha A + \beta B]_{ij,hk} \;=\; \alpha [A]_{ij,hk} + \beta [B]_{ij,hk} \qquad \forall\, i, j, h, k\,.$$

Prodotto tensoriale di operatori lineari su $\mathbb{R}^a$ e $\mathbb{R}^b$
Dati due operatori lineari, A su $\mathbb{R}^a$ e B su $\mathbb{R}^b$

$$A \;:\; \mathbb{R}^a \longrightarrow \mathbb{R}^a \qquad\qquad B \;:\; \mathbb{R}^b \longrightarrow \mathbb{R}^b$$

descritti rispetto alle relative basi $\{e_1, \ldots, e_a\}$ e $\{f_1, \ldots, f_b\}$ dalle matrici rappresentative $[A]$ e $[B]$

$$Ae_i = \sum_{h=1}^{a} [A]_{hi} e_h \quad 1 \le i \le a \qquad Bf_j = \sum_{k=1}^{b} [B]_{kj} f_k \quad 1 \le j \le b\,,$$

si definisce prodotto tensoriale degli operatori l'operatore lineare $A \otimes B$ di $\mathbb{R}^a \otimes \mathbb{R}^b$ in sè la cui matrice rappresentativa rispetto alla base $\{e_i \otimes f_j,\, 1 \le i \le a,\, 1 \le j \le b\}$ si scrive

$$[A \otimes B]_{hk,ij} = [A]_{hi}[B]_{kj} \qquad 1 \le h,k \le a \quad 1 \le i,j \le b\,.$$

Introducendo per brevità la convenzione di somma sugli indici ripetuti, l'azione dell'operatore $A \otimes B$ su un generico vettore $x \in \mathbb{R}^a \otimes \mathbb{R}^b$ è descritta da

$$A \otimes Bx = x_{ij} A \otimes B(e_i \otimes f_j) = x_{ij} [A \otimes B]_{hk,ij} e_h \otimes f_k = x_{ij} [A]_{hk} [B]_{ij} e_h \otimes f_k$$

per cui

$$(A \otimes Bx)_{ij} = [A \otimes B]_{ij,hk} x_{hk} = A_{ih} B_{jk} x_{hk}\,.$$

Si osservi che se applicato ad un prodotto tensoriale di vettori x e y, il prodotto tensoriale ha l'azione seguente

$$\begin{aligned} A \otimes B(x \otimes y) &= [A \otimes B]_{hk,ij} x_i y_j e_h \otimes f_k = x_i y_j [A]_{hi} [B]_{kj} e_h \otimes f_k = \\ &= (Ax)_h (By)_k e_h \otimes f_k = (Ax) \otimes (By)\,. \end{aligned}$$

Aggiunto di un prodotto tensoriale di operatori lineari

L'aggiunto in $\mathbb{R}^a \otimes \mathbb{R}^b$ del prodotto tensoriale $A \otimes B$ degli operatori lineari A in $\mathbb{R}^a$ e B in $\mathbb{R}^b$ è dato dal prodotto tensoriale degli operatori aggiunti

$$(A \otimes B)^+ = A^+ \otimes B^+\,.$$

Basta infatti osservare che per ogni $x \in \mathbb{R}^a \otimes \mathbb{R}^b$ vale

$$\langle y | A \otimes Bx \rangle = y_{ij} (A \otimes Bx)_{ij} = y_{ij} [A \otimes B]_{ij,hk} x_{hk} = y_{ij} [A]_{ih} [B]_{jk} x_{hk}$$

e poiché le matrici rappresentative degli aggiunti di A e B nelle basi rispettive soddisfano le condizioni

$$\begin{aligned} \alpha_i [A]_{ij} \beta_j &= [A^+]_{ji} \alpha_i \beta_j \qquad \forall\, (\alpha_1, \ldots, \alpha_a)\,,\ (\beta_1, \ldots, \beta_a) \in \mathbb{R}^a \\ \xi_i [B]_{ij} \eta_j &= [B^+]_{ji} \xi_i \eta_j \qquad \forall\, (\xi_1, \ldots, \xi_b)\,,\ (\eta_1, \ldots, \eta_b) \in \mathbb{R}^b \end{aligned}$$

si conclude che

$$\begin{aligned} \langle y | A \otimes Bx \rangle &= [A^+]_{hi} y_{ij} [B]_{jk} x_{hk} = [A^+]_{hi} [B^+]_{kj} y_{ij} x_{hk} = \\ &= (A^+ \otimes B^+ y)_{hk} x_{hk} = \langle A^+ \otimes B^+ y | x \rangle \end{aligned}$$

per cui $(A \otimes B)^+ = A^+ \otimes B^+$, come affermato.

Composizione di operatori prodotto tensoriale
Si considerino in $\mathbb{R}^a \otimes \mathbb{R}^b$ i prodotti tensoriali

$$A \otimes B \qquad \text{e} \qquad C \otimes D$$

degli operatori lineari A, C in $\mathbb{R}^a$ e B, D in $\mathbb{R}^b$. Vale allora l'identità

$$(A \otimes B)\,(C \otimes D) \;=\; AC \,\otimes\, BD\,. \tag{26.1}$$

Rispetto alla base degli $e_i \otimes f_j$ risulta in effetti

$$\begin{aligned}
[(A \otimes B)\,(C \otimes D)]_{ij,hk} \;&=\; [A \otimes B]_{ij,pq}[C \otimes D]_{pq,hk} \;= \\
&=\; [A]_{ip}[B]_{jq}[C]_{ph}[D]_{qk} \;=\; [A]_{ip}[C]_{ph}[B]_{jq}[D]_{qk} \;= \\
&=\; [AC]_{ih}[BD]_{jk} \;=\; [AC \otimes BD]_{ij,hk}
\end{aligned}$$

e dall'arbitrarietà degli indici i, j, h, k segue l'asserto.

Proprietà (pseudo)distributiva del prodotto tensoriale rispetto alla somma di operatori lineari
Dati gli operatori lineari

$$A\,,\; B \;:\; \mathbb{R}^a \longrightarrow \mathbb{R}^a \qquad \text{e} \qquad C \;:\; \mathbb{R}^b \longrightarrow \mathbb{R}^b$$

vale la proprietà di tipo distributivo

$$(A + B) \otimes C \;=\; A \otimes C + B \otimes C \tag{26.2}$$

ove si convenga che nell'espressione a secondo membro i prodotti tensoriali abbiano la precedenza sulla somma. Rispetto alla base degli $e_i \otimes f_j$ si ha infatti, per un qualsiasi $x \in \mathbb{R}^a \otimes \mathbb{R}^b$,

$$\begin{aligned}
((A + B) \otimes Cx)_{ij} \;&=\; [(A + B) \otimes C]_{ij,hk}x_{hk} \;=\; [A + B]_{ih}[C]_{jk}x_{hk} \;= \\
&=\; ([A]_{ih} + [B]_{ih})[C]_{jk}x_{hk} \;=\; [A]_{ih}[C]_{jk}x_{hk} + [B]_{ih})[C]_{jk}x_{hk} \;= \\
&=\; [A \otimes C]_{ij,hk}x_{hk} + [B \otimes C]_{ij,hk}x_{hk} \;=\; [A \otimes C + B \otimes C]_{ij,hk}x_{hk} \;= \\
&=\; ((A \otimes C + B \otimes C)x)_{ij}
\end{aligned}$$

come richiesto.

27. Operatori di proiezione ortogonale utili per l'analisi della varianza

Sottospazio vettoriale N_n

Nello spazio vettoriale $\mathbb{R}^n$ si considera la base canonica $\{e_1, \ldots, e_n\}$ e si indicano con $x_1, \ldots, x_n$ le componenti di un generico vettore x rispetto a tale base, in modo che

$$x = (x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i e_i \,.$$

Lo spazio vettoriale si intende munito dell'usuale prodotto scalare

$$\langle x|y\rangle = \sum_{i=1}^{n} x_i y_i \qquad \forall\, x = \sum_{i=1}^{n} x_i e_i \,, \quad y = \sum_{i=1}^{n} y_i e_i \,.$$

Si introduce il sottospazio N_n dei vettori di $\mathbb{R}^n$ le cui componenti rispetto alla base canonica hanno somma nulla

$$N_n = \left\{ x = \sum_{i=1}^{n} x_i e_i \in \mathbb{R}^n \;:\; \sum_{i=1}^{n} x_i = 0 \right\}$$

ed il relativo complemento ortogonale in $\mathbb{R}^n$

$$N_n^{\perp} = \{ x \in \mathbb{R}^n \;:\; \langle y|x\rangle = 0 \quad \forall\, y \in N_n \} \,.$$

Struttura di $N_n^{\perp}$

Si verifica che

$$N_n^{\perp} = \{ x = (x_1, \ldots, x_n) \in \mathbb{R}^n \;:\; x_i = x_j \quad \forall\, i, j = 1, \ldots, n \}$$

cioè che il complemento ortogonale di N_n è il sottospazio dei vettori le cui componenti rispetto alla base canonica sono tutte eguali fra loro. Se infatti $x \in N_n^{\perp}$ si ha che

$$\langle y|x\rangle = \sum_{i=1}^{n} x_i y_i = 0 \qquad \forall\, y_1, \ldots, y_n \;:\; \sum_{i=1}^{n} y_i = 0$$

e siccome una scelta possibile dei vettori y è costituita dagli $n-1$ vettori (linearmente indipendenti)

$$\begin{aligned} y^{(1)} &= (1, 0, \ldots, 0, -1) \\ y^{(2)} &= (0, 1, \ldots, 0, -1) \\ &\cdots\cdots \\ y^{(n-1)} &= (0, 0, \ldots, 1, -1) \end{aligned}$$

ne deriva che le componenti di x devono soddisfare le relazioni

$$\begin{cases} x_1 - x_n = 0 \\ x_2 - x_n = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ x_{n-1} - x_n = 0 \end{cases}$$

per cui $x_i = x_n \; \forall\, i = 1, \ldots, n-1$ e di conseguenza

$$x_i \,=\, x_j \qquad \forall\, i,j = 1,\ldots,n\,.$$

Risulta in tal modo verificata l'inclusione

$$N_n^\perp \subseteq \{x \,=\, (x_1,\ldots,x_n) \in \mathbb{R}^n \;:\; x_i \,=\, x_j \quad \forall\, i,j=1,\ldots,n\}\,.$$

Viceversa, supposto che x abbia componenti tutte eguali ad un $a \in \mathbb{R}$, si avrà

$$\langle y|x\rangle \,=\, \sum_{i=1}^n y_i a \,=\, a\sum_{i=1}^n y_i \,=\, 0 \qquad \forall\, y \in N_n$$

per cui $x \in N_n^\perp$ e quindi

$$\{x \,=\, (x_1,\ldots,x_n) \in \mathbb{R}^n \;:\; x_i \,=\, x_j \quad \forall\, i,j=1,\ldots,n\} \subseteq N_n^\perp$$

col che l'affermazione è dimostrata.

Dimensione di N_n e $N_n^\perp$
La dimensione di $N_n^\perp$ segue immediatamente dalla definizione

$$\dim N_n^\perp \,=\, \dim\{x \,=\, (x_1,\ldots,x_n) \in \mathbb{R}^n \;:\; x_i \,=\, x_j \quad \forall\, i,j=1,\ldots,n\} \,=\, 1$$

Poiché inoltre le dimensioni di N_n e $N_n^\perp$ sono l'una il complemento a n dell'altra

$$\dim N_n + \dim N_n^\perp \,=\, n\,.$$

si deduce che

$$\dim N_n \,=\, n - \dim N_n^\perp \,=\, n-1\,.$$

Scomposizione di $\mathbb{R}^n$ in somma diretta di N_n e $N_n^\perp$
Lo spazio $\mathbb{R}^n$ può essere scritto come somma diretta dei sottospazi ortogonali N_n e $N_n^\perp$. Sebbene questa affermazione segua dai teoremi generali sul complemento ortogonale, conviene verificare esplicitamente il risultato in modo da ricavare la forma assunta dalle proiezioni su N_n e $N_n^\perp$ di un generico vettore $x \in \mathbb{R}^n$. Per ogni $x = (x_1,\ldots,x_n) \in \mathbb{R}^n$ si può scrivere infatti

$$x_i \,=\, x_i - \frac{1}{n}\sum_{j=1}^n x_j + \frac{1}{n}\sum_{j=1}^n x_j \qquad \forall\, i = 1,\ldots,n$$

con i vettori v e w, di componenti rispettive

$$v_i \,=\, x_i - \frac{1}{n}\sum_{j=1}^n x_j \qquad \forall\, i=1,\ldots,n$$

$$w_i \,=\, \frac{1}{n}\sum_{j=1}^n x_j\,,$$

che appartengono l'uno allo spazio N_n e l'altro al complemento ortogonale $N_n^\perp$. Vale in effetti

$$\sum_{i=1}^n v_i \,=\, \sum_{i=1}^n \Big(x_i - \frac{1}{n}\sum_{j=1}^n x_j \Big) \,=\, \sum_{i=1}^n x_i - \sum_{j=1}^n x_j \,=\, 0$$

mentre la componente $w_i = \sum_{j=1}^n x_j/n$ è indipendente dall'indice i, per cui

$$\mathbb{R}^n \,=\, N_n + N_n^\perp \,.$$

Dal momento che inoltre

$$x \in N_n \cap N_n^\perp \quad \Longrightarrow \quad \langle x|x\rangle \,=\, 0 \quad \Longleftrightarrow \quad x \,=\, 0$$

per cui

$$N_n \cap N_n^\perp \,=\, \{0\}\,,$$

si conclude che

$$\mathbb{R}^n \,=\, N_n \oplus N_n^\perp \,.$$

Operatore di proiezione ortogonale su N_n
La scomposizione in somma diretta $\mathbb{R}^n = N_n \oplus N_n^\perp$ implica che $\forall\, x \in \mathbb{R}^n$ esistono un unico vettore $v \in N_n$ ed un unico vettore $w \in N_n^\perp$ per i quali risulta $x = v + w$. È così definita l'applicazione

$$P \;:\; x \in \mathbb{R}^n \;\longrightarrow\; v \in N_n \,.$$

Linearità dell'operatore di proiezione P
La linearità di P segue immediatamente dalla definizione di somma diretta. Se infatti

$$\begin{aligned} x \,&=\, v + w \quad v \in N_n \,, \quad w \in N_n^\perp \\ x' \,&=\, v' + w' \quad v' \in N_n \,, \quad w' \in N_n^\perp \end{aligned}$$

allora per definizione $v = Px$ e $v' = Px'$. Inoltre, $\forall \lambda, \lambda' \in \mathbb{R}$ si ha

$$\begin{aligned} \lambda x + \lambda' x' \,&=\, \lambda(v+w) + \lambda'(v'+w') \,= \\ &=\, \lambda v + \lambda' v' \,+\, \lambda w + \lambda' w' \end{aligned}$$

ma poiché $\lambda v + \lambda' v' \in N_n$ e $\lambda w + \lambda' w' \in N_n^\perp$, essendo N_n e $N_n^\perp$ sottospazi vettoriali di $\mathbb{R}^n$, si conclude che

$$P(\lambda x + \lambda' x') \,=\, \lambda v + \lambda' v' \,=\, \lambda P x + \lambda' P x' \,.$$

Idempotenza dell'operatore P
Per $x \in \mathbb{R}^n$ arbitrario vale

$$Px \,=\, Px + 0$$

con $Px \in N_n$ e $0 \in N_n^\perp$. Dalla definizione di P segue pertanto che

$$P^2 x = P(Px) = Px \qquad \forall\, x \in \mathbb{R}^n$$

e che quindi l'operatore è idempotente

$$P^2 = P\,.$$

Simmetria dell'operatore P
Per ogni $x, y \in \mathbb{R}^n$ risulta

$$\langle x/Py\rangle = \langle Px + x - Px|Py\rangle = \langle Px|Py\rangle + \langle x - Px|y\rangle$$

e siccome $x - Px \in N_n^\perp$, $Py \in N_n$, il prodotto scalare $\langle x - Px|y\rangle$ si annulla e l'espressione si semplifica in

$$\langle x/Py\rangle = \langle Px|Py\rangle\,.$$

L'ortogonalità dei vettori $Px \in N_n$ e $y - Py \in N_n^\perp$ consente poi di scrivere

$$\langle x|Py\rangle = \langle Px|Py\rangle + \langle Px|y - Py\rangle = \langle Px|y - Py + Py\rangle = \langle Px|y\rangle$$

e di provare così il risultato.

Operatore di proiezione ortogonale su $N_n^\perp$
È evidente che l'operatore $\mathbb{I} - P$ definito per mezzo della relazione

$$(\mathbb{I} - P)x = x - Px \qquad \forall\, x \in \mathbb{R}^n$$

costituisce l'operatore di proiezione ortogonale sul sottospazio $N_n^\perp$. Ciò in virtù della identità

$$x = Px + x - Px = Px + \mathbb{I}x - Px = Px + (\mathbb{I} - P)x$$

nella quale $Px \in N_n$ e $(\mathbb{I} - P)x \in N_n^\perp$.

Rango degli operatori P e $\mathbb{I} - P$
Il rango dell'operatore P concide con la dimensione del suo range

$$\text{rank}(P) = \text{range}(P) = \text{dim}N_n = n - 1$$

e lo stesso dicasi per l'operatore $\mathbb{I} - P$

$$\text{rank}(\mathbb{I} - P) = \text{dim}N_n^\perp = 1\,.$$

Nullità degli operatori P e $\mathbb{I} - P$
Le nullità degli operatori vengono dedotte dalle relazioni

$$\dim\ker(P) + \dim\text{range}(P) = \dim(\mathbb{R}^n)$$

$$\dim\ker(\mathbb{I}-P)+\dim\mathrm{range}(\mathbb{I}-P)\;=\;\dim(\mathbb{R}^n)$$

ovvero dal fatto che range e kernel dei due operatori si corrispondono scambievolmente

$$\ker(\mathbb{I}-P)\;=\;\mathrm{range}(P)$$

$$\mathrm{range}(\mathbb{I}-P)\;=\;\ker(P)\,.$$

In effetti

$$x\in\ker(\mathbb{I}-P)\quad\Longrightarrow\quad(\mathbb{I}-P)x\;=\;0\quad\Longrightarrow\quad x\;=\;Px\quad\Longrightarrow\quad x\in\mathrm{range}(P)$$

e, viceversa,

$$\begin{aligned}&x\in\mathrm{range}(P)\quad\Longrightarrow\quad x\;=\;Py\,,\;y\in\mathbb{R}^n\quad\Longrightarrow\\&(\mathbb{I}-P)x\;=\;(\mathbb{I}-P)Py\;=\;Py-P^2y\;=\;Py-Py\;=\;0\quad\Longrightarrow\quad x\in\ker(\mathbb{I}-P)\end{aligned}$$

per cui $\ker(\mathbb{I}-P)=\mathrm{range}(P)$. In modo analogo

$$\begin{aligned}&x\in\mathrm{range}(\mathbb{I}-P)\quad\Longrightarrow\quad x\;=\;(\mathbb{I}-P)y\,,\;y\in\mathbb{R}^n\quad\Longrightarrow\\&Px\;=\;P(\mathbb{I}-P)y\;=\;Py-P^2y\;=\;Py-Py\;=\;0\quad\Longrightarrow\quad x\in\ker(P)\end{aligned}$$

e

$$\begin{aligned}&x\in\ker(P)\quad\Longrightarrow\quad Px\;=\;0\quad\Longrightarrow\\&x\;=\;x-0\;=\;x-Px\;=\;(\mathbb{I}-P)x\quad\Longrightarrow\quad x\in\mathrm{range}(\mathbb{I}-P)\end{aligned}$$

con la conseguente identità $\mathrm{range}(\mathbb{I}-P)=\ker(P)$. In ogni caso le nullità valgono

$$\dim\ker(P)\;=\;1\qquad\text{e}\qquad\dim\ker(\mathbb{I}-P)\;=\;n-1\,.$$

Prodotti tensoriali di operatori P e $\mathbb{I}-P$
Si considerino in $\mathbb{R}^a$ e $\mathbb{R}^b$, rispettivamente, gli operatori lineari di proiezione ortogonale su $N_a\subset\mathbb{R}^a$ e $N_b\subset\mathbb{R}^b$

$$P_1\;:\;\mathbb{R}^a\longrightarrow N_a\qquad\qquad P_2\;:\;\mathbb{R}^b\longrightarrow N_b$$

secondo la precedente definizione. Si può allora scrivere la seguente scomposizione

$$\mathbb{I}\;=\;\mathbb{I}\otimes\mathbb{I}\;=\;(\mathbb{I}-P_1)\otimes(\mathbb{I}-P_2)+P_1\otimes(\mathbb{I}-P_2)+(\mathbb{I}-P_1)\otimes P_2+P_1\otimes P_2 \tag{27.1}$$

con gli operatori lineari

$$\begin{aligned}(\mathbb{I}-P_1)\otimes(\mathbb{I}-P_2)\;&:\;\mathbb{R}^a\otimes\mathbb{R}^b\longrightarrow N_a^\perp\otimes N_b^\perp\\P_1\otimes(\mathbb{I}-P_2)\;&:\;\mathbb{R}^a\otimes\mathbb{R}^b\longrightarrow N_a\otimes N_b^\perp\\(\mathbb{I}-P_1)\otimes P_2\;&:\;\mathbb{R}^a\otimes\mathbb{R}^b\longrightarrow N_a^\perp\otimes N_b\\P_1\otimes P_2\;&:\;\mathbb{R}^a\otimes\mathbb{R}^b\longrightarrow N_a\otimes N_b\end{aligned} \tag{27.2}$$

simmetrici, idempotenti e di ranghi rispettivi

$$\begin{aligned}
\dim(N_a^{\perp} \otimes N_b^{\perp}) &= \dim N_a^{\perp} \cdot \dim N_b^{\perp} = (a-1)(b-1) \\
\dim(N_a \otimes N_b^{\perp}) &= \dim N_a \cdot \dim N_b^{\perp} = 1\,(b-1) = b-1 \\
\dim(N_a^{\perp} \otimes N_b) &= \dim N_a^{\perp} \cdot \dim N_b = (a-1)\,1 = a-1 \\
\dim(N_a \otimes N_b) &= \dim N_a \cdot \dim N_b = 1\,;
\end{aligned}$$

inoltre, la composizione di due qualsiasi degli operatori predetti è nulla.
La relazione (27.1) è una immediata conseguenza della proprietà distributiva (26.2)

$$\begin{aligned}
\mathbb{I} = \mathbb{I} \otimes \mathbb{I} &= (\mathbb{I} - P_1 + P_1) \otimes (\mathbb{I} - P_2 + P_2) = \\
&= (\mathbb{I} - P_1 + P_1) \otimes (\mathbb{I} - P_2) + (\mathbb{I} - P_1 + P_1) \otimes P_2 = \\
&= (\mathbb{I} - P_1) \otimes (\mathbb{I} - P_2) + P_1 \otimes (\mathbb{I} - P_2) + (\mathbb{I} - P_1) \otimes P_2 + P_1 \otimes P_2
\end{aligned}$$

mentre i ranghi degli operatori coincidono con le dimensioni dei rispettivi codomini. La simmetria degli operatori (27.2) si deduce dall'espressione dell'aggiunto di un prodotto tensoriale e dalla simmetria di P_1, P_2, per esempio

$$((\mathbb{I} - P_1) \otimes (\mathbb{I} - P_2))^+ = (\mathbb{I} - P_1)^+ \otimes (\mathbb{I} - P_2)^+ = (\mathbb{I} - P_1) \otimes (\mathbb{I} - P_2)$$

(la dimostrazione è analoga negli altri tre casi). L'idempotenza si verifica usando la formula (26.1) di composizione fra prodotti tensoriali di operatori lineari e l'idempotenza di P_1, P_2

$$((\mathbb{I} - P_1) \otimes (\mathbb{I} - P_2))^2 = (\mathbb{I} - P_1)^2 \otimes (\mathbb{I} - P_2)^2 = (\mathbb{I} - P_1) \otimes (\mathbb{I} - P_2)\,;$$

la stessa formula consente di provare l'annullarsi di tutte le composizioni fra gli operatori (27.2), ad esempio

$$((\mathbb{I} - P_1) \otimes (\mathbb{I} - P_2))\,(P_1 \otimes (\mathbb{I} - P_2)) = ((\mathbb{I} - P_1)P_1) \otimes (\mathbb{I} - P_2)^2 = 0 \otimes (\mathbb{I} - P_2) = 0\,.$$

28. Analisi della varianza (ANOVA)

L'analisi di varianza (ANalysis Of VAriance, ANOVA) si applica a set di dati organizzati secondo lo schema seguente

$$x_{ijk} \qquad 1 \le i \le a\,, \quad 1 \le j \le b\,, \quad 1 \le k \le c\,.$$

Le variabili

$$\{x_{ijk}\,,\ 1 \le k \le c\} \tag{28.1}$$

rappresentano i risultati di c misure ripetute di una grandezza x sotto particolari condizioni quantitative o qualitative specificate dai valori assegnati degli indici i e j. Tali condizioni quantitative o qualitative sono note come **fattori**: l'indice i si riferisce al primo fattore, suscettibile di assumere a diversi livelli, mentre l'indice j è relativo al secondo fattore, per il quale ricorrono b livelli differenti. A titolo di esempio, si può considerare il caso di un materiale di cui si determina una particolare proprietà fisica, quale potrebbe essere l'energia libera superficiale. Campioni identici del materiale vengono sottoposti ad un trattamento consistente nell'attacco chimico

(i) mediante un reagente scelto in un set di a reagenti prestabiliti,

(ii) eseguito ad una data temperatura, scelta fra b temperature di riferimento predefinite.

L'energia libera superficiale del materiale è la grandezza x. Si individua il tipo di reagente utilizzato mediante un particolare valore dell'indice $i = 1, \ldots, a$. Il reagente costituisce il **primo fattore**. La temperatura del trattamento viene specificata assegnando il valore del secondo indice $j = 1, \ldots, b$; tale temperatura di trattamento rappresenta il **secondo fattore** del sistema di dati. A reagente e temperatura di trattamento assegnati (i e j fissati) si eseguono c misure ripetute dell'energia libera superficiale del campione, ciascuna individuata da un valore del terzo indice $k = 1, \ldots, c$. I risultati vengono rappresentati nella forma (28.1). Non è necessario assumere, in generale, che il numero di prove ripetute sia sempre lo stesso — c — per ogni scelta dei livelli i e j. L'ipotesi viene introdotta soltanto per semplicità di notazione.

Si può ritenere che l'effetto del trattamento sia quello di modificare il valore dell'energia libera superficiale del materiale, in modo che i valori medi delle variabili (28.1) vengano a dipendere dagli indici i e j che specificano il trattamento. Le variazioni delle variabili all'interno di ciascuno dei set (28.1) si assumono puramente casuali, non potendo essere interpretate o "spiegate"come dovute ai diversi trattamenti.

In definitiva, un modello statistico semplice del sistema di dati è il seguente

$$x_{ijk} = \mu_{ij} + \varepsilon_{ijk} \qquad i = 1, \ldots, a\,, \quad j = 1, \ldots, b\,, \quad k = 1, \ldots, c$$

dove:

- le ε_{ijk} si assumono variabili casuali gaussiane indipendenti identicamente distribuite, di media nulla e varianza σ^2;
- ciascuno dei coefficienti μ_{ij} si identifica con il comune valore medio di tutte le variabili x_{ijk}, $k = 1, \ldots, c$.

In conseguenza del modello assunto, è evidente che tutti i dati x_{ijk} sono intesi come varibili casuali gaussiane indipendenti di eguale varianza σ^2 e media μ_{ij}, quest'ultima dipendente — almeno in linea di principio — dal particolare trattamento (i, j).

Dal punto di vista formale conviene introdurre i vettori $x, \mu \in \mathbb{R}^a \otimes \mathbb{R}^b \otimes \mathbb{R}^c$ di componenti ovvie

$$\mu_{ijk} = \mu_{ij} \qquad \text{e} \qquad x_{ijk} \qquad \forall\, i, j, k$$

unitamente agli operatori di proiezione ortogonale già definiti al punto precedente:

$$\begin{aligned} P_1 &: \mathbb{R}^a \longrightarrow N_a \subset \mathbb{R}^a \\ P_2 &: \mathbb{R}^b \longrightarrow N_b \subset \mathbb{R}^b \\ P_3 &: \mathbb{R}^c \longrightarrow N_c \subset \mathbb{R}^c\,. \end{aligned}$$

Scarti, devianze e rispettive scompozioni

Dall'identità

$$\mathbb{I}\otimes\mathbb{I} = (\mathbb{I}-P_1)\otimes(\mathbb{I}-P_2) + P_1\otimes(\mathbb{I}-P_2) + (\mathbb{I}-P_1)\otimes P_2 + P_1\otimes P_2 \tag{28.2}$$

e dalla proprietà distributiva del prodotto tensoriale di operatori rispetto alla somma segue che

$$\begin{aligned}\mathbb{I}\otimes\mathbb{I}\otimes\mathbb{I} &= \mathbb{I}\otimes\mathbb{I}\otimes(\mathbb{I}-P_3) + \mathbb{I}\otimes\mathbb{I}\otimes P_3 = \\ &= [(\mathbb{I}-P_1)\otimes(\mathbb{I}-P_2) + P_1\otimes(\mathbb{I}-P_2) + (\mathbb{I}-P_1)\otimes P_2 + P_1\otimes P_2]\otimes(\mathbb{I}-P_3)+ \\ &+ \mathbb{I}\otimes\mathbb{I}\otimes P_3\end{aligned}$$

per cui

$$\begin{aligned}x &= (\mathbb{I}-P_1)\otimes(\mathbb{I}-P_2)\otimes(\mathbb{I}-P_3)x+ \\ &+ P_1\otimes(\mathbb{I}-P_2)\otimes(\mathbb{I}-P_3)x+ \\ &+ (\mathbb{I}-P_1)\otimes P_2\otimes(\mathbb{I}-P_3)x+ \\ &+ P_1\otimes P_2\otimes(\mathbb{I}-P_3)x+ \\ &+ \mathbb{I}\otimes\mathbb{I}\otimes P_3x\end{aligned} \tag{28.3}$$

e conseguentemente

$$\begin{aligned}x-(\mathbb{I}-P_1)\otimes(\mathbb{I}-P_2)\otimes(\mathbb{I}-P_3)x &= P_1\otimes(\mathbb{I}-P_2)\otimes(\mathbb{I}-P_3)x+ \\ &+ (\mathbb{I}-P_1)\otimes P_2\otimes(\mathbb{I}-P_3)x+ \\ &+ P_1\otimes P_2\otimes(\mathbb{I}-P_3)x+ \\ &+ \mathbb{I}\otimes\mathbb{I}\otimes P_3x\ .\end{aligned} \tag{28.4}$$

Calcolando il modulo quadrato delle espressioni a primo e a secondo membro e usando le proprietà degli operatori di proiezione si perviene infine all'identità

$$\begin{aligned}\big|x-(\mathbb{I}-P_1)\otimes(\mathbb{I}-P_2)\otimes(\mathbb{I}-P_3)x\big|^2 &= \big|P_1\otimes(\mathbb{I}-P_2)\otimes(\mathbb{I}-P_3)x\big|^2+ \\ &+ \big|(\mathbb{I}-P_1)\otimes P_2\otimes(\mathbb{I}-P_3)x\big|^2+ \\ &+ \big|P_1\otimes P_2\otimes(\mathbb{I}-P_3)x\big|^2+ \\ &+ \big|\mathbb{I}\otimes\mathbb{I}\otimes P_3x\big|^2\ .\end{aligned} \tag{28.5}$$

Con la definizione delle medie totale e parziali

$$x_{\bullet\bullet\bullet} \stackrel{\text{def}}{=} \frac{1}{abc}\sum_i\sum_j\sum_k x_{ijk}$$

$$x_{i\bullet\bullet} \stackrel{\text{def}}{=} \frac{1}{bc}\sum_j\sum_k x_{ijk}$$

$$x_{\bullet j \bullet} \stackrel{\text{def}}{=} \frac{1}{ac} \sum_i \sum_k x_{ijk}$$

$$x_{ij\bullet} \stackrel{\text{def}}{=} \frac{1}{c} \sum_k x_{ijk}$$

si hanno le relazioni

$$\begin{aligned}
&[(\mathbb{I}-P_1)\otimes(\mathbb{I}-P_2)\otimes(\mathbb{I}-P_3)x]_{ijk} \;=\; x_{\bullet\bullet\bullet} \\
&[P_1\otimes(\mathbb{I}-P_2)\otimes(\mathbb{I}-P_3)x]_{ijk} \;=\; x_{i\bullet\bullet}-x_{\bullet\bullet\bullet} \\
&[(\mathbb{I}-P_1)\otimes P_2\otimes(\mathbb{I}-P_3)x]_{ijk} \;=\; x_{\bullet j\bullet}-x_{\bullet\bullet\bullet} \\
&[P_1\otimes P_2\otimes(\mathbb{I}-P_3)x]_{ijk} \;=\; x_{ij\bullet}-x_{i\bullet\bullet}-x_{\bullet j\bullet}+x_{\bullet\bullet\bullet} \\
&[\mathbb{I}\otimes\mathbb{I}\otimes P_3x]_{ijk} \;=\; x_{ijk}-x_{ij\bullet}
\end{aligned} \tag{28.6}$$

in modo che le equazioni (28.3), (28.4) e (28.5) diventano

$$\begin{aligned}
x \;=\;& x_{\bullet\bullet\bullet}+ \\
&+x_{i\bullet\bullet}-x_{\bullet\bullet\bullet}+ \\
&+x_{\bullet j\bullet}-x_{\bullet\bullet\bullet}+ \\
&+x_{ij\bullet}-x_{i\bullet\bullet}-x_{\bullet j\bullet}+x_{\bullet\bullet\bullet}+ \\
&+x_{ijk}-x_{ij\bullet}
\end{aligned} \tag{28.7}$$

$$\begin{aligned}
x-x_{\bullet\bullet\bullet} \;=\;& x_{i\bullet\bullet}-x_{\bullet\bullet\bullet}+ \\
&+x_{\bullet j\bullet}-x_{\bullet\bullet\bullet}+ \\
&+x_{ij\bullet}-x_{i\bullet\bullet}-x_{\bullet j\bullet}+x_{\bullet\bullet\bullet}+ \\
&+x_{ijk}-x_{ij\bullet}
\end{aligned} \tag{28.8}$$

$$\begin{aligned}
\sum_{ijk}(x_{ijk}-x_{\bullet\bullet\bullet})^2 \;=\;& \sum_{ijk}(x_{i\bullet\bullet}-x_{\bullet\bullet\bullet})^2+ \\
&+\sum_{ijk}(x_{\bullet j\bullet}-x_{\bullet\bullet\bullet})^2+ \\
&+\sum_{ijk}(x_{ij\bullet}-x_{i\bullet\bullet}-x_{\bullet j\bullet}+x_{\bullet\bullet\bullet})^2+ \\
&+\sum_{ijk}(x_{ijk}-x_{ij\bullet})^2\,.
\end{aligned} \tag{28.9}$$

L'equazione (28.8) permette di esprimere lo scarto totale $x_{ijk}-x_{\bullet\bullet\bullet}$ come somma degli scarti:

$x_{i\bullet\bullet}-x_{\bullet\bullet\bullet}$	"spiegato" dal primo fattore (relativo all'indice i)
$x_{\bullet j\bullet}-x_{\bullet\bullet\bullet}$	"spiegato" dal secondo fattore (relativo all'indice j)
$x_{ij\bullet}-x_{i\bullet\bullet}-x_{\bullet j\bullet}+x_{\bullet\bullet\bullet}$	"spiegato" dall'interazione fra i due fattori
$x_{ijk}-x_{ij\bullet}$	"non spiegato"

che ne forniscono la scomposizione. In forza della (28.9), un risultato analogo vale per la devianza SS, o somma dei quadrati degli scarti, esprimibile come somma delle devianze

$$\begin{aligned}
SS_1 &= \sum_{ijk}(x_{i\bullet\bullet} - x_{\bullet\bullet\bullet})^2 && \text{“spiegata” dal primo fattore} \\
SS_2 &= \sum_{ijk}(x_{\bullet j\bullet} - x_{\bullet\bullet\bullet})^2 && \text{“spiegata” dal secondo fattore} \\
SS_{12} &= \sum_{ijk}(x_{ij\bullet} - x_{i\bullet\bullet} - x_{\bullet j\bullet} + x_{\bullet\bullet\bullet})^2 && \text{“spiegata” dall'interazione fra i due fattori} \\
SS_u &= \sum_{ijk}(x_{ijk} - x_{ij\bullet})^2 && \text{“non spiegata”}
\end{aligned}$$

per cui si ha

$$SS = SS_1 + SS_2 + SS_{12} + SS_u\,.$$

Parametri del modello

L'identità (28.2) consente di riesprimere il vettore dei valori medi μ nella forma

$$\mu = \mathbb{I}\otimes\mathbb{I}\mu = (\mathbb{I}-P_1)\otimes(\mathbb{I}-P_2)\mu + P_1\otimes(\mathbb{I}-P_2)\mu + (\mathbb{I}-P_1)\otimes P_2\mu + P_1\otimes P_2\mu$$

che per componenti si riduce a:

$$\mu_{ij} = \mu_{\bullet\bullet} + (\mu_{i\bullet} - \mu_{\bullet\bullet}) + (\mu_{\bullet j} - \mu_{\bullet\bullet}) + (\mu_{ij} - \mu_{i\bullet} - \mu_{\bullet j} + \mu_{\bullet\bullet})\,.$$

Per semplicità conviene introdurre la seguente notazione per le medie totale e parziali delle componenti di μ:

$$\begin{aligned}
\mu_{\bullet\bullet} &\stackrel{\text{def}}{=} \frac{1}{a}\sum_{i=1}^{a}\frac{1}{b}\sum_{j=1}^{b}\mu_{ij} = \frac{1}{ab}\sum_{ij}\mu_{ij} \\
\mu_{i\bullet} &\stackrel{\text{def}}{=} \frac{1}{b}\sum_{j=1}^{b}\mu_{ij} = \frac{1}{b}\sum_{j}\mu_{ij} \\
\mu_{\bullet j} &\stackrel{\text{def}}{=} \frac{1}{a}\sum_{i=1}^{a}\mu_{ij} = \frac{1}{a}\sum_{i}\mu_{ij}
\end{aligned}$$

e definire $\forall\, i = 1,\ldots,a,\ j = 1,\ldots,b$ i simboli:

$$\begin{aligned}
\mu_0 &\stackrel{\text{def}}{=} [(\mathbb{I}-P_1)\otimes(\mathbb{I}-P_2)\mu]_{ij} = \mu_{\bullet\bullet} \\
\alpha_i &\stackrel{\text{def}}{=} [P_1\otimes(\mathbb{I}-P_2)\mu]_{ij} = \mu_{i\bullet} - \mu_{\bullet\bullet} \\
\beta_j &\stackrel{\text{def}}{=} [(\mathbb{I}-P_1)\otimes P_2\mu]_{ij} = \mu_{\bullet j} - \mu_{\bullet\bullet} \\
(\alpha\beta)_{ij} &\stackrel{\text{def}}{=} [P_1\otimes P_2\mu]_{ij} = \mu_{ij} - \mu_{i\bullet} - \mu_{\bullet j} + \mu_{\bullet\bullet}
\end{aligned} \tag{28.10}$$

in modo che risulta

$$\mu_{ij} \;=\; \mu_0 + \alpha_i + \beta_j + (\alpha\beta)_{ij} \qquad \forall\, i = 1, \ldots, a\,,\; \forall\, j = 1, \ldots, b\,.$$

Il coefficiente μ_0 individua una stima complessiva di μ_{ij} che non tiene conto nel dettaglio dei fattori i e j; i parametri α_i descrivono l'effetto del primo fattore (associato all'indice i) sul valore medio μ_{ij}, indipendentemente dal secondo fattore; analogamente, i coefficienti β_j specificano l'effetto del secondo fattore (relativo all'indice j) sul valore medio μ_{ij} a prescindere dal primo fattore; gli $(\alpha\beta)_{ij}$ rappresentano infine dei termini di interazione fra i due fattori (responsabili del fatto che le differenze del tipo $\mu_{ij} - \mu_{ik}$ dipendono dall'indice fissato i, o che, analogamente, $\mu_{ij} - \mu_{hj}$ dipende da j).

Il modello statistico del sistema diventà pertanto

$$x_{ijk} \;=\; \mu_0 + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

con

$$\begin{aligned} &\sum_i \alpha_i \;=\; 0 \qquad\quad \sum_j \beta_j \;=\; 0 \\ &\sum_i (\alpha\beta)_{ij} \;=\; 0 \quad \forall\, j \qquad \sum_j (\alpha\beta)_{ij} \;=\; 0 \quad \forall\, i \end{aligned} \tag{28.11}$$

e le ε_{ijk} variabili casuali indipendenti gaussiane di media nulla e varianza σ^2.

Stima di best-fit dei parametri

La stima campionaria dei parametri α_i, β_j, $(\alpha\beta)_{ij}$ viene eseguita usando il metodo dei minimi quadrati. Si tratta di esprimere in termini di tali parametri la funzione di merito

$$\mathbb{S}(\mu) \;=\; \sum_{ijk} (\mu_{ij} - x_{ijk})^2 \;=\; \langle \mu - x | \mu - x \rangle$$

vale a dire la somma dei quadrati degli scarti delle variabili casuali indipendenti x_{ijk} rispetto ai relativi valori medi μ_{ij}. A questo scopo conviene osservare preliminarmente che siccome $(\mathbb{I} - P_3)\mu \;=\; \mu$ e $P_3\mu = 0$, deve aversi

$$\begin{aligned} \mu \;&=\; \mathbb{I} \otimes \mathbb{I} \otimes (\mathbb{I} - P_3)\mu \;+\; \mathbb{I} \otimes \mathbb{I} \otimes P_3\mu \;= \\ &=\; (\mathbb{I} - P_1) \otimes (\mathbb{I} - P_2) \otimes (\mathbb{I} - P_3)\mu + P_1 \otimes (\mathbb{I} - P_2) \otimes (\mathbb{I} - P_3)\mu + \\ &+\; (\mathbb{I} - P_1) \otimes P_2 \otimes (\mathbb{I} - P_3)\mu + P_1 \otimes P_2 \otimes (\mathbb{I} - P_3)\mu + \mathbb{I} \otimes \mathbb{I} \otimes P_3\mu \end{aligned}$$

per cui, sottraendo membro a membro dalla (28.3)

$$\begin{aligned} x - \mu \;=\; &(\mathbb{I} - P_1) \otimes (\mathbb{I} - P_2) \otimes (\mathbb{I} - P_3)(x - \mu) + \\ &+ P_1 \otimes (\mathbb{I} - P_2) \otimes (\mathbb{I} - P_3)(x - \mu) + \\ &+ (\mathbb{I} - P_1) \otimes P_2 \otimes (\mathbb{I} - P_3)(x - \mu) + \\ &+ P_1 \otimes P_2 \otimes (\mathbb{I} - P_3)(x - \mu) + \\ &+ \mathbb{I} \otimes \mathbb{I} \otimes P_3(x - \mu) \end{aligned} \tag{28.12}$$

e per il carattere autoaggiunto e idempotente degli operatori di proiezione

$$
\begin{aligned}
\langle x-\mu|x-\mu\rangle \;=\;& \langle x-\mu|(\mathbb{I}-P_1)\otimes(\mathbb{I}-P_2)\otimes(\mathbb{I}-P_3)(x-\mu)\rangle+ \\
&+\langle x-\mu|P_1\otimes(\mathbb{I}-P_2)\otimes(\mathbb{I}-P_3)(x-\mu)\rangle+ \\
&+\langle x-\mu|(\mathbb{I}-P_1)\otimes P_2\otimes(\mathbb{I}-P_3)(x-\mu)\rangle+ \\
&+\langle x-\mu|P_1\otimes P_2\otimes(\mathbb{I}-P_3)(x-\mu)\rangle+ \\
&+\langle x-\mu|\mathbb{I}\otimes\mathbb{I}\otimes P_3(x-\mu)\rangle
\end{aligned}
$$

sicché

$$
\begin{aligned}
|x-\mu|^2 \;=\;& \left|(\mathbb{I}-P_1)\otimes(\mathbb{I}-P_2)\otimes(\mathbb{I}-P_3)(x-\mu)\right|^2+ \\
&+\left|P_1\otimes(\mathbb{I}-P_2)\otimes(\mathbb{I}-P_3)(x-\mu)\right|^2+ \\
&+\left|(\mathbb{I}-P_1)\otimes P_2\otimes(\mathbb{I}-P_3)(x-\mu)\right|^2+ \\
&+\left|P_1\otimes P_2\otimes(\mathbb{I}-P_3)(x-\mu)\right|^2+ \\
&+\left|\mathbb{I}\otimes\mathbb{I}\otimes P_3(x-\mu)\right|^2 .
\end{aligned}
$$

Basta allora inserire le definizioni (28.6) e (28.10) per ottenere la funzione di merito richiesta

$$
\begin{aligned}
\mathbb{S}(\mu_0,\alpha,\beta,\alpha\beta) \;=\; \sum_{ijk}(x_{ijk}-\mu_{ij})^2 \;=\;& \sum_{ijk}(x_{\bullet\bullet\bullet}-\mu_0)^2+ \\
&+\sum_{ijk}(x_{i\bullet\bullet}-x_{\bullet\bullet\bullet}-\alpha_i)^2+ \\
&+\sum_{ijk}(x_{\bullet j\bullet}-x_{\bullet\bullet\bullet}-\beta_j)^2+ \\
&+\sum_{ijk}[x_{ij\bullet}-x_{i\bullet\bullet}-x_{\bullet j\bullet}+x_{\bullet\bullet\bullet}-(\alpha\beta)_{ij}]^2+ \\
&+\sum_{ijk}(x_{ijk}-x_{ij\bullet})^2
\end{aligned}
\tag{28.13}
$$

il cui mimino deve essere determinato tenendo conto delle condizioni (28.11). È immediato verificare che detto minimo si ottiene per la seguente scelta dei parametri:

$$
\begin{aligned}
\mu_0 &= x_{\bullet\bullet\bullet} \\
\alpha_i &= x_{i\bullet\bullet}-x_{\bullet\bullet\bullet} \qquad \forall\, i \\
\beta_j &= x_{\bullet j\bullet}-x_{\bullet\bullet\bullet} \qquad \forall\, j \\
(\alpha\beta)_{ij} &= x_{ij\bullet}-x_{i\bullet\bullet}-x_{\bullet j\bullet}+x_{\bullet\bullet\bullet} \qquad \forall\, i,j\,.
\end{aligned}
$$

È sufficiente osservare che i valori precedenti rendono minima la funzione obiettivo (28.13), annullandone i primi quattro termini a secondo membro, e che nel contempo soddisfano i vincoli (28.11):

$$
\begin{aligned}
\sum_i \alpha_i &= \sum_i (x_{i\bullet\bullet} - x_{\bullet\bullet\bullet}) = a x_{\bullet\bullet\bullet} - a x_{\bullet\bullet\bullet} = 0 \\
\sum_j \beta_j &= \sum_j (x_{\bullet j\bullet} - x_{\bullet\bullet\bullet}) = b x_{\bullet\bullet\bullet} - b x_{\bullet\bullet\bullet} = 0 \\
\sum_i (\alpha\beta)_{ij} &= \sum_i (x_{ij\bullet} - x_{i\bullet\bullet} - x_{\bullet j\bullet} + x_{\bullet\bullet\bullet}) = a x_{\bullet j\bullet} - a x_{\bullet\bullet\bullet} - a x_{\bullet j\bullet} + a x_{\bullet\bullet\bullet} = 0 \quad \forall\, j \\
\sum_j (\alpha\beta)_{ij} &= \sum_j (x_{ij\bullet} - x_{i\bullet\bullet} - x_{\bullet j\bullet} + x_{\bullet\bullet\bullet}) = b x_{i\bullet\bullet} - b x_{i\bullet\bullet} - b x_{\bullet\bullet\bullet} + b x_{\bullet\bullet\bullet} = 0 \quad \forall\, i\,.
\end{aligned}
$$

Incidentalmente, il valore minimo della funzione obiettivo coincide con la devianza non spiegata

$$
\min_{\mu_0,\alpha,\beta,\alpha\beta} \mathbb{S}(\mu_0,\alpha,\beta,\alpha\beta) = \sum_{ijk}(x_{ijk} - x_{ij\bullet})^2 = SS_u
$$

ovvero con l'ultimo termine quadratico, non negativo, della (28.13).
Allo stesso risultato si perviene, ovviamente, applicando il metodo dei moltiplicatori di Lagrange.

Variabili di chi-quadrato
Dalla relazione (28.13) si deduce:

$$
\begin{aligned}
\frac{1}{\sigma^2}\sum_{ijk}(x_{ijk} - \mu_{ij})^2 &= \frac{1}{\sigma^2}\sum_{ijk}(x_{\bullet\bullet\bullet} - \mu_0)^2 + \\
&+ \frac{1}{\sigma^2}\sum_{ijk}(x_{i\bullet\bullet} - x_{\bullet\bullet\bullet} - \alpha_i)^2 + \\
&+ \frac{1}{\sigma^2}\sum_{ijk}(x_{\bullet j\bullet} - x_{\bullet\bullet\bullet} - \beta_j)^2 + \\
&+ \frac{1}{\sigma^2}\sum_{ijk}[x_{ij\bullet} - x_{i\bullet\bullet} - x_{\bullet j\bullet} + x_{\bullet\bullet\bullet} - (\alpha\beta)_{ij}]^2 + \\
&+ \frac{1}{\sigma^2}\sum_{ijk}(x_{ijk} - x_{ij\bullet})^2
\end{aligned}
\tag{28.13}
$$

Ma poiché, con ovvio significato dei simboli, deve aversi

$$
\begin{aligned}
x_{\bullet\bullet\bullet} &= \mu_0 + \varepsilon_{\bullet\bullet\bullet} \\
x_{i\bullet\bullet} &= \mu_0 + \alpha_i + \varepsilon_{i\bullet\bullet} \\
x_{\bullet j\bullet} &= \mu_0 + \beta_j + \varepsilon_{\bullet j\bullet} \\
x_{ij\bullet} &= \mu_0 + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ij\bullet}
\end{aligned}
$$

risulta

$$
\begin{aligned}
\frac{x_{\bullet\bullet\bullet} - \mu_0}{\sigma} &= \frac{\varepsilon_{\bullet\bullet\bullet}}{\sigma} \\
\frac{x_{i\bullet\bullet} - x_{\bullet\bullet\bullet} - \alpha_i}{\sigma} &= \frac{\varepsilon_{i\bullet\bullet} - \varepsilon_{\bullet\bullet\bullet}}{\sigma} \\
\frac{x_{\bullet j\bullet} - x_{\bullet\bullet\bullet} - \beta_j}{\sigma} &= \frac{\varepsilon_{\bullet j\bullet} - \varepsilon_{\bullet\bullet\bullet}}{\sigma} \\
\frac{x_{ij\bullet} - x_{i\bullet\bullet} - x_{\bullet j\bullet} + x_{\bullet\bullet\bullet} - (\alpha\beta)_{ij}}{\sigma} &= \frac{\varepsilon_{ij\bullet} - \varepsilon_{i\bullet\bullet} - \varepsilon_{\bullet j\bullet} + \varepsilon_{\bullet\bullet\bullet}}{\sigma} \\
\frac{x_{ijk} - x_{ij\bullet}}{\sigma} &= \frac{\varepsilon_{ijk} - \varepsilon_{ij\bullet}}{\sigma}
\end{aligned}
$$

e quindi

$$
\frac{1}{\sigma^2}\sum_{ijk}(x_{\bullet\bullet\bullet} - \mu_0)^2 = \sum_{ijk}\Big(\frac{\varepsilon_{\bullet\bullet\bullet}}{\sigma}\Big)^2 = \Big\langle \frac{\varepsilon}{\sigma}\Big|(\mathbb{I} - P_1)\otimes(\mathbb{I} - P_2)\otimes(\mathbb{I} - P_3)\frac{\varepsilon}{\sigma}\Big\rangle
$$

$$
\frac{1}{\sigma^2}\sum_{ijk}(x_{i\bullet\bullet} - x_{\bullet\bullet\bullet} - \alpha_i)^2 = \sum_{ijk}\Big(\frac{\varepsilon_{i\bullet\bullet} - \varepsilon_{\bullet\bullet\bullet}}{\sigma}\Big)^2 = \Big\langle \frac{\varepsilon}{\sigma}\Big|P_1\otimes(\mathbb{I} - P_2)\otimes(\mathbb{I} - P_3)\frac{\varepsilon}{\sigma}\Big\rangle
$$

$$
\frac{1}{\sigma^2}\sum_{ijk}(x_{\bullet j\bullet} - x_{\bullet\bullet\bullet} - \beta_j)^2 = \sum_{ijk}\Big(\frac{\varepsilon_{\bullet j\bullet} - \varepsilon_{\bullet\bullet\bullet}}{\sigma}\Big)^2 = \Big\langle \frac{\varepsilon}{\sigma}\Big|(\mathbb{I} - P_1)\otimes P_2\otimes(\mathbb{I} - P_3)\frac{\varepsilon}{\sigma}\Big\rangle
$$

$$
\begin{aligned}
\frac{1}{\sigma^2}\sum_{ijk}[x_{ij\bullet} - x_{i\bullet\bullet} - x_{\bullet j\bullet} + x_{\bullet\bullet\bullet} - (\alpha\beta)_{ij}]^2 &= \sum_{ijk}\Big(\frac{\varepsilon_{ij\bullet} - \varepsilon_{i\bullet\bullet} - \varepsilon_{\bullet j\bullet} + \varepsilon_{\bullet\bullet\bullet}}{\sigma}\Big)^2 = \\
&= \Big\langle \frac{\varepsilon}{\sigma}\Big|P_1\otimes P_2\otimes(\mathbb{I} - P_3)\frac{\varepsilon}{\sigma}\Big\rangle
\end{aligned}
$$

$$
\frac{1}{\sigma^2}\sum_{ijk}(x_{ijk} - x_{ij\bullet})^2 = \sum_{ijk}\Big(\frac{\varepsilon_{ijk} - \varepsilon_{ij\bullet}}{\sigma}\Big)^2 = \Big\langle \frac{\varepsilon}{\sigma}\Big|\mathbb{I}\otimes\mathbb{I}\otimes P_3\frac{\varepsilon}{\sigma}\Big\rangle
$$

con le variabili casuali ε/σ gaussiane, standard e stocasticamente indipendenti. Si riconosce allora che i termini a secondo membro nella (28.13) sono forme quadratiche delle variabili ε/σ con operatori di rappresentazione idempotenti e a composizione nulla

$$
\begin{aligned}
&(\mathbb{I} - P_1)\otimes(\mathbb{I} - P_2)\otimes(\mathbb{I} - P_3) \\
&P_1\otimes(\mathbb{I} - P_2)\otimes(\mathbb{I} - P_3) \\
&(\mathbb{I} - P_1)\otimes P_2\otimes(\mathbb{I} - P_3) \\
&P_1\otimes P_2\otimes(\mathbb{I} - P_3) \\
&\mathbb{I}\otimes\mathbb{I}\otimes P_3
\end{aligned}
$$

e che costituiscono pertanto delle variabili di chi-quadrato stocasticamente indipendenti con gradi di libertà rispettivi:

$$
\begin{aligned}
&\text{g.d.l.}\Big\langle \frac{\varepsilon}{\sigma}\Big|(\mathbb{I}-P_1)\otimes(\mathbb{I}-P_2)\otimes(\mathbb{I}-P_3)\frac{\varepsilon}{\sigma}\Big\rangle \;=\; \mathrm{tr}[(\mathbb{I}-P_1)\otimes(\mathbb{I}-P_2)\otimes(\mathbb{I}-P_3)] \;=\; 1\\
&\text{g.d.l.}\Big\langle \frac{\varepsilon}{\sigma}\Big|P_1\otimes(\mathbb{I}-P_2)\otimes(\mathbb{I}-P_3)\frac{\varepsilon}{\sigma}\Big\rangle \;=\; \mathrm{tr}[P_1\otimes(\mathbb{I}-P_2)\otimes(\mathbb{I}-P_3)] \;=\; a-1\\
&\text{g.d.l.}\Big\langle \frac{\varepsilon}{\sigma}\Big|(\mathbb{I}-P_1)\otimes P_2\otimes(\mathbb{I}-P_3)\frac{\varepsilon}{\sigma}\Big\rangle \;=\; \mathrm{tr}[(\mathbb{I}-P_1)\otimes P_2\otimes(\mathbb{I}-P_3)] \;=\; b-1\\
&\text{g.d.l.}\Big\langle \frac{\varepsilon}{\sigma}\Big|P_1\otimes P_2\otimes(\mathbb{I}-P_3)\frac{\varepsilon}{\sigma}\Big\rangle \;=\; \mathrm{tr}[P_1\otimes P_2\otimes(\mathbb{I}-P_3)] \;=\; (a-1)(b-1)\\
&\text{g.d.l.}\Big\langle \frac{\varepsilon}{\sigma}\Big|\mathbb{I}\otimes\mathbb{I}\otimes P_3\frac{\varepsilon}{\sigma}\Big\rangle \;=\; \mathrm{tr}[\mathbb{I}\otimes\mathbb{I}\otimes P_3] \;=\; ab(c-1)\,.
\end{aligned}
$$

Nel seguito, per brevità, si farà uso delle notazioni:

$$
\begin{aligned}
\mathcal{X}^2{}_1 \;&=\; \frac{1}{\sigma^2}\sum_{ijk}(x_{\bullet\bullet\bullet}-\mu_0)^2 \;=\; \frac{1}{\sigma^2}\sum_{ijk}\varepsilon_{\bullet\bullet\bullet}{}^2\\
\mathcal{X}^2{}_{a-1} \;&=\; \frac{1}{\sigma^2}\sum_{ijk}(x_{i\bullet\bullet}-x_{\bullet\bullet\bullet}-\alpha_i)^2 \;=\; \frac{1}{\sigma^2}\sum_{ijk}(\varepsilon_{i\bullet\bullet}-\varepsilon_{\bullet\bullet\bullet})^2\\
\mathcal{X}^2{}_{b-1} \;&=\; \frac{1}{\sigma^2}\sum_{ijk}(x_{\bullet j\bullet}-x_{\bullet\bullet\bullet}-\beta_j)^2 \;=\; \frac{1}{\sigma^2}\sum_{ijk}(\varepsilon_{\bullet j\bullet}-\varepsilon_{\bullet\bullet\bullet})^2\\
\mathcal{X}^2{}_{(a-1)(b-1)} \;&=\; \frac{1}{\sigma^2}\sum_{ijk}[x_{ij\bullet}-x_{i\bullet\bullet}-x_{\bullet j\bullet}+x_{\bullet\bullet\bullet}-(\alpha\beta)_{ij}]^2 \;=\\
&=\; \frac{1}{\sigma^2}\sum_{ijk}[\varepsilon_{ij\bullet}-\varepsilon_{i\bullet\bullet}-\varepsilon_{\bullet j\bullet}+\varepsilon_{\bullet\bullet\bullet}]^2\\
\mathcal{X}^2{}_{ab(c-1)} \;&=\; \frac{1}{\sigma^2}\sum_{ijk}(x_{ijk}-x_{ij\bullet})^2 \;=\; \frac{1}{\sigma^2}\sum_{ijk}(\varepsilon_{ijk}-\varepsilon_{ij\bullet})^2\,.
\end{aligned}
\tag{28.14}
$$

Variabili di Fisher
Da quanto dimostrato al punto precedente appare chiaro che la variabile casuale

$$
F_{\mu_0} \;=\; \frac{ab(c-1)\mathcal{X}^2{}_1}{\mathcal{X}^2{}_{ab(c-1)}} \;=\; ab(c-1)\frac{\displaystyle\sum_{ijk}(x_{\bullet\bullet\bullet}-\mu_0)^2}{\displaystyle\sum_{ijk}(x_{ijk}-x_{ij\bullet})^2}
$$

è il quoziente di due variabili di chi-quadrato ridotte e indipendenti, rispettivamente a 1 e $ab(c-1)$ gradi di libertà, per cui seguirà una distribuzione di Fisher a $(1, ab(c-1))$ gradi di libertà. In modo analogo si riconosce in

$$
F_\alpha \;=\; \frac{ab(c-1)\mathcal{X}^2{}_{a-1}}{(a-1)\mathcal{X}^2{}_{ab(c-1)}} \;=\; \frac{ab(c-1)}{a-1}\frac{\displaystyle\sum_{ijk}(x_{i\bullet\bullet}-x_{\bullet\bullet\bullet}-\alpha_i)^2}{\displaystyle\sum_{ijk}(x_{ijk}-x_{ij\bullet})^2}
$$

una variabile di Fisher a $(a-1, ab(c-1))$ gradi di libertà, mentre

$$F_\beta = \frac{ab(c-1)\mathcal{X}^2{}_{b-1}}{(b-1)\mathcal{X}^2{}_{ab(c-1)}} = \frac{ab(c-1)}{b-1}\frac{\displaystyle\sum_{ijk}(x_{\bullet j\bullet} - x_{\bullet\bullet\bullet} - \beta_j)^2}{\displaystyle\sum_{ijk}(x_{ijk} - x_{ij\bullet})^2}$$

è una variabile F a $(b-1, ab(c-1))$ gradi di libertà. Infine:

$$\begin{aligned} F_{\alpha\beta} &= \frac{ab(c-1)\mathcal{X}^2{}_{(a-1)(b-1)}}{(a-1)(b-1)\mathcal{X}^2{}_{ab(c-1)}} = \\ &= \frac{ab(c-1)}{(a-1)(b-1)}\frac{\displaystyle\sum_{ijk}[x_{ij\bullet} - x_{i\bullet\bullet} - x_{\bullet j\bullet} + x_{\bullet\bullet\bullet} - (\alpha\beta)_{ij}]^2}{\displaystyle\sum_{ijk}(x_{ijk} - x_{ij\bullet})^2} \end{aligned}$$

segue una distribuzione di Fisher a $((a-1)(b-1), ab(c-1))$ gradi di libertà.

Variabili di Student
I risultati dei punti precedenti permettono di scrivere le relazioni:

$$\begin{aligned} \frac{x_{\bullet\bullet\bullet} - \mu_0}{\sigma} &= \Big[(\mathbb{I} - P_1)\otimes(\mathbb{I} - P_2)\otimes(\mathbb{I} - P_3)\frac{\varepsilon}{\sigma}\Big]_{ijk} \\ \frac{x_{i\bullet\bullet} - x_{\bullet\bullet\bullet} - \alpha_i}{\sigma} &= \Big[P_1\otimes(\mathbb{I} - P_2)\otimes(\mathbb{I} - P_3)\frac{\varepsilon}{\sigma}\Big]_{ijk} \\ \frac{x_{\bullet j\bullet} - x_{\bullet\bullet\bullet} - \beta_j}{\sigma} &= \Big[(\mathbb{I} - P_1)\otimes P_2\otimes(\mathbb{I} - P_3)\frac{\varepsilon}{\sigma}\Big]_{ijk} \\ \frac{x_{ij\bullet} - x_{i\bullet\bullet} - x_{\bullet j\bullet} + x_{\bullet\bullet\bullet} - (\alpha\beta)_{ij}}{\sigma} &= \Big[P_1\otimes P_2\otimes(\mathbb{I} - P_3)\frac{\varepsilon}{\sigma}\Big]_{ijk} \end{aligned}$$

che, convenendo la somma sugli indici ripetuti, possono riesprimersi in forma matriciale:

$$\begin{aligned} \frac{x_{\bullet\bullet\bullet} - \mu_0}{\sigma} &= \big[(\mathbb{I} - P_1)\otimes(\mathbb{I} - P_2)\otimes(\mathbb{I} - P_3)\big]_{ijk,i_1j_1k_1}\frac{\varepsilon_{i_1j_1k_1}}{\sigma} \qquad \forall\, i,j,k \\ \frac{x_{i\bullet\bullet} - x_{\bullet\bullet\bullet} - \alpha_i}{\sigma} &= \big[P_1\otimes(\mathbb{I} - P_2)\otimes(\mathbb{I} - P_3)\big]_{ijk,i_1j_1k_1}\frac{\varepsilon_{ijk,i_1j_1k_1}}{\sigma} \\ \frac{x_{\bullet j\bullet} - x_{\bullet\bullet\bullet} - \beta_j}{\sigma} &= \big[(\mathbb{I} - P_1)\otimes P_2\otimes(\mathbb{I} - P_3)\big]_{ijk,i_1j_1k_1}\frac{\varepsilon_{ijk,i_1j_1k_1}}{\sigma} \\ \frac{x_{ij\bullet} - x_{i\bullet\bullet} - x_{\bullet j\bullet} + x_{\bullet\bullet\bullet} - (\alpha\beta)_{ij}}{\sigma} &= \big[P_1\otimes P_2\otimes(\mathbb{I} - P_3)\big]_{ijk,i_1j_1k_1}\frac{\varepsilon_{,i_1j_1k_1}}{\sigma}\,. \end{aligned}$$

Queste combinazioni lineari delle variabili normali standard ε_{ijk}/σ sono distribuite normalmente e con varianze rispettive:

$$\begin{aligned} &\sum_{i_1j_1k_1}\big[(\mathbb{I} - P_1)\otimes(\mathbb{I} - P_2)\otimes(\mathbb{I} - P_3)\big]^2_{ijk,i_1j_1k_1} = \\ &= \big[(\mathbb{I} - P_1)\otimes(\mathbb{I} - P_2)\otimes(\mathbb{I} - P_3)\big]_{ijk,ijk} = \frac{1}{abc} \end{aligned}$$

$$\sum_{i_1 j_1 k_1} \big[P_1 \otimes (\mathbb{I}-P_2) \otimes (\mathbb{I}-P_3)\big]^2_{ijk,i_1j_1k_1} =$$
$$= \big[P_1 \otimes (\mathbb{I}-P_2) \otimes (\mathbb{I}-P_3)\big]_{ijk,ijk} = \Big(1-\frac{1}{a}\Big)\frac{1}{bc} = \frac{a-1}{abc}$$

$$\sum_{i_1 j_1 k_1} \big[(\mathbb{I}-P_1) \otimes P_2 \otimes (\mathbb{I}-P_3)\big]^2_{ijk,i_1j_1k_1} =$$
$$= \big[(\mathbb{I}-P_1) \otimes P_2 \otimes (\mathbb{I}-P_3)\big]_{ijk,ijk} = \Big(1-\frac{1}{b}\Big)\frac{1}{ac} = \frac{b-1}{abc}$$

$$\sum_{i_1 j_1 k_1} \big[P_1 \otimes P_2 \otimes (\mathbb{I}-P_3)\big]^2_{ijk,i_1j_1k_1} =$$
$$= \big[P_1 \otimes P_2 \otimes (\mathbb{I}-P_3)\big]_{ijk,ijk} = \Big(1-\frac{1}{a}\Big)\Big(1-\frac{1}{b}\Big)\frac{1}{c} = \frac{(a-1)(b-1)}{abc}$$

$\forall\, i,j,k$ fissati. Elevando al quadrato si ottiene poi:

$$\Big(\frac{x_{\bullet\bullet\bullet}-\mu_0}{\sigma}\Big)^2 = \big[(\mathbb{I}-P_1) \otimes (\mathbb{I}-P_2) \otimes (\mathbb{I}-P_3)\big]_{ijk,i_1j_1k_1} \big[(\mathbb{I}-P_1) \otimes (\mathbb{I}-P_2) \otimes (\mathbb{I}-P_3)\big]_{ijk,i_2j_2k_2} \frac{\varepsilon_{i_1j_1k_1}}{\sigma}\frac{\varepsilon_{i_2j_2k_2}}{\sigma}$$

$$\Big(\frac{x_{i\bullet\bullet}-x_{\bullet\bullet\bullet}-\alpha_i}{\sigma}\Big)^2 = \big[P_1 \otimes (\mathbb{I}-P_2) \otimes (\mathbb{I}-P_3)\big]_{ijk,i_1j_1k_1} \big[P_1 \otimes (\mathbb{I}-P_2) \otimes (\mathbb{I}-P_3)\big]_{ijk,i_2j_2k_2} \frac{\varepsilon_{i_1j_1k_1}}{\sigma}\frac{\varepsilon_{i_2j_2k_2}}{\sigma}$$

$$\Big(\frac{x_{\bullet j\bullet}-x_{\bullet\bullet\bullet}-\beta_j}{\sigma}\Big)^2 = \big[(\mathbb{I}-P_1) \otimes P_2 \otimes (\mathbb{I}-P_3)\big]_{ijk,i_1j_1k_1} \big[(\mathbb{I}-P_1) \otimes P_2 \otimes (\mathbb{I}-P_3)\big]_{ijk,i_2j_2k_2} \frac{\varepsilon_{i_1j_1k_1}}{\sigma}\frac{\varepsilon_{i_2j_2k_2}}{\sigma}$$

$$\Big(\frac{x_{ij\bullet}-x_{i\bullet\bullet}-x_{\bullet j\bullet}+x_{\bullet\bullet\bullet}-(\alpha\beta)_{ij}}{\sigma}\Big)^2 = \big[P_1 \otimes P_2 \otimes (\mathbb{I}-P_3)\big]_{ijk,i_1j_1k_1} \big[P_1 \otimes P_2 \otimes (\mathbb{I}-P_3)\big]_{ijk,i_2j_2k_2} \frac{\varepsilon_{,i_1j_1k_1}}{\sigma}\frac{\varepsilon_{,i_2j_2k_2}}{\sigma}\,.$$

Si tratta di forme quadratiche delle variabili gaussiane standard e indipendenti ε_{ijk}/σ, con matrici di rappresentazione rispettive:

$$\big[(\mathbb{I}-P_1) \otimes (\mathbb{I}-P_2) \otimes (\mathbb{I}-P_3)\big]_{ijk,i_1j_1k_1} \big[(\mathbb{I}-P_1) \otimes (\mathbb{I}-P_2) \otimes (\mathbb{I}-P_3)\big]_{ijk,i_2j_2k_2}$$

$$\big[P_1 \otimes (\mathbb{I}-P_2) \otimes (\mathbb{I}-P_3)\big]_{ijk,i_1j_1k_1} \big[P_1 \otimes (\mathbb{I}-P_2) \otimes (\mathbb{I}-P_3)\big]_{ijk,i_2j_2k_2}$$

$$\big[(\mathbb{I}-P_1) \otimes P_2 \otimes (\mathbb{I}-P_3)\big]_{ijk,i_1j_1k_1} \big[(\mathbb{I}-P_1) \otimes P_2 \otimes (\mathbb{I}-P_3)\big]_{ijk,i_2j_2k_2}$$

$$\big[P_1 \otimes P_2 \otimes (\mathbb{I}-P_3)\big]_{ijk,i_1j_1k_1} \big[P_1 \otimes P_2 \otimes (\mathbb{I}-P_3)\big]_{ijk,i_2j_2k_2}$$

$\forall\, i_1j_1k_1, i_2j_2k_2$. Per contro, si è già verificato che

$$\frac{1}{\sigma^2}\sum_{ijk}(x_{ijk}-x_{ij\bullet})^2 \;=\; \big[\mathbb{I}\otimes\mathbb{I}\otimes P_3\big]_{i_1j_1k_1,i_2j_2k_2}\frac{\varepsilon_{i_1j_1k_1}}{\sigma}\frac{\varepsilon_{i_2j_2k_2}}{\sigma}$$

è una forma quadratica delle ε_{ijk}/σ con matrice rapresentativa

$$\big[\mathbb{I}\otimes\mathbb{I}\otimes P_3\big]_{i_1j_1k_1,i_2j_2k_2} \qquad \forall\, i_1j_1k_1, i_2j_2k_2\,.$$

Si potrà allora verificare l'indipendenza stocastica delle variabili casuali

$$\Big(\frac{x_{\bullet\bullet\bullet}-\mu_0}{\sigma}\Big)^2 \qquad \text{e} \qquad \frac{1}{\sigma^2}\sum_{ijk}(x_{ijk}-x_{ij\bullet})^2$$

accertando l'annullarsi del prodotto delle relative matrici rappresentative

$$\begin{aligned}
&\big[(\mathbb{I}-P_1)\otimes(\mathbb{I}-P_2)\otimes(\mathbb{I}-P_3)\big]_{ijk,i_1j_1k_1}\\
&\big[(\mathbb{I}-P_1)\otimes(\mathbb{I}-P_2)\otimes(\mathbb{I}-P_3)\big]_{ijk,i_2j_2k_2}\big[\mathbb{I}\otimes\mathbb{I}\otimes P_3\big]_{i_2j_2k_2,i_3j_3k_3} =\\
=&\big[(\mathbb{I}-P_1)\otimes(\mathbb{I}-P_2)\otimes(\mathbb{I}-P_3)\big]_{ijk,i_1j_1k_1}\\
&\big[(\mathbb{I}-P_1)\otimes(\mathbb{I}-P_2)\otimes(\mathbb{I}-P_3)\,\cdot\,\mathbb{I}\otimes\mathbb{I}\otimes P_3\big]_{ijk,i_3j_3k_3} =\\
=&\big[(\mathbb{I}-P_1)\otimes(\mathbb{I}-P_2)\otimes(\mathbb{I}-P_3)\big]_{ijk,i_1j_1k_1}0 \;=\; 0 \qquad \forall i_1j_1k_1, i_3j_3k_3
\end{aligned}$$

in accordo con il teorema di Craig. In modo analogo si dimostra l'indipendenza stocastica delle coppie di variabii

$$\Big(\frac{x_{i\bullet\bullet}-x_{\bullet\bullet\bullet}-\alpha_i}{\sigma}\Big)^2 \qquad \text{e} \qquad \frac{1}{\sigma^2}\sum_{ijk}(x_{ijk}-x_{ij\bullet})^2$$

$$\Big(\frac{x_{\bullet j\bullet}-x_{\bullet\bullet\bullet}-\beta_j}{\sigma}\Big)^2 \qquad \text{e} \qquad \frac{1}{\sigma^2}\sum_{ijk}(x_{ijk}-x_{ij\bullet})^2$$

$$\Big(\frac{x_{ij\bullet}-x_{i\bullet\bullet}-x_{\bullet j\bullet}+x_{\bullet\bullet\bullet}-(\alpha\beta)_{ij}}{\sigma}\Big)^2 \qquad \text{e} \qquad \frac{1}{\sigma^2}\sum_{ijk}(x_{ijk}-x_{ij\bullet})^2\,.$$

È così dato di definire le variabili di Student:

$$t_{\mu_0} \;=\; \sqrt{ab(c-1)}\,\frac{\dfrac{x_{\bullet\bullet\bullet}-\mu_0}{\sigma}\Big(\dfrac{1}{abc}\Big)^{-1/2}}{\sqrt{\dfrac{1}{\sigma^2}\displaystyle\sum_{ijk}(x_{ijk}-x_{ij\bullet})^2}} \;=\; ab\sqrt{c(c-1)}\,\frac{x_{\bullet\bullet\bullet}-\mu_0}{\sqrt{\displaystyle\sum_{ijk}(x_{ijk}-x_{ij\bullet})^2}}$$

$$t_{\alpha_i} \;=\; \sqrt{ab(c-1)}\,\frac{\dfrac{x_{i\bullet\bullet}-x_{\bullet\bullet\bullet}-\alpha_i}{\sigma}\Big(\dfrac{a-1}{abc}\Big)^{-1/2}}{\sqrt{\dfrac{1}{\sigma^2}\displaystyle\sum_{ijk}(x_{ijk}-x_{ij\bullet})^2}} \;=\; ab\sqrt{\frac{c(c-1)}{a-1}}\,\frac{x_{i\bullet\bullet}-x_{\bullet\bullet\bullet}-\alpha_i}{\sqrt{\displaystyle\sum_{ijk}(x_{ijk}-x_{ij\bullet})^2}}$$

$$t_{\beta_j} = \sqrt{ab(c-1)}\frac{\dfrac{x_{\bullet j\bullet}-x_{\bullet\bullet\bullet}-\beta_j}{\sigma}\Big(\dfrac{b-1}{abc}\Big)^{-1/2}}{\sqrt{\dfrac{1}{\sigma^2}\displaystyle\sum_{ijk}(x_{ijk}-x_{ij\bullet})^2}} = ab\sqrt{\frac{c(c-1)}{b-1}}\frac{x_{\bullet j\bullet}-x_{\bullet\bullet\bullet}-\beta_j}{\sqrt{\displaystyle\sum_{ijk}(x_{ijk}-x_{ij\bullet})^2}}$$

$$t_{(\alpha\beta)_{ij}} = \sqrt{ab(c-1)}\frac{\dfrac{x_{ij\bullet}-x_{i\bullet\bullet}-x_{\bullet j\bullet}+x_{\bullet\bullet\bullet}-(\alpha\beta)_{ij}}{\sigma}\Big(\dfrac{(a-1)(b-1)}{abc}\Big)^{-1/2}}{\sqrt{\dfrac{1}{\sigma^2}\displaystyle\sum_{ijk}(x_{ijk}-x_{ij\bullet})^2}} =$$

$$= ab\sqrt{\frac{c(c-1)}{(a-1)(b-1)}}\frac{x_{ij\bullet}-x_{i\bullet\bullet}-x_{\bullet j\bullet}+x_{\bullet\bullet\bullet}-(\alpha\beta)_{ij}}{\sqrt{\displaystyle\sum_{ijk}(x_{ijk}-x_{ij\bullet})^2}}$$

tutte a $ab(c-1)$ gradi di libertà. Le stesse variabili possono anche indicarsi nella forma più compatta:

$$\begin{aligned} t_{\mu_0} &= \sqrt{abc}\,\frac{x_{\bullet\bullet\bullet}-\mu_0}{s} \\ t_{\alpha_i} &= \sqrt{\frac{abc}{a-1}}\,\frac{x_{i\bullet\bullet}-x_{\bullet\bullet\bullet}-\alpha_i}{s} \\ t_{\beta_j} &= \sqrt{\frac{abc}{b-1}}\,\frac{x_{\bullet j\bullet}-x_{\bullet\bullet\bullet}-\beta_j}{s} \\ t_{(\alpha\beta)_{ij}} &= \sqrt{\frac{abc}{(a-1)(b-1)}}\,\frac{x_{ij\bullet}-x_{i\bullet\bullet}-x_{\bullet j\bullet}-x_{\bullet\bullet\bullet}-(\alpha\beta)_{ij}}{s} \end{aligned} \tag{28.15}$$

dove

$$s^2 = \frac{1}{ab(c-1)}SS_u = \frac{1}{ab(c-1)}\sigma^2\mathcal{X}^2{}_{ab(c-1)} = \frac{1}{ab(c-1)}\sum_{ijk}(x_{ijk}-x_{ij\bullet})^2 \tag{28.16}$$

è la **varianza** del sistema di dati **non spiegata** dal modello a due fattori.

Intervalli di confidenza dei parametri
Gli intervalli di confidenza dei parametri possono essere determinati sia utilizzando le variabili di Fisher sopra individuate, sia facendo ricorso alle variabili di Student discusse al punto precedente.

Poiché t_{μ_0} segue una distribuzione di Student a $ab(c-1)$ gradi di libertà, $\forall\,\alpha \in (0,1)$ assegnato la disequazione

$$t_{[\frac{\alpha}{2}](ab(c-1))} \le t_{\mu_0} \le t_{[1-\frac{\alpha}{2}](ab(c-1))}$$

è verificata con probabilità $1-\alpha$. Esplicitando la variabile casuale e ricordata la simmetria della distribuzione di Student, la disequazione assume la forma

$$-t_{[1-\frac{\alpha}{2}](ab(c-1))} \le \sqrt{abc}\,\frac{x_{\bullet\bullet\bullet}-\mu_0}{s} \le t_{[1-\frac{\alpha}{2}](ab(c-1))}$$

dalla quale si deduce l'espressione per l'intervallo di confidenza del valor medio generale μ_0

$$x_{\bullet\bullet\bullet} - \sqrt{\frac{1}{abc}}\, s\, t_{[1-\frac{\alpha}{2}](ab(c-1))} \le \mu_0 \le x_{\bullet\bullet\bullet} + \sqrt{\frac{1}{abc}}\, s\, t_{[1-\frac{\alpha}{2}](ab(c-1))}$$

con livello di confidenza $1-\alpha$.
In modo analogo si ricavano gli intervalli di confidenza per i coefficienti α_i, β_j e $(\alpha\beta)_{ij}$:

$$x_{i\bullet\bullet} - x_{\bullet\bullet\bullet} - \sqrt{\frac{a-1}{abc}}\, s\, t_{[1-\frac{\alpha}{2}](ab(c-1))} \le \alpha_i \le x_{i\bullet\bullet} - x_{\bullet\bullet\bullet} + \sqrt{\frac{a-1}{abc}}\, s\, t_{[1-\frac{\alpha}{2}](ab(c-1))}$$

$$x_{\bullet j\bullet} - x_{\bullet\bullet\bullet} - \sqrt{\frac{b-1}{abc}}\, s\, t_{[1-\frac{\alpha}{2}](ab(c-1))} \le \beta_j \le x_{\bullet j\bullet} - x_{\bullet\bullet\bullet} + \sqrt{\frac{b-1}{abc}}\, s\, t_{[1-\frac{\alpha}{2}](ab(c-1))}$$

$$x_{ij\bullet} - x_{i\bullet\bullet} - x_{\bullet j\bullet} - x_{\bullet\bullet\bullet} - \sqrt{\frac{(a-1)(b-1)}{abc}}\, s\, t_{[1-\frac{\alpha}{2}](ab(c-1))} \le (\alpha\beta)_{ij} \le$$

$$\le x_{ij\bullet} - x_{i\bullet\bullet} - x_{\bullet j\bullet} - x_{\bullet\bullet\bullet} + \sqrt{\frac{(a-1)(b-1)}{abc}}\, s\, t_{[1-\frac{\alpha}{2}](ab(c-1))}$$

sempre con pari livello di confidenza $1-\alpha$. Si osservi che le variabili casuali t_{μ_0}, t_{α_i}, t_{β_j}, $(\alpha\beta)_{ij}$ non risultano stocasticamente indipendenti, per cui gli eventi:

$$\begin{aligned}
&\mu_0 \in x_{\bullet\bullet\bullet} \pm \sqrt{\frac{1}{abc}}\, s\, t_{[1-\frac{\alpha}{2}](ab(c-1))} \\
&\alpha_i \in x_{i\bullet\bullet} - x_{\bullet\bullet\bullet} \pm \sqrt{\frac{a-1}{abc}}\, s\, t_{[1-\frac{\alpha}{2}](ab(c-1))} \\
&\beta_j \in x_{\bullet j\bullet} - x_{\bullet\bullet\bullet} \pm \sqrt{\frac{b-1}{abc}}\, s\, t_{[1-\frac{\alpha}{2}](ab(c-1))} \\
&(\alpha\beta)_{ij} \in x_{ij\bullet} - x_{i\bullet\bullet} - x_{\bullet j\bullet} - x_{\bullet\bullet\bullet} \pm \sqrt{\frac{(a-1)(b-1)}{abc}}\, s\, t_{[1-\frac{\alpha}{2}](ab(c-1))}
\end{aligned}$$

non possono essere considerati a loro volta indipendenti. Ciò implica che il simultaneo ricorrere di k condizioni del tipo precedente — appartenenza di k parametri del modello ai rispettivi intervalli di confidenza — non può individuarsi con il semplice prodotto $(1-\alpha)^k$ dei singoli livelli di confidenza.

Intervalli di confidenza simultanei

Un intervallo di confidenza per il parametro μ_0 può essere determinato per mezzo della statistica F_{μ_0}, che segue una distribuzione di Fisher a $(1, ab(c-1))$ gradi di libertà. La diseguaglianza

$$F_{\mu_0} \le F_{[1-\alpha](1,ab(c-1))} \qquad \Longleftrightarrow \qquad \frac{1}{s^2}\sum_{ijk}(x_{\bullet\bullet\bullet}-\mu_0)^2 \le F_{[1-\alpha](1,ab(c-1))}$$

è dunque verificata con probabilità $1-\alpha \in (0,1)$ preassegnata a piacere e conduce all'intervallo di confidenza

$$x_{\bullet\bullet\bullet} - s\frac{1}{\sqrt{abc}}\sqrt{F_{[1-\alpha](1,ab(c-1))}} \leq \mu_0 \leq x_{\bullet\bullet\bullet} + s\frac{1}{\sqrt{abc}}\sqrt{F_{[1-\alpha](1,ab(c-1))}}$$

con livello di confidenza $1-\alpha$.
Analogamente, con probabilità $1-\alpha$ è soddisfatta la diseguaglianza

$$F_\alpha \leq F_{[1-\alpha](a-1,ab(c-1))} \qquad \Longleftrightarrow \qquad \frac{1}{s^2}\frac{1}{a-1}\sum_{ijk}(x_{i\bullet\bullet} - x_{\bullet\bullet\bullet} - \alpha_i)^2 \leq F_{[1-\alpha](a-1,ab(c-1))}$$

dalla quale si ricava

$$\sum_i (x_{i\bullet\bullet} - x_{\bullet\bullet\bullet} - \alpha_i)^2 \leq s^2\frac{a-1}{bc}F_{[1-\alpha](a-1,ab(c-1))}$$

che implica

$$|x_{i\bullet\bullet} - x_{\bullet\bullet\bullet} - \alpha_i| \leq s\sqrt{\frac{a-1}{bc}}\sqrt{F_{[1-\alpha](a-1,ab(c-1))}}$$

uniformemente in $i = 1,\ldots,a$. Si ottengono così gli intervalli di confidenza **simultanei**

$$\alpha_i \in x_{i\bullet\bullet} - x_{\bullet\bullet\bullet} \pm s\sqrt{\frac{a-1}{bc}}\sqrt{F_{[1-\alpha](a-1,ab(c-1))}} \qquad i = 1,\ldots,a$$

con livello di confidenza complessivo $1-\alpha$ (la probabilità che tutti gli α_i appartengano simultaneamente agli intervalli sopra individuati è uguale a $1-\alpha$). Le diseguaglianze

$$F_\beta \leq F_{[1-\alpha](b-1,ab(c-1))} \qquad \Longleftrightarrow \qquad \frac{1}{s^2}\frac{1}{b-1}\sum_{ijk}(x_{\bullet j\bullet} - x_{\bullet\bullet\bullet} - \beta_j)^2 \leq F_{[1-\alpha](b-1,ab(c-1))}$$

$$F_{\alpha\beta} \leq F_{[1-\alpha]((a-1)(b-1),ab(c-1))} \qquad \Longleftrightarrow$$
$$\frac{1}{s^2}\frac{1}{(a-1)(b-1)}\sum_{ijk}(x_{ij\bullet} - x_{i\bullet\bullet} - x_{\bullet j\bullet} - x_{\bullet\bullet\bullet} - (\alpha\beta)_{ij})^2 \leq F_{[1-\alpha]((a-1)(b-1),ab(c-1))}$$

porgono infine gli intervalli di confidenza simultanei per i coefficienti β_j e $(\alpha\beta)_{ij}$:

$$\beta_j \in x_{\bullet j\bullet} - x_{\bullet\bullet\bullet} \pm s\sqrt{\frac{b-1}{ac}}\sqrt{F_{[1-\alpha](b-1,ab(c-1))}} \qquad j = 1,\ldots,b$$

$$(\alpha\beta)_{ij} \in x_{ij\bullet} - x_{i\bullet\bullet} - x_{\bullet j\bullet} - x_{\bullet\bullet\bullet} \pm s\sqrt{\frac{(a-1)(b-1)}{c}}\sqrt{F_{[1-\alpha]((a-1)(b-1),ab(c-1))}}$$
$$i = 1,\ldots,a,\ j = 1,\ldots,b.$$

Osservazione
Intervalli di confidenza simultanei per **tutti** i parametri possono essere ottenuti mediante la variabile

$$F = \frac{c-1}{\sum_{ijk}(x_{ijk}-x_{ij\bullet})^2}\sum_{ijk}\Big[(x_{\bullet\bullet\bullet}-\mu_0)^2+(x_{i\bullet\bullet}-x_{\bullet\bullet\bullet}-\alpha_i)^2+$$

$$+(x_{\bullet j\bullet}-x_{\bullet\bullet\bullet}-\beta_j)^2+[x_{ij\bullet}-x_{i\bullet\bullet}-x_{\bullet j\bullet}+x_{\bullet\bullet\bullet}-(\alpha\beta)_{ij}]^2\Big]$$

che segue una distribuzione di Fisher ad $(ab, ab(c-1))$ gradi di libertà. Le variabili di chi-quadrato $\mathcal{X}^2{}_1$, $\mathcal{X}^2{}_{a-1}$, $\mathcal{X}^2{}_{b-1}$ e $\mathcal{X}^2{}_{(a-1)(b-1)}$ definite in (28.14) sono infatti stocasticamente indipendenti fra loro e rispetto alla forma quadratica

$$\mathcal{X}^2{}_{ab(c-1)} = ab(c-1)\frac{s^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{ijk}(x_{ijk}-x_{ij\bullet})^2$$

variabile di $\mathcal{X}^2$ a $ab(c-1)$ gradi di libertà. Se ne conclude che la somma

$$\mathcal{X}^2{}_1+\mathcal{X}^2{}_{a-1}+\mathcal{X}^2{}_{b-1}+\mathcal{X}^2{}_{(a-1)(b-1)}$$

costituisce una variabile di $\mathcal{X}^2$ a $1+(a-1+(b-1)+(a-1)(b-1) = ab$ gradi di libertà stocasticamente indipendente da $ab(c-1)s^2/\sigma^2$. Di qui l'asserto, dal momento che

$$F = \frac{\mathcal{X}^2{}_1+\mathcal{X}^2{}_{a-1}+\mathcal{X}^2{}_{b-1}+\mathcal{X}^2{}_{(a-1)(b-1)}}{ab}\,\frac{ab(c-1)}{ab(c-1)s^2/\sigma^2}\,.$$

Comunque si fissi un $\alpha \in (0,1)$ la probabilità che si abbia $F \le F_{[1-\alpha](ab,ab(c-1))}$ sarà uguale a $1-\alpha$, la stessa probabilità che sia verificata la diseguaglianza

$$\frac{1}{ab}\frac{1}{s^2}\,\sigma^2\big[\mathcal{X}^2{}_1+\mathcal{X}^2{}_{a-1}+\mathcal{X}^2{}_{b-1}+\mathcal{X}^2{}_{(a-1)(b-1)}\big] \le F_{[1-\alpha](ab,ab(c-1))}$$

ovvero

$$\sigma\big[\mathcal{X}^2{}_1+\mathcal{X}^2{}_{a-1}+\mathcal{X}^2{}_{b-1}+\mathcal{X}^2{}_{(a-1)(b-1)}\big]^{1/2} \le s\sqrt{ab}\sqrt{F_{[1-\alpha](ab,ab(c-1))}}\,.$$

Di conseguenza, con probabilità $1-\alpha$ risultano simultaneamente soddisfatte tutte le diseguaglianze:

$$|x_{\bullet\bullet\bullet}-\mu_0| \le s\sqrt{ab}\sqrt{F_{[1-\alpha](ab,ab(c-1))}}$$
$$|x_{i\bullet\bullet}-x_{\bullet\bullet\bullet}| \le s\sqrt{ab}\sqrt{F_{[1-\alpha](ab,ab(c-1))}}$$
$$|x_{\bullet j\bullet}-x_{\bullet\bullet\bullet}| \le s\sqrt{ab}\sqrt{F_{[1-\alpha](ab,ab(c-1))}}$$
$$|x_{ij\bullet}-x_{i\bullet\bullet}-x_{\bullet j\bullet}+x_{\bullet\bullet\bullet}| \le s\sqrt{ab}\sqrt{F_{[1-\alpha](ab,ab(c-1))}}$$

e gli intervalli di confidenza simultanei di tutti i parametri assumono pertanto la forma:

$$\mu_0 \in x_{\bullet\bullet\bullet} \pm s\sqrt{ab}\sqrt{F_{[1-\alpha](ab,ab(c-1))}}$$
$$\alpha_i \in x_{i\bullet\bullet} - x_{\bullet\bullet\bullet} \pm s\sqrt{ab}\sqrt{F_{[1-\alpha](ab,ab(c-1))}}$$
$$x_{\bullet j\bullet} - x_{\bullet\bullet\bullet} \in \pm s\sqrt{ab}\sqrt{F_{[1-\alpha](ab,ab(c-1))}}$$
$$x_{ij\bullet} - x_{i\bullet\bullet} - x_{\bullet j\bullet} + x_{\bullet\bullet\bullet} \in \pm s\sqrt{ab}\sqrt{F_{[1-\alpha](ab,ab(c-1))}}$$

con livello di confidenza $1-\alpha$.

Test delle ipotesi sui singoli parametri

Criteri per verificare l'annullarsi di singoli parametri del modello ANOVA possono essere sviluppati immediatamente per mezzo delle variabili di Student (28.15). Così ad esempio l'ipotesi nulla $H_0 : \mu_0 = 0$ contro l'alternativa $H_1 : \mu_0 \neq 0$ verrà accettata con livello di significatività α se

$$\left|\sqrt{abc}\frac{x_{\bullet\bullet\bullet}}{s}\right| \leq t_{[1-\frac{\alpha}{2}](ab(c-1))}$$

e rigettata in caso contrario. Del tutto simili sono le condizioni di accettazione dell'ipotesi nulla $H_0 : \alpha_i = 0$ contro $H_1 : \alpha_i \neq 0$

$$\left|\sqrt{\frac{abc}{a-1}}\frac{x_{i\bullet\bullet} - x_{\bullet\bullet\bullet}}{s}\right| \leq t_{[1-\frac{\alpha}{2}](ab(c-1))}$$

come pure quelle per l'accettazione di $H_0 : \beta_j = 0$ contro $H_1 : \beta_j \neq 0$

$$\left|\sqrt{\frac{abc}{b-1}}\frac{x_{\bullet j\bullet} - x_{\bullet\bullet\bullet}}{s}\right| \leq t_{[1-\frac{\alpha}{2}](ab(c-1))} \,.$$

L'ipotesi $H_0 : (\alpha\beta)_{ij} = 0$ viene preferita all'alternativa $H_1 : (\alpha\beta)_{ij} \neq 0$ qualora risulti

$$\left|\sqrt{\frac{abc}{(a-1)(b-1)}}\frac{x_{ij\bullet} - x_{i\bullet\bullet} - x_{\bullet j\bullet} + x_{\bullet\bullet\bullet}}{s}\right| \leq t_{[1-\frac{\alpha}{2}](ab(c-1))}$$

sempre con livello di significatività α.
In modo analogo si possono costruire test per verificare valori presunti e non nulli dei parametri.

Variabili casuali utili per i test delle ipotesi su set di parametri

Nella elaborazione dei test delle ipotesi relativi ai set di parametri α_i, β_j o $(\alpha\beta)_{ij}$ — vedi seguito — si fa uso delle variabili casuali

$$
\begin{aligned}
\sum_{ijk} x_{\bullet\bullet\bullet}{}^2 &= \sum_{ijk} (\varepsilon_{\bullet\bullet\bullet} + \mu_0)^2 \\
SS_1 &= \sum_{ijk} (x_{i\bullet\bullet} - x_{\bullet\bullet\bullet})^2 = \sum_{ijk} (\varepsilon_{i\bullet\bullet} - \varepsilon_{\bullet\bullet\bullet} + \alpha_i)^2 \\
SS_2 &= \sum_{ijk} (x_{\bullet j\bullet} - x_{\bullet\bullet\bullet})^2 = \sum_{ijk} (\varepsilon_{\bullet j\bullet} - \varepsilon_{\bullet\bullet\bullet} + \beta_j)^2 \\
SS_{12} &= \sum_{ijk} (x_{ij\bullet} - x_{i\bullet\bullet} - x_{\bullet j\bullet} + x_{\bullet\bullet\bullet})^2 = \sum_{ijk} [\varepsilon_{ij\bullet} - \varepsilon_{i\bullet\bullet} - \varepsilon_{\bullet j\bullet} + \varepsilon_{\bullet\bullet\bullet} + (\alpha\beta)_{ij}]^2 \\
SS_u &= \sum_{ijk} (x_{ij\bullet} - x_{\bullet\bullet\bullet})^2 = \sum_{ijk} (\varepsilon_{ij\bullet} - \varepsilon_{\bullet\bullet\bullet})^2 .
\end{aligned}
$$

Di queste ha interesse determinare il valore medio, considerando che le ε_{ijk} sono per ipotesi variabili gaussiane indipendenti di media nulla e varianza σ^2, e che di conseguenza le variabili quadratiche:

$$
\begin{aligned}
\mathcal{X}^2{}_1 &= \frac{1}{\sigma^2} \sum_{ijk} \varepsilon_{\bullet\bullet\bullet}{}^2 \\
\mathcal{X}^2{}_{a-1} &= \frac{1}{\sigma^2} \sum_{ijk} (\varepsilon_{i\bullet\bullet} - \varepsilon_{\bullet\bullet\bullet})^2 \\
\mathcal{X}^2{}_{b-1} &= \frac{1}{\sigma^2} \sum_{ijk} (\varepsilon_{\bullet j\bullet} - \varepsilon_{\bullet\bullet\bullet})^2 \\
\mathcal{X}^2{}_{(a-1)(b-1)} &= \frac{1}{\sigma^2} \sum_{ijk} (\varepsilon_{ij\bullet} - \varepsilon_{i\bullet\bullet} - \varepsilon_{\bullet j\bullet} + \varepsilon_{\bullet\bullet\bullet})^2 \\
\mathcal{X}^2{}_{ab(c-1)} &= \frac{1}{\sigma^2} \sum_{ijk} (\varepsilon_{ij\bullet} - \varepsilon_{\bullet\bullet\bullet})^2
\end{aligned}
$$

sono variabili di $\mathcal{X}^2$ rispettivamente a 1, $a-1$, $b-1$, $(a-1)(b-1)$ e $ab(c-1)$ gradi di libertà. Si ha così

$$
\begin{aligned}
\mathbb{E}\Big[\sum_{ijk} x_{\bullet\bullet\bullet}{}^2\Big] &= \mathbb{E}\Big[\sum_{ijk} (\varepsilon_{\bullet\bullet\bullet} + \mu_0)^2\Big] = \\
&= \sigma^2 \mathbb{E}\Big[\frac{1}{\sigma^2} \sum_{ijk} \varepsilon_{\bullet\bullet\bullet}{}^2\Big] + 2 \sum_{ijk} \mathbb{E}[\varepsilon_{\bullet\bullet\bullet}]\mu_0 + \sum_{ijk} \mathbb{E}[\mu_0^2] = \\
&= \sigma^2 1 + 2 \sum_{ijk} 0 \mu_0 + \sum_{ijk} \mu_0^2 = \sigma^2 + \sum_{ijk} \mu_0^2
\end{aligned}
$$

e analogamente

$$
\begin{aligned}
\mathbb{E}\Big[\sum_{ijk}(x_{i\bullet\bullet}-x_{\bullet\bullet\bullet})^2\Big] &= \mathbb{E}\Big[\sum_{ijk}(\varepsilon_{i\bullet\bullet}-\varepsilon_{\bullet\bullet\bullet}+\alpha_i)^2\Big] = \\
&= \sigma^2\mathbb{E}\Big[\frac{1}{\sigma^2}\sum_{ijk}(\varepsilon_{i\bullet\bullet}-\varepsilon_{\bullet\bullet\bullet})^2\Big]+2\sum_{ijk}\mathbb{E}[\varepsilon_{i\bullet\bullet}-\varepsilon_{\bullet\bullet\bullet}]\alpha_i+\sum_{ijk}\alpha_i^2 = \\
&= \sigma^2(a-1)+2\sum_{ijk}0\alpha_i+\sum_{ijk}\alpha_i^2 = \sigma^2(a-1)+\sum_{ijk}\alpha_i^2\,.
\end{aligned}
$$

Vale inoltre

$$
\begin{aligned}
\mathbb{E}\Big[\sum_{ijk}(x_{\bullet j\bullet}-x_{\bullet\bullet\bullet})^2\Big] &= \mathbb{E}\Big[\sum_{ijk}(\varepsilon_{\bullet j\bullet}-\varepsilon_{\bullet\bullet\bullet}+\beta_j)^2\Big] = \\
&= \sigma^2\mathbb{E}\Big[\frac{1}{\sigma^2}\sum_{ijk}(\varepsilon_{\bullet j\bullet}-\varepsilon_{\bullet\bullet\bullet})^2\Big]+2\sum_{ijk}\mathbb{E}[\varepsilon_{\bullet j\bullet}-\varepsilon_{\bullet\bullet\bullet}]\beta_j+\sum_{ijk}\beta_j^2 = \\
&= \sigma^2(b-1)+2\sum_{ijk}0\beta_j+\sum_{ijk}\beta_j^2 = \sigma^2(b-1)+\sum_{ijk}\beta_j^2
\end{aligned}
$$

mentre

$$
\begin{aligned}
&\mathbb{E}\Big[\sum_{ijk}(x_{ij\bullet}-x_{i\bullet\bullet}-x_{\bullet j\bullet}+x_{\bullet\bullet\bullet})^2\Big] = \\
&=\mathbb{E}\Big[\sum_{ijk}[\varepsilon_{ij\bullet}-\varepsilon_{i\bullet\bullet}-\varepsilon_{\bullet j\bullet}+\varepsilon_{\bullet\bullet\bullet}+(\alpha\beta)_{ij}]^2\Big] = \\
&=\sigma^2\mathbb{E}\Big[\frac{1}{\sigma^2}\sum_{ijk}(\varepsilon_{ij\bullet}-\varepsilon_{i\bullet\bullet}-\varepsilon_{\bullet j\bullet}+\varepsilon_{\bullet\bullet\bullet})^2\Big]+ \\
&+2\sum_{ijk}\mathbb{E}[\varepsilon_{ij\bullet}-\varepsilon_{i\bullet\bullet}-\varepsilon_{\bullet j\bullet}+\varepsilon_{\bullet\bullet\bullet}](\alpha\beta)_{ij}+\sum_{ijk}(\alpha\beta)_{ij}^2 = \\
&=\sigma^2(a-1)(b-1)+2\sum_{ijk}0(\alpha\beta)_{ij}+\sum_{ijk}(\alpha\beta)_{ij}^2 = \\
&=\sigma^2(a-1)(b-1)+\sum_{ijk}(\alpha\beta)_{ij}^2
\end{aligned}
$$

ed infine

$$
\mathbb{E}\Big[\sum_{ijk}(x_{ijk}-x_{ij\bullet})^2\Big] = \sigma^2\mathbb{E}\Big[\frac{1}{\sigma^2}\sum_{ijk}(\varepsilon_{ijk}-\varepsilon_{ij\bullet})^2\Big] = \sigma^2 ab(c-1)\,.
$$

Pertanto:

$$
\mathbb{E}\Big[\sum_{ijk}{x_{\bullet\bullet\bullet}}^2\Big] = \sigma^2+\sum_{ijk}\mu_0^2
$$

$$
\begin{aligned}
&\mathbb{E}\left[\frac{1}{a-1}\sum_{ijk}(x_{i\bullet\bullet}-x_{\bullet\bullet\bullet})^2\right] \,=\, \sigma^2+\frac{1}{a-1}\sum_{ijk}\alpha_i^2 \\
&\mathbb{E}\left[\frac{1}{b-1}\sum_{ijk}(x_{\bullet j\bullet}-x_{\bullet\bullet\bullet})^2\right] \,=\, \sigma^2+\frac{1}{b-1}\sum_{ijk}\beta_j^2 \\
&\mathbb{E}\left[\frac{1}{(a-1)(b-1)}\sum_{ijk}(x_{ij\bullet}-x_{i\bullet\bullet}-x_{\bullet j\bullet}+x_{\bullet\bullet\bullet})^2\right] \,=\, \sigma^2+\frac{1}{(a-1)(b-1)}\sum_{ijk}(\alpha\beta)_{ij}^2 \\
&\mathbb{E}\left[\frac{1}{ab(c-1)}\sum_{ijk}(x_{ijk}-x_{ij\bullet})^2\right] \,=\, \sigma^2\,.
\end{aligned}
\tag{28.17}
$$

Test delle ipotesi sull'intero set $(\alpha_1,\ldots,\alpha_a)$
Si vuole testare l'ipotesi nulla che tutti i coefficienti α_i siano nulli

$$H_0 \;:\; (\alpha_1,\ldots,\alpha_a) \,=\, 0$$

contro l'ipotesi alternativa che ciò non abbia luogo, con almeno un parametro α_i diverso da zero

$$H_1 \;:\; (\alpha_1,\ldots,\alpha_a) \neq 0\,.$$

È sufficiente osservare che la variabile casuale F_α segue una distribuzione di Fisher a $(a-1, ab(c-1)$ gradi di libertà, in modo che per $(\alpha_1,\ldots,\alpha_a)=0$ il quoziente

$$F_{\alpha=0} \,=\, \frac{ab(c-1)}{a-1}\,\frac{\displaystyle\sum_{ijk}(x_{i\bullet\bullet}-x_{\bullet\bullet\bullet})^2}{\displaystyle\sum_{ijk}(x_{ijk}-x_{ij\bullet})^2} \,=\, \frac{1}{a-1}\,\frac{1}{s^2}\sum_{ijk}(x_{i\bullet\bullet}-x_{\bullet\bullet\bullet})^2$$

obbedisce anch'esso ad una distribuzione di probabilità dello stesso tipo. In virtù delle relazioni (28.17), valori intorno ad 1 o inferiori della statistica $F_{\alpha=0}$ si potranno ritenere indicativi della correttezza di H_0, mentre per $F_{\alpha=0} \gg 1$ indurranno a preferire l'ipotesi alternativa H_1. La regione di rigetto di H_0 viene quindi definita come

$$\{F_{\alpha=0} > F_{[1-\alpha](a-1,ab(c-1))}\}$$

con livello di significatività α.

Test delle ipotesi sull'intero set $(\beta_1,\ldots,\beta_b)$
Si vuole ora sottoporre a verifica l'ipotesi che i parametri β_j siano tutti nulli

$$H_0 \;:\; (\beta_1,\ldots,\beta_b) \,=\, 0$$

contro l'ipotesi alternativa che almeno uno di essi sia diverso da zero

$$H_1 \;:\; (\beta_1,\ldots,\beta_b) \neq 0\,.$$

Anche in questo caso il test è reso possibile dall'introduzione di una opportuna variabile di Fisher, nella fattispecie la statistica F_β a $(b-1, ab(c-1))$ gradi di libertà. Per H_0 vera, anche la variabile casuale

$$F_{\beta=0} = \frac{ab(c-1)}{b-1} \frac{\displaystyle\sum_{ijk}(x_{\bullet j\bullet} - x_{\bullet\bullet\bullet})^2}{\displaystyle\sum_{ijk}(x_{ijk} - x_{ij\bullet})^2} = \frac{1}{b-1}\,\frac{1}{s^2}\,\sum_{ijk}(x_{\bullet j\bullet} - x_{\bullet\bullet\bullet})^2$$

segue la stessa distribuzione; ciò consente di ritenere H_0 accettabile per $F_{\beta=0} \lesssim 1$, mentre valori di $F_{\beta=0}$ più elevati inducono a rigettare l'ipotesi in favore dell'alternativa H_1 — si ricordino le equazioni (28.17). Si definisce pertanto la regione di rigetto H_0 come

$$\{F_{\beta=0} > F_{[1-\alpha](b-1,ab(c-1))}\}$$

ancora con livello di significatività α prefissato a piacere.

Test delle ipotesi sull'intero set $\{(\alpha\beta)_{ij},\ 1 \leq i \leq a,\ 1 \leq j \leq b\}$
Si sottopone a test l'ipotesi che non si abbia interazione fra i due fattori, ovvero che siano nulli tutti i termini di interazione $(\alpha\beta)_{ij}$

$$H_0 \ :\ (\alpha\beta)_{ij} = 0 \qquad \forall\, i = 1,\ldots,a \quad \forall\, j = 1,\ldots,b$$

contro l'ipotesi alternativa che ve ne sia almeno uno diverso da zero

$$H_1 \ :\ \exists\, i \in \{1,\ldots,a\}\,,\ j \in \{1,\ldots,b\} \ :\ (\alpha\beta)_{ij} \neq 0\,.$$

Poiché la variabile casuale

$$\begin{aligned} F_{\alpha\beta} &= \frac{ab(c-1)}{(a-1)(b-1)} \frac{\displaystyle\sum_{ijk}[x_{ij\bullet} - x_{i\bullet\bullet} - x_{\bullet j\bullet} + x_{\bullet\bullet\bullet} - (\alpha\beta)_{ij}]^2}{\displaystyle\sum_{ijk}(x_{ijk} - x_{ij\bullet})^2} = \\ &= \frac{1}{(a-1)(b-1)}\,\frac{1}{s^2}\,\sum_{ijk}[x_{ij\bullet} - x_{i\bullet\bullet} - x_{\bullet j\bullet} + x_{\bullet\bullet\bullet} - (\alpha\beta)_{ij}]^2 \end{aligned}$$

segue una distribuzione di Fisher a $((a-1)(b-1), ab(c-1))$ gradi di libertà, si può affermare che una distribuzione dello stesso tipo caratterizzerà anche la variabile casuale

$$\begin{aligned} F_{\alpha\beta=0} &= \frac{ab(c-1)}{(a-1)(b-1)} \frac{\displaystyle\sum_{ijk}(x_{ij\bullet} - x_{i\bullet\bullet} - x_{\bullet j\bullet + x_{\bullet\bullet\bullet}})^2}{\displaystyle\sum_{ijk}(x_{ijk} - x_{ij\bullet})^2} = \\ &= \frac{1}{(a-1)(b-1)}\,\frac{1}{s^2}\,\sum_{ijk}(x_{ij\bullet} - x_{i\bullet\bullet} - x_{\bullet j\bullet} + x_{\bullet\bullet\bullet})^2 \end{aligned}$$

qualora l'ipotesi nulla sia verificata. Questa si potrà allora ritenere accettabile per $F_{\alpha\beta=0} \lesssim 1$, mentre ad essa si dovrà preferire l'ipotesi alternativa H_1 nel caso risulti $F_{\alpha\beta=0} > 1$ — si faccia ancora riferimento alle relazioni (28.17). La regione di rigetto di H_0 viene perciò definita come

$$\{F_{\alpha\beta=0} > F_{[1-\alpha]((a-1)(b-1),ab(c-1))}\}$$

con livello di significatività $\alpha \in (0,1)$.

Teorema di Scheffé per i parametri α_i
Per ogni a-upla di coefficienti reali $(C_1, C_2, \ldots, C_a)$ e per ogni $\alpha \in (0,1)$, la diseguaglianza

$$\Big|\sum_i C_i(x_{i\bullet\bullet} - x_{\bullet\bullet\bullet} - \alpha_i)\Big|^2 \le F_{[\alpha](a-1,ab(c-1))}\,\frac{a-1}{bc}s^2\sum_h C_h^2$$

è verificata con probabilità α. Se poi $\sum_i C_i = 0$, allora la diseguaglianza precedente si riduce a

$$\Big|\sum_i C_i(x_{i\bullet\bullet} - \alpha_i)\Big|^2 \le F_{[\alpha](a-1,ab(c-1))}\,\frac{a-1}{bc}s^2\sum_h C_h^2. \tag{28.18}$$

Dalla definizione della variabile di chi-quadrato

$$\mathcal{X}^2{}_{a-1} = \frac{1}{\sigma^2}\sum_{ijk}(x_{i\bullet\bullet} - x_{\bullet\bullet\bullet} - \alpha_i)^2 = \frac{bc}{\sigma^2}\sum_i(x_{i\bullet\bullet} - x_{\bullet\bullet\bullet} - \alpha_i)^2$$

si ha infatti, applicando la diseguaglianza di Cauchy-Schwarz

$$\sum_h C_h^2\,\mathcal{X}^2{}_{a-1} = \frac{bc}{\sigma^2}\sum_h C_h^2\sum_i(x_{i\bullet\bullet} - x_{\bullet\bullet\bullet} - \alpha_i)^2 \ge \frac{bc}{\sigma^2}\Big|\sum_i C_i(x_{i\bullet\bullet} - x_{\bullet\bullet\bullet} - \alpha_i)\Big|^2$$

per cui risulta — si ricordi la definizione (28.16) della varianza non spiegata s^2 —

$$\begin{aligned}\sum_h C_h^2\,\frac{ab(c-1)}{a-1}\frac{\mathcal{X}^2{}_{a-1}}{\mathcal{X}^2{}_{ab(c-1)}} &\ge \frac{ab(c-1)}{a-1}\frac{1}{\mathcal{X}^2{}_{ab(c-1)}}\frac{bc}{\sigma^2}\Big|\sum_i C_i(x_{i\bullet\bullet} - x_{\bullet\bullet\bullet} - \alpha_i)\Big|^2 = \\ &= \frac{bc}{a-1}\frac{1}{s^2}\Big|\sum_i C_i(x_{i\bullet\bullet} - x_{\bullet\bullet\bullet} - \alpha_i)\Big|^2.\end{aligned}$$

D'altra parte si è già verificato che il quoziente

$$F_\alpha = \frac{ab(c-1)}{a-1}\frac{\mathcal{X}^2{}_{a-1}}{\mathcal{X}^2{}_{ab(c-1)}}$$

segue una distribuzione di Fisher a $(a-1, ab(c-1))$ gradi di libertà, in modo che la diseguaglianza

$$\frac{ab(c-1)}{a-1}\frac{\mathcal{X}^2{}_{a-1}}{\mathcal{X}^2{}_{ab(c-1)}} \le F_{[\alpha](a-1,ab(c-1))}$$

è verificata con probabilità $\alpha \in (0,1)$, la stessa probabilità con la quale

$$\sum_h C_h^2 \frac{ab(c-1)}{a-1} \frac{\mathcal{X}^2{}_{a-1}}{\mathcal{X}^2{}_{ab(c-1)}} \le F_{[\alpha](a-1,ab(c-1))} \sum_h C_h^2$$

e quindi

$$\left|\sum_i C_i(x_{i\bullet\bullet} - x_{\bullet\bullet\bullet} - \alpha_i)\right|^2 \le F_{[\alpha](a-1,ab(c-1))} \frac{a-1}{bc} s^2 \sum_h C_h^2$$

come affermato. Qualora si abbia $\sum_i C_i = 0$, vale inoltre

$$\sum_i C_i(x_{i\bullet\bullet} - x_{\bullet\bullet\bullet} - \alpha_i) = \sum_i C_i(x_{i\bullet\bullet} - \alpha_i) - \sum_i C_i x_{\bullet\bullet\bullet} = \sum_i C_i(x_{i\bullet\bullet} - \alpha_i)$$

in accordo con la (28.18).

Teorema di Scheffé per i parametri β_j
Per ogni b-upla di coefficienti reali $(C_1, C_2, \ldots, C_b)$ e per ogni $\alpha \in (0,1)$, la diseguaglianza

$$\left|\sum_j C_i(x_{\bullet j\bullet} - x_{\bullet\bullet\bullet} - \beta_j)\right|^2 \le F_{[\alpha](b-1,ab(c-1))} \frac{b-1}{ac} s^2 \sum_h C_h^2$$

è verificata con probabilità α. Per $\sum_j C_j = 0$, la stessa diseguaglianza diventa

$$\left|\sum_j C_i(x_{\bullet j\bullet} - \beta_j)\right|^2 \le F_{[\alpha](b-1,ab(c-1))} \frac{b-1}{ac} s^2 \sum_h C_h^2 . \tag{28.19}$$

La prova del risultato è analoga a quella precedente. Si tratta soltanto di considerare la variabile di chi-quadrato

$$\mathcal{X}^2{}_{b-1} = \frac{1}{\sigma^2} \sum_{ijk} (x_{\bullet j\bullet} - x_{\bullet\bullet\bullet} - \beta_j)^2 = \frac{ac}{\sigma^2} \sum_j (x_{\bullet j\bullet} - x_{\bullet\bullet\bullet} - \beta_j)^2$$

moltiplicarla per la somma dei quadrati dei coefficienti ed applicare la diseguaglianza di Cauchy-Schwarz

$$\sum_h C_h^2 \mathcal{X}^2{}_{b-1} = \frac{ac}{\sigma^2} \sum_h C_h^2 \sum_j (x_{\bullet j\bullet} - x_{\bullet\bullet\bullet} - \beta_j)^2 \ge \frac{ac}{\sigma^2} \left|\sum_j C_j(x_{\bullet j\bullet} - x_{\bullet\bullet\bullet} - \beta_j)\right|^2$$

per ottenere

$$\sum_h C_h^2 \frac{ab(c-1)}{b-1} \frac{\mathcal{X}^2{}_{b-1}}{\mathcal{X}^2{}_{ab(c-1)}} \ge \frac{ab(c-1)}{b-1} \frac{1}{\mathcal{X}^2{}_{ab(c-1)}} \frac{ac}{\sigma^2} \left|\sum_j C_j(x_{\bullet j\bullet} - x_{\bullet\bullet\bullet} - \beta_j)\right|^2 =$$

$$= \frac{ac}{b-1} \frac{1}{s^2} \left|\sum_j C_j(x_{\bullet j\bullet} - x_{\bullet\bullet\bullet} - \beta_j)\right|^2 .$$

Poiché il quoziente

$$F_\beta = \frac{ab(c-1)}{b-1}\frac{\mathcal{X}^2{}_{b-1}}{\mathcal{X}^2{}_{ab(c-1)}}$$

costituisce una variabile di Fisher a $(b-1, ab(c-1))$ gradi di libertà, le diseguaglianze equivalenti

$$\frac{ab(c-1)}{b-1}\frac{\mathcal{X}^2{}_{b-1}}{\mathcal{X}^2{}_{ab(c-1)}} \le F_{[\alpha](b-1,ab(c-1))}$$

$$\sum_h C_h^2 \frac{ab(c-1)}{b-1}\frac{\mathcal{X}^2{}_{b-1}}{\mathcal{X}^2{}_{ab(c-1)}} \le F_{[\alpha](b-1,ab(c-1))} \sum_h C_h^2$$

sono verificate con probabilità $\alpha \in (0,1)$, la stessa con la quale vale

$$\left|\sum_j C_j(x_{\bullet j\bullet} - x_{\bullet\bullet\bullet} - \beta_j)\right|^2 \le F_{[\alpha](b-1,ab(c-1))} \frac{b-1}{ac} s^2 \sum_h C_h^2 \,.$$

La disequazione (28.19) per $\sum_j C_j = 0$ si ricava come nel caso precedente

$$\sum_j C_j(x_{\bullet j\bullet} - x_{\bullet\bullet\bullet} - \beta_j) = \sum_j C_j(x_{\bullet j\bullet} - \beta_j) - \sum_j C_j x_{\bullet\bullet\bullet} = \sum_j C_j(x_{\bullet j\bullet} - \beta_j)\,.$$

Teorema di Scheffé per i parametri $(\alpha\beta)_{ij}$
Per ogni (ab)*-upla di coefficienti reali* $\{C_{ij},\, i = 1,\ldots,a,\, j = 1,\ldots,b\}$ *e per ogni* $\alpha \in (0,1)$*, la diseguaglianza*

$$\left|\sum_{ij} C_{ij}[x_{ij\bullet} - x_{i\bullet\bullet} - x_{\bullet j\bullet} + x_{\bullet\bullet\bullet} - (\alpha\beta)_{ij}]\right|^2 \le F_{[\alpha]((a-1)(b-1),ab(c-1))} \frac{(a-1)(b-1)}{c} s^2 \sum_{hl} C_{hl}^2$$

è verificata con probabilità α*. Se inoltre i coefficienti* C_{ij} *sono assegnati in modo che si abbia*

$$\sum_i C_{ij} = 0 \quad \forall j \qquad \sum_j C_{ij} = 0 \quad \forall i$$

la disequazione precedente assume la forma

$$\left|\sum_{ij} C_{ij}[x_{ij\bullet} - (\alpha\beta)_{ij}]\right|^2 \le F_{[\alpha]((a-1)(b-1),ab(c-1))} \frac{(a-1)(b-1)}{c} s^2 \sum_{hl} C_{hl}^2$$

Come prima, la definizione della variabile $\mathcal{X}^2{}_{(a-1)(b-1)}$

$$\begin{aligned}\mathcal{X}^2{}_{(a-1)(b-1)} &= \frac{1}{\sigma^2}\sum_{ijk}[x_{ij\bullet} - x_{i\bullet\bullet} - x_{\bullet j\bullet} + x_{\bullet\bullet\bullet} - (\alpha\beta)_{ij}]^2 = \\ &= \frac{c}{\sigma^2}\sum_{ij}[x_{ij\bullet} - x_{i\bullet\bullet} - x_{\bullet j\bullet} + x_{\bullet\bullet\bullet} - (\alpha\beta)_{ij}]^2\end{aligned}$$

e la diseguaglianza di Cauchy-Schwarz porgono

$$\begin{aligned}\sum_{hl} C_{hl}^2\, \mathcal{X}^2{}_{(a-1)(b-1)} &= \frac{c}{\sigma^2}\sum_{hl} C_{hl}^2 \sum_{ij}[x_{ij\bullet} - x_{i\bullet\bullet} - x_{\bullet j\bullet} + x_{\bullet\bullet\bullet} - (\alpha\beta)_{ij}]^2 \geq \\ &\geq \frac{c}{\sigma^2}\left|\sum_{ij} C_{ij}[x_{ij\bullet} - x_{i\bullet\bullet} - x_{\bullet j\bullet} + x_{\bullet\bullet\bullet} - (\alpha\beta)_{ij}]\right|^2\end{aligned}$$

sicché

$$\begin{aligned}&\sum_{hl} C_{hl}^2\, \frac{ab(c-1)}{(a-1)(b-1)}\frac{\mathcal{X}^2{}_{(a-1)(b-1)}}{\mathcal{X}^2{}_{ab(c-1)}} \geq \\ &\geq \frac{c}{\sigma^2}\,\frac{1}{\mathcal{X}^2{}_{ab(c-1)}}\,\frac{ab(c-1)}{(a-1)(b-1)}\left|\sum_{ij} C_{ij}[x_{ij\bullet} - x_{i\bullet\bullet} - x_{\bullet j\bullet} + x_{\bullet\bullet\bullet} - (\alpha\beta)_{ij}]\right|^2 = \\ &= \frac{c}{(a-1)(b-1)}\,\frac{1}{s^2}\left|\sum_{ij} C_{ij}[x_{ij\bullet} - x_{i\bullet\bullet} - x_{\bullet j\bullet} + x_{\bullet\bullet\bullet} - (\alpha\beta)_{ij}]\right|^2 .\end{aligned}$$

La variabile di Fisher a $((a-1)(b-1), ab(c-1))$ gradi di libertà

$$F_{\alpha\beta} = \frac{ab(c-1)}{(a-1)(b-1)}\frac{\mathcal{X}^2{}_{(a-1)(b-1)}}{\mathcal{X}^2{}_{ab(c-1)}}$$

assicura che le diseguaglianze equivalenti

$$\frac{ab(c-1)}{(a-1)(b-1)}\frac{\mathcal{X}^2{}_{(a-1)(b-1)}}{\mathcal{X}^2{}_{ab(c-1)}} \leq F_{[\alpha]((a-1)(b-1),ab(c-1))}$$

$$\sum_{hl} C_{hl}^2\, \frac{ab(c-1)}{(a-1)(b-1)}\frac{\mathcal{X}^2{}_{(a-1)(b-1)}}{\mathcal{X}^2{}_{ab(c-1)}} \leq F_{[\alpha]((a-1)(b-1),ab(c-1))} \sum_{hl} C_{hl}^2$$

siano soddisfatte con probabilità α, al pari di

$$\left|\sum_{ij} C_{ij}[x_{ij\bullet} - x_{i\bullet\bullet} - x_{\bullet j\bullet} + x_{\bullet\bullet\bullet} - (\alpha\beta)_{ij}]\right|^2 \leq F_{[\alpha]((a-1)(b-1),ab(c-1))}\,\frac{(a-1)(b-1)}{c} s^2 \sum_{hl} C_{hl}^2$$

Il caso di $\sum_i C_{ij} = \sum_j C_{ij} = 0$ è immediato, in quanto

$$\begin{aligned}&\sum_{ij} C_{ij}[x_{ij\bullet} - x_{i\bullet\bullet} - x_{\bullet j\bullet} + x_{\bullet\bullet\bullet} - (\alpha\beta)_{ij}] = \\ &= \sum_{ij} C_{ij}[x_{ij\bullet} - (\alpha\beta)_{ij}] + \sum_{ij} C_{ij}[-x_{i\bullet\bullet} - x_{\bullet j\bullet} + x_{\bullet\bullet\bullet}] = \sum_{ij} C_{ij}[x_{ij\bullet} - (\alpha\beta)_{ij}] .\end{aligned}$$

Intervalli di confidenza (simultanei) per le differenze $\alpha_{i_1} - \alpha_{i_2}$
Una importante applicazione del teorema di Scheffé è la stima della differenza di due parametri α_i qualsiasi, nel qual caso tutti i pesi C_h sono nulli salvo due, rispettivamente eguali a 1 e -1

$$|x_{i_1\bullet\bullet} - x_{i_2\bullet\bullet} - (\alpha_{i_1} - \alpha_{i_2})|^2 \,=\, |x_{i_1\bullet\bullet} - \alpha_{i_1} - (x_{i_2\bullet\bullet} - \alpha_{i_2})|^2 \leq F_{[\alpha](a-1,ab(c-1))}\, s^2\, 2\frac{a-1}{bc}$$

e quindi con livello di confidenza α **tutte** le differenze $\alpha_{i_1} - \alpha_{i_2}$ sono comprese entro gli intervalli di confidenza

$$x_{i_1\bullet\bullet} - x_{i_2\bullet\bullet} \pm \sqrt{F_{[\alpha](a-1,ab(c-1))}\, s^2\, 2\frac{a-1}{bc}}\,.$$

Intervalli di confidenza (simultanei) per le differenze $\beta_{j_1} - \beta_{j_2}$
Si ricavano in modo analogo a quanto visto nel punto precedente. Si ha in effetti, con probabilità $\alpha \in (0,1)$ prefissata,

$$|x_{\bullet j_1\bullet} - x_{\bullet j_2\bullet} - (\beta_{j_1} - \beta_{j_2})|^2 \,=\, |x_{\bullet j_1\bullet} - \beta_{j_1} - (x_{\bullet j_2\bullet} - \beta_{j_2})|^2 \leq F_{[\alpha](b-1,ab(c-1))}\, s^2\, 2\frac{b-1}{ac}$$

e pertanto l'intervallo di confidenza di $\beta_{j_1} - \beta_{j_2}$ è dato da

$$x_{\bullet j_1\bullet} - x_{\bullet j_2\bullet} \pm \sqrt{F_{[\alpha](b-1,ab(c-1))}\, s^2\, 2\frac{b-1}{ac}}\,.$$

con livello di confidenza α.

ANOVA a due fattori senza interazione
Omissis

ANOVA a un fattore
Omissis

29. Definizione. Regressione lineare con il metodo del chi quadrato

Sia

$$\big\{(x_i, y_i) \in \mathbb{R}^2\,,\ 1 \leq i \leq n\big\}$$

un insieme di dati soddisfacente le seguenti ipotesi:

(i) per ogni $i = 1, \ldots, n$ il dato y_i è la realizzazione di una variabile casuale gaussiana di varianza nota σ_i^2. Il valore medio di tale variabile casuale è a priori incognito, ma si assume dipendere esclusivamente dall'ascissa x_i — di cui sarà funzione;

(ii) l'insieme dei dati è descritto dal modello matematico

$$\zeta_i \,=\, \sum_{j=1}^{p} \alpha_j \phi_j(x_i) \,+\, \varepsilon_i \quad , \qquad 1 \leq i \leq n\,, \tag{29.1}$$

nel quale le $\phi_j : \mathbb{R} \longrightarrow \mathbb{R}$, $1 \leq j \leq p$, sono $p < n$ funzioni preassegnate, nessuna costantemente nulla, e gli $\alpha_1, \ldots, \alpha_p$ dei parametri costanti da stimare convenientemente, mentre le ε_i costituiscono un insieme di n variabili gaussiane indipendenti, di media nulla e varianza σ_i^2:

$$\mathbb{E}(\varepsilon_i) = 0 \qquad \mathbb{E}(\varepsilon_i \varepsilon_k) = \delta_{ik}\sigma_k^2 \quad 1 \leq i, k \leq n \ .$$

Il valore y_i deve dunque intendersi come una realizzazione della variabile casuale ζ_i, per ogni i.

Si definisce regressione lineare del set di dati, con il metodo del chi quadrato, l'espressione

$$R(x) = \sum_{j=1}^{p} a_j \phi_j(x) \quad , \qquad x \in \mathbb{R} \ ,$$

nella quale gli a_j sono le stime dei parametri costanti α_j ottenute per mezzo del campione $\{(x_i, y_i)\,,\ 1 \leq i \leq n\}$ minimizzando il funzionale quadratico $\mathfrak{L} : \mathbb{R}^p \longrightarrow \mathbb{R}$ definito da

$$\mathfrak{L}(\xi_1, \ldots, \xi_p) = \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \left[-y_i + \sum_{j=1}^{p} \xi_j \phi_j(x_i) \right]^2 . \tag{29.2}$$

Nell'ipotesi che tutte le varianze σ_i^2 siano eguali, specificarne il comune valore diventa irrilevante ai fini della ottimizzazione del funzionale $\mathfrak{L}$, che si può perciò ridurre ad una sommatoria dei soli quadrati degli scarti. In tal caso la procedura di regressione lineare prende il nome di **metodo dei minimi quadrati**.

30. Equazione normale e sua soluzione

I punti critici del funzionale quadratico (29.2) *sono tutte e sole le soluzioni dell'equazione lineare — cosiddetta equazione normale —*

$$Fa = \Phi \sigma^{-1} y \ , \tag{30.1}$$

nella quale le matrici F e Φ sono definite da

$$F_{jk} = \sum_{i=1}^{n} \frac{\phi_j(x_i)\phi_k(x_i)}{\sigma_i^2} \qquad \Phi_{ki} = \frac{\phi_k(x_i)}{\sigma_i} \ , \quad 1 \leq j, k \leq p\,,\ 1 \leq i \leq n\,, \tag{30.2}$$

mentre

$$a = \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix} \qquad \sigma^{-1} = \begin{pmatrix} 1/\sigma_1 & & & \\ & 1/\sigma_2 & & \mathbb{O} \\ & & \ddots & \\ & \mathbb{O} & & 1/\sigma_n \end{pmatrix} \qquad y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} . \tag{30.3}$$

Vale in particolare che $F = \Phi\Phi^T$, matrice reale simmetrica semidefinita positiva. Nell'ipotesi che la matrice $p \times n$ Φ abbia rango massimo p, allora F risulta definita positiva e l'unica soluzione dell'equazione normale (30.1) *è data da*

$$a \,=\, F^{-1}\Phi\sigma^{-1}y \;. \tag{30.4}$$

Dimostrazione
I punti critici del funzionale $\mathfrak{L}$ sono per definizione le soluzioni del sistema di equazioni

$$\frac{\partial\mathfrak{L}}{\partial\xi_k}(a_1,\ldots,a_p) \,=\, \sum_{i=1}^n \frac{1}{\sigma_i^2}\,2\left[-y_i+\sum_{j=1}^p a_j\phi_j(x_i)\right]\phi_k(x_i) \,=\, 0\;, \quad 1\le k\le p\,,$$

ossia

$$\sum_{j=1}^p a_j \sum_{i=1}^n \frac{\phi_j(x_i)\phi_k(x_i)}{\sigma_i^2} \,=\, \sum_{i=1}^n \frac{\phi_k(x_i)}{\sigma_i}\,\frac{1}{\sigma_i}\,y_i\;, \quad 1\le k\le p\,,$$

e quindi, introducendo le matrici (30.2)-(30.3),

$$(Fa)_k \,=\, (\Phi\sigma^{-1}y)_k\;, \quad 1\le k\le p\,, \qquad \Longleftrightarrow \qquad Fa \,=\, \Phi\sigma^{-1}y\;.$$

La matrice F è sempre reale simmetrica e semidefinita positiva. Sul fatto che si tratti di matrice reale non sussiste dubbio, mentre la simmetria è banale:

$$F_{jk} \,=\, \sum_{i=1}^n \frac{\phi_j(x_i)\phi_k(x_i)}{\sigma_i^2} \,=\, \sum_{i=1}^n \frac{\phi_k(x_i)\phi_j(x_i)}{\sigma_i^2} \,=\, F_{kj}\;, \quad 1\le j,k\le p\;.$$

Il carattere semidefinito positivo segue anch'esso piuttosto facilmente, notando che per ogni vettore $t\in\mathbb{R}^p$ deve aversi:

$$\begin{aligned} t^TFt \,&=\, \sum_{j,k=1}^p t_jt_kF_{jk} \,=\, \sum_{j,k=1}^p t_jt_k\sum_{i=1}^n \frac{\phi_j(x_i)\phi_k(x_i)}{\sigma_i^2} \,= \\ &=\, \sum_{i=1}^n\frac{1}{\sigma_i^2}\sum_{j=1}^p t_j\phi_j(x_i)\sum_{k=1}^p t_k\phi_k(x_i) \,=\, \sum_{i=1}^n\frac{1}{\sigma_i^2}\left[\sum_{j=1}^p t_j\phi_j(x_i)\right]^2 \,\ge\, 0\,. \end{aligned}$$

Inoltre $t^TFt = 0$ se e soltanto se $\displaystyle\sum_{i=1}^n\frac{1}{\sigma_i^2}\left[\sum_{j=1}^p t_j\phi_j(x_i)\right]^2 = 0$ e quindi

$$\sum_{j=1}^p t_j\phi_j(x_i) \,=\, 0\;, \quad 1\le i\le n\,, \qquad \Longleftrightarrow \qquad \sum_{j=1}^p t_j\phi_{ji} \,=\, 0\;, \quad 1\le i\le n\,,$$

condizione che può essere soddisfatta per un $t\neq 0$ nel solo caso che il rango — riga — di Φ sia minore di quello massimo p. Ciò comporta in particolare che l'ipotesi $\mathrm{rank}(\Phi) = p$

implica $t^T Ft = 0$ se e soltanto se $t = 0$, pertanto F è definita positiva e di conseguenza invertibile. La (30.4) segue allora dalla (30.1) moltiplicando a sinistra membro a membro per F^{-1}. Quanto all'identità $F = \Phi\Phi^T$, per verificarla bastano pochi passaggi algebrici:

$$F_{jk} = \sum_{i=1}^{n} \frac{\phi_j(x_i)\phi_k(x_i)}{\sigma_i^2} = \sum_{i=1}^{n} \frac{\phi_j(x_i)}{\sigma_i}\,\frac{\phi_k(x_i)}{\sigma_i} = \sum_{i=1}^{n} \Phi_{ji}\Phi_{ki} = \sum_{i=1}^{n} \Phi_{ji}(\Phi^T)_{ik} = (\Phi\Phi^T)_{jk} \,,$$

essendo $1 \leq j,k \leq p$. $\square$

31. I parametri di regressione lineare, con il metodo del chi quadrato, come variabili casuali. Media e matrice di covarianza

Se Φ ha rango massimo p, la stima (30.4) dei parametri di regressione costituisce un insieme di p variabili gaussiane con valori medi

$$\mathbb{E}(a_j) \,=\, \alpha_j \;, \quad 1 \leq j \leq p\,,$$

e matrice di covarianza

$$C_a \,=\, \mathbb{E}\Big[\big[a - \mathbb{E}(a)\big]\big[a - \mathbb{E}(a)\big]^T\Big] \,=\, F^{-1} \;. \tag{31.1}$$

I valori stimati dei parametri di regressione sono stocasticamente indipendenti se e soltanto se i vettori colonna $\Phi_1, \ldots, \Phi_p \in \mathbb{R}^n$ definiti da

$$\Phi_j \in \mathbb{R}^n \;:\quad (\Phi_j)_i \,=\, \Phi_{ji} \,=\, \frac{1}{\sigma_i}\phi_j(x_i) \quad 1 \leq j \leq p \quad 1 \leq i \leq n$$

costituiscono un sistema ortogonale in $\mathbb{R}^n$

$$\langle \Phi_j | \Phi_k \rangle \stackrel{\text{def}}{=} \sum_{i=1}^{n} (\Phi_j)_i (\Phi_k)_i \,=\, 0 \qquad \forall\, j \neq k\,.$$

Dimostrazione

Dalla relazione $a = F^{-1}\Phi\sigma^{-1}y$ segue immediatamente che le a_j risultano da combinazioni lineari delle variabili gaussiane y_i e sono quindi variabili gaussiane a propria volta. Prendendo poi il valor medio di ambo i membri di $a = F^{-1}\Phi\sigma^{-1}y$ si deduce che $\mathbb{E}(a) = F^{-1}\Phi\sigma^{-1}\mathbb{E}(y)$ e dunque

$$\begin{aligned} \mathbb{E}(a_j) \,&=\, \sum_{i=1}^{n} (F^{-1}\Phi)_{ji} \frac{1}{\sigma_i}\mathbb{E}(y_i) \,=\, \sum_{i=1}^{n} (F^{-1}\Phi)_{ji} \frac{1}{\sigma_i} \sum_{j'=1}^{p} \alpha_{j'}\Phi_{j'}(x_i) \,= \\ &=\, \sum_{i=1}^{n} (F^{-1}\Phi)_{ji} \sum_{j'=1}^{p} \alpha_{j'}(\Phi^T)_{ij'} \,=\, (F^{-1}\Phi\Phi^T\alpha)_j \,=\, \alpha_j \;, \quad 1 \leq j \leq p\,, \end{aligned}$$

in cui ovviamente $\alpha^T = (\alpha_1 \ldots \alpha_p)$ e si è fatto uso delle relazioni $F = \Phi\Phi^T$ ed $\mathbb{E}(y_i) = \mathbb{E}\big[\sum_{j=1}^{p} \alpha_j\phi_j(x_i) + \varepsilon_i\big] = \sum_{j=1}^{p} \alpha_j\phi_j(x_i)$, $1 \leq i \leq n$. Per il calcolo della matrice di covarianza si osservi preliminarmente che

$$a - \mathbb{E}(a) \;=\; F^{-1}\Phi\sigma^{-1}y - \mathbb{E}(F^{-1}\Phi\sigma^{-1}y) \;=\; F^{-1}\Phi\big[\sigma^{-1}y - \mathbb{E}(\sigma^{-1}y)\big]$$

per cui

$$\big[a-\mathbb{E}(a)\big]\big[a-\mathbb{E}(a)\big]^T \;=\; F^{-1}\Phi\big[\sigma^{-1}y-\mathbb{E}(\sigma^{-1}y)\big]\,\big[\sigma^{-1}y-\mathbb{E}(\sigma^{-1}y)\big]^T\Phi^T(F^{-1})^T$$

e quindi

$$C_a \;=\; \mathbb{E}\Big[\big[a-\mathbb{E}(a)\big]\big[a-\mathbb{E}(a)\big]^T\Big] \;= \tag{31.2}$$

$$= F^{-1}\Phi\,\mathbb{E}\Big[\big[\sigma^{-1}y-\mathbb{E}(\sigma^{-1}y)\big]\big[\sigma^{-1}y-\mathbb{E}(\sigma^{-1}y)\big]^T\Big]\Phi^T(F^{-1})^T \;=\; F^{-1}\Phi\,C_{\sigma^{-1}y}\Phi^T(F^{-1})^T\,;$$

la matrice di covarianza $C_{\sigma^{-1}y}$ delle variabili $\sigma^{-1}y$ si determina agevolmente in forza delle ipotesi assunte nella definizione di regressione, ed in particolare per tramite della (29.1):

$$(C_{\sigma^{-1}y})_{ik} = \mathbb{E}\bigg[\Big[\frac{y_i}{\sigma_i}-\mathbb{E}\Big(\frac{y_i}{\sigma_i}\Big)\Big]\Big[\frac{y_k}{\sigma_k}-\mathbb{E}\Big(\frac{y_k}{\sigma_k}\Big)\Big]\bigg]\;,$$

osservato che

$$\frac{y_i}{\sigma_i}-\mathbb{E}\Big(\frac{y_i}{\sigma_i}\Big) \;=\; \sum_{j=1}^{p}\alpha_j\frac{\phi_j(x_i)}{\sigma_i}+\frac{\varepsilon_i}{\sigma_i}-\mathbb{E}\Big[\sum_{j=1}^{p}\alpha_j\frac{\phi_j(x_i)}{\sigma_i}+\frac{\varepsilon_i}{\sigma_i}\Big] \;=\; \frac{\varepsilon_i}{\sigma_i}-\mathbb{E}\Big[\frac{\varepsilon_i}{\sigma_i}\Big] \;=\; \frac{\varepsilon_i}{\sigma_i}$$

per cui

$$(C_{\sigma^{-1}y})_{ik} \;=\; \mathbb{E}\Big[\frac{\varepsilon_i}{\sigma_i}\frac{\varepsilon_k}{\sigma_k}\Big] \;=\; \frac{1}{\sigma_i\sigma_k}\mathbb{E}(\varepsilon_i\varepsilon_k) \;=\; \frac{1}{\sigma_i\sigma_k}\sigma_i^2\delta_{ik} \;=\; \delta_{ik}\;,\quad 1\le i,k\le n\,.$$

Inserendo in (31.2) la relazione $C_{\sigma^{-1}y}=\mathbb{I}$ appena ottenuta, si perviene al risultato richiesto — equazione (31.1):

$$C_a \;=\; F^{-1}\Phi\,C_{\sigma^{-1}y}\Phi^T(F^{-1})^T \;=\; F^{-1}\Phi\Phi^T(F^{-1})^T \;=\; (F^{-1})^T \;=\; F^{-1}\,.$$

Trattandosi di variabili gaussiane, l'indipendenza stocastica dei parametri stimati a_j è equivalente alla mancanza di correlazione, ossia alla struttura diagonale della matrice di covarianza $C_a = F^{-1}$ appena determinata. D'altra parte, la matrice F^{-1} è diagonale se e soltanto se lo è F, e siccome

$$F_{jk} \;=\; \sum_{i=1}^{n}\Phi_{ji}(\Phi)^T_{ik} \;=\; \sum_{i=1}^{n}\Phi_{ji}(\Phi)_{ki} \;=\; \sum_{i=1}^{n}(\Phi_j)_i(\Phi_k)_i \;=\; \langle\Phi_j|\Phi_k\rangle$$

si conclude che la struttura diagonale della F ricorre se e soltanto se il sistema dei vettori $\Phi_1,\ldots,\Phi_p$ è ortogonale in $\mathbb{R}^n$. $\square$

Osservazione

Per il teorema di Gram-Schmidt è sempre possibile introdurre opportune combinazioni lineari dei vettori non nulli $\Phi_1,\ldots,\Phi_p$ in modo da definire un sistema ortogonale. Ne consegue che è sempre dato di sostituire ai parametri di regressione $\alpha_1,\ldots,\alpha_p$ nuovi parametri $\beta_1,\ldots,\beta_p$ definiti da appropriate combinazioni lineari degli $\alpha_1,\ldots,\alpha_p$ in modo che le stime di chi-quadrato dei $\beta_1,\ldots,\beta_p$ risultino variabili gaussiane stocasticamente indipendenti.

32. Somma dei quadrati degli scarti attorno alla regressione

Nell'ipotesi che Φ abbia rango massimo $p < n$, la somma normalizzata dei quadrati degli scarti attorno alla regressione

$$\text{NSSAR} = \sum_{i=1}^{n} \frac{1}{\sigma_i^2}\Big[-y_i + \sum_{j=1}^{p}\phi_j(x_i)a_j\Big]^2 \tag{32.1}$$

costituisce una variabile di $\mathcal{X}^2$ a $n-p$ gradi di libertà.

Dimostrazione

Si osserva anzitutto che

$$\mathbb{E}(y_i) = \mathbb{E}\Big[\sum_{j=1}^{p}\alpha_j\phi_j(x_i) + \varepsilon_i\Big] = \sum_{j=1}^{p}\alpha_j\phi_j(x_i) + \mathbb{E}(\varepsilon_i) = \sum_{j=1}^{p}\alpha_j\phi_j(x_i)$$

mentre

$$\mathbb{E}\Big[\sum_{j=1}^{p}a_j\phi_j(x_i)\Big] = \sum_{j=1}^{p}\mathbb{E}(a_j)\,\phi_j(x_i) = \sum_{j=1}^{p}\alpha_j\,\phi_j(x_i)$$

in modo che le quantità entro parentesi quadre nella (32.1) possono riscriversi come somme di — realizzazioni di — variabili casuali a media nulla:

$$-y_i + \sum_{j=1}^{p}\phi_j(x_i)a_j = -\Big[y_i - \sum_{j=1}^{p}\alpha_j\,\phi_j(x_i)\Big] + \Big[\sum_{j=1}^{p}\phi_j(x_i)a_j - \sum_{j=1}^{p}\alpha_j\,\phi_j(x_i)\Big]$$

e ricordando che $a = F^{-1}\Phi\sigma^{-1}y$ ed $\alpha = \mathbb{E}(a) = \mathbb{E}(F^{-1}\Phi\sigma^{-1}y) = F^{-1}\Phi\,\mathbb{E}(\sigma^{-1}y)$:

$$\frac{1}{\sigma_i}\Big[-y_i + \sum_{j=1}^{p}\phi_j(x_i)a_j\Big] = -\Big[\frac{y_i}{\sigma_i} - \frac{\mathbb{E}(y_i)}{\sigma_i}\Big] + \sum_{j=1}^{p}\frac{\phi_j(x_i)}{\sigma_i}(a_j - \alpha_j) =$$

$$\begin{aligned}
&= -\Big[\frac{y_i}{\sigma_i} - \frac{\mathbb{E}(y_i)}{\sigma_i}\Big] + \sum_{j=1}^{p}\Phi_{ji}(a_j - \alpha_j) = \big\{-(\sigma^{-1}y)^T + \mathbb{E}[(\sigma^{-1}y)^T] + (a-\alpha)^T\Phi\big\}_i = \\
&= \big\{-(\sigma^{-1}y)^T + \mathbb{E}[(\sigma^{-1}y)^T] + [F^{-1}\Phi(\sigma^{-1}y - \mathbb{E}(\sigma^{-1}y))]^T\Phi\big\}_i = \\
&= \big\{-(\sigma^{-1}y)^T + \mathbb{E}[(\sigma^{-1}y)^T] + \big[(\sigma^{-1}y)^T - \mathbb{E}[(\sigma^{-1}y)^T]\big]\Phi^T(F^{-1})^T\Phi\big\}_i .
\end{aligned}$$

Posto per brevità $z = \sigma^{-1}y - \mathbb{E}(\sigma^{-1}y)$, la (32.1) diventa allora

$$\begin{aligned}
\text{NSSAR} = \sum_{i=1}^{n}\Big[\frac{y_i}{\sigma_i} - \sum_{j=1}^{p}\frac{\phi_j(x_i)}{\sigma_i}a_j\Big]^2 &= \sum_{i=1}^{n}\big\{z^T - z^T\Phi^T(F^{-1})^T\Phi\big\}_i^2 = \\
&= \big[z^T - z^T\Phi^T(F^{-1})^T\Phi\big]\big[z^T - z^T\Phi^T(F^{-1})^T\Phi\big]^T = \\
&= \big[z^T - z^T\Phi^T(F^{-1})^T\Phi\big]\big[z - \Phi^TF^{-1}\Phi z\big] = \\
&= z^T\big[\mathbb{I} - \Phi^T(F^{-1})^T\Phi\big]\big[\mathbb{I} - \Phi^TF^{-1}\Phi\big]\,z = z^TBz
\end{aligned} \tag{32.2}$$

ed è quindi riconoscibile come una forma quadratica semidefinita positiva di z, con matrice rappresentativa reale e simmetrica

$$\begin{aligned} B &= \big[\mathbb{I} - \Phi^T (F^{-1})^T \Phi\big]\big[\mathbb{I} - \Phi^T F^{-1}\Phi\big] = \\ &= \mathbb{I} - \Phi^T (F^{-1})^T \Phi - \Phi^T F^{-1}\Phi + \Phi^T (F^{-1})^T \Phi\Phi^T F^{-1}\Phi = \mathbb{I} - \Phi^T F^{-1}\Phi \end{aligned} \tag{32.3}$$

in cui si è fatto uso della relazione $\Phi\Phi^T = F$. Il vettore z costituisce un sistema di n variabili gaussiane indipendenti a media nulla e con varianza unitaria:

$$\mathbb{E}(z_i z_k) = \mathbb{E}\Big(\frac{\varepsilon_i}{\sigma_i}\frac{\varepsilon_k}{\sigma_k}\Big) = \frac{1}{\sigma_i\sigma_k}\mathbb{E}(\varepsilon_i\varepsilon_k) = \frac{1}{\sigma_i\sigma_k}\sigma_i^2\delta_{ik} = \delta_{ik}\ , \quad 1 \le i,k \le n\,. \tag{32.4}$$

Si tratta ora di dimostrare che per mezzo di una appropriata trasformazione lineare ortogonale, la matrice B può ricondursi alla forma diagonale con $n-p$ elementi diagonali uguali ad 1 ed i restanti a 0. A questo scopo si osserva in primo luogo che B è una matrice idempotente:

$$B^2 = [\mathbb{I} - \Phi^T F^{-1}\Phi]\,[\mathbb{I} - \Phi^T F^{-1}\Phi] = \mathbb{I} - \Phi^T F^{-1}\Phi - \Phi^T F^{-1}\Phi + \Phi^T F^{-1}\Phi\Phi^T F^{-1}\Phi =$$

$$= \mathbb{I} - \Phi^T F^{-1}\Phi - \Phi^T F^{-1}\Phi + \Phi^T F^{-1}\Phi = \mathbb{I} - \Phi^T F^{-1}\Phi = B$$

per cui i suoi autovalori, certamente reali per via della simmetria, possono essere soltanto 0 ed 1. Infatti l'equazione $Bv = \lambda v$, con $v \in \mathbb{R}^n \setminus \{0\}$ e $\lambda \in \mathbb{R}$, implica

$$\lambda v = Bv = B^2 v = \lambda^2 v$$

e conseguentemente $\lambda(1-\lambda)v = 0$. Si verifica poi che 0 ed 1 sono entrambi effettivamente autovalori di B costruendo in modo esplicito i corrispondenti autospazi. Gli autovettori di B relativi all'autovalore 1 soddisfano la condizione

$$Bv = v\ , \quad v \neq 0\,,$$

ossia, in forza di (32.3),

$$[\mathbb{I} - \Phi^T F^{-1}\Phi]v = v\ , \quad v \neq 0\,,$$

e quindi

$$-\Phi^T F^{-1}\Phi v = 0\ , \quad v \neq 0\ .$$

Da questa relazione è chiaro che per ogni $v \in \ker\Phi$ vale $v \in \ker(B - \mathbb{I})$, sicché

$$\ker\Phi \subseteq \ker(B - \mathbb{I})\ ; \tag{32.5}$$

per contro, se $v \in \ker(B - \mathbb{I})$ risulta

$$-\Phi^T F^{-1}\Phi v = 0$$

e moltiplicando a sinistra ambo i membri per la matrice Φ si ottiene

$$-\Phi\Phi^T F^{-1}\Phi v = 0$$

da cui segue $\Phi v = 0$, ossia $v \in \ker\Phi$. Di qui l'inclusione

$$\ker(B - \mathbb{I}) \subseteq \ker\Phi$$

che unitamente alla (32.5) implica l'equazione

$$\ker(B - \mathbb{I}) \,=\, \ker\Phi \,.$$

Nell'ipotesi assunta che Φ abbia rango massimo $p < n$, si ha certamente che

$$\ker\Phi \,=\, \{v \in \mathbb{R}^n \,:\, \Phi v \,=\, 0\}$$

è un sottospazio vettoriale reale di dimensione $n - p < n$. Lo stesso dovrà perciò valere per $\ker(B - \mathbb{I})$:

$$\dim\ker(B - \mathbb{I}) \,=\, \dim\ker\Phi \,=\, n - p < n$$

e di conseguenza, data la diagonalizzabilità della matrice simmetrica B, anche 0 risulterà autovalore di B, con

$$\dim\ker B \,=\, n - \dim\ker(B - \mathbb{I}) \,=\, n - (n - p) \,=\, p$$

in quanto

$$\ker B \oplus \ker(B - \mathbb{I}) \,=\, \mathbb{R}^n \qquad \text{e} \qquad \ker B \,=\, [\ker(B - \mathbb{I})]^{\perp}$$

— si ricorda che gli autospazi di una matrice reale e simmetrica sono sempre ortogonali fra loro. Esisterà pertanto una matrice ortogonale R tale che

$$R^T B R \,=\, \begin{pmatrix} \boxed{\begin{matrix} \mathbb{I} \\ (n-p)\times(n-p) \end{matrix}} & \boxed{\begin{matrix} \mathbb{O} \\ (n-p)\times p \end{matrix}} \\ \boxed{\begin{matrix} \mathbb{O} \\ p\times(n-p) \end{matrix}} & \boxed{\begin{matrix} \mathbb{O} \\ p\times p \end{matrix}} \end{pmatrix} \,=\, D$$

e sarà quindi sufficiente introdurre in (32.2) la trasformazione ortogonale $z \,=\, Rw$ per concludere che

$$\text{NSSAR} \,=\, w^T R^T B R w \,=\, w^T D w \,=\, \sum_{i=1}^{n-p} w_i^2$$

con le w variabili gaussiane, quali combinazioni lineari delle variabili gaussiane z, stocasticamente indipendenti con media nulla e varianza unitaria:

$$\mathbb{E}(w) \,=\, \mathbb{E}(R^T z) \,=\, R^T \mathbb{E}(z) \,=\, R^T 0 \,=\, 0$$

$$\mathbb{E}(ww^T) = \mathbb{E}(R^T zz^T R) = R^T \mathbb{E}(zz^T) R = R^T \mathbb{I} R = R^T R = \mathbb{I}$$

in cui si è fatto ricorso alla (32.4). Risulta così provato che NSSAR è la somma dei quadrati di $n-p$ variabili gaussiane indipendenti a media nulla e varianza unitaria, cioè, precisamente, una variabile di $\mathcal{X}^2$ a $n-p$ gradi di libertà. $\square$

Osservazione sul calcolo della NSSAR
Poiché i punti critici del funzionale di chi-quadrato $\mathfrak{L}$ sono le soluzioni del sistema di equazioni algebriche

$$\frac{\partial \mathfrak{L}}{\partial \xi_h}(a_1, \ldots, a_p) = 2 \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \left[-y_i + \sum_{j=1}^{p} a_j \phi_j(x_i) \right] \phi_h(x_i) = 0 \qquad \forall\, h = 1, \ldots, p$$

si ha immediatamente che

$$\begin{aligned}
\text{NSSAR} &= \mathfrak{L}(a_1, \ldots, a_p) = \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \left[-y_i + \sum_{j=1}^{p} a_j \phi_j(x_i) \right]^2 = \\
&= \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \left[-y_i + \sum_{h=1}^{p} a_h \phi_h(x_i) \right] \left[-y_i + \sum_{j=1}^{p} a_j \phi_j(x_i) \right] = \\
&= \sum_{i=1}^{n} \frac{1}{\sigma_i^2} y_i \left[y_i - \sum_{j=1}^{p} a_j \phi_j(x_i) \right] + \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \sum_{h=1}^{p} a_h \phi_h(x_i) \left[-y_i + \sum_{j=1}^{p} a_j \phi_j(x_i) \right] = \\
&= \sum_{i=1}^{n} \frac{1}{\sigma_i^2} y_i \left[y_i - \sum_{j=1}^{p} a_j \phi_j(x_i) \right] + \sum_{h=1}^{p} a_h \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \left[-y_i + \sum_{j=1}^{p} a_j \phi_j(x_i) \right] \phi_h(x_i) = \\
&= \sum_{i=1}^{n} \frac{1}{\sigma_i^2} y_i \left[y_i - \sum_{j=1}^{p} a_j \phi_j(x_i) \right] + \sum_{h=1}^{p} a_h \, 0
\end{aligned}$$

per cui la somma normalizzata dei quadrati degli scarti attorno alla regressione assume la forma semplificata

$$\text{NSSAR} = \sum_{i=1}^{n} \frac{1}{\sigma_i^2} y_i \left[y_i - \sum_{j=1}^{p} a_j \phi_j(x_i) \right] \qquad (32.6)$$

valida sotto qualsiasi ipotesi circa il rango di Φ — che quindi può anche non essere massimo.

33. Indipendenza stocastica di NSSAR e parametri di regressione stimati a_j

Nell'ipotesi che la matrice Φ abbia rango massimo p, la somma normalizzata dei quadrati degli scarti attorno alla regressione NSSAR *è stocasticamente indipendente da ciascuna delle stime a_j, $j = 1, \ldots, p$.*

Dimostrazione
Provare l'indipendenza stocastica di NSSAR da a_j equivale a dimostrare che le variabili casuali

$$(a_j - \alpha_j)^2 \qquad \text{e} \qquad \text{NSSAR}$$

sono stocasticamente indipendenti. Il vantaggio offerto da questa strategia consiste nel fatto che le variabili casuali $a - \alpha$ sono combinazioni lineari

$$a - \alpha \,=\, a - \mathbb{E}(a) \,=\, F^{-1}\Phi\sigma^{-1}(y - \mathbb{E}(y)) \,=\, F^{-1}\Phi z \tag{33.1}$$

delle variabili normali indipendenti $z = \sigma^{-1}(y - \mathbb{E}(y))$ ed i loro quadrati costituiscono perciò delle forme quadratiche semidefinite posivive delle stesse variabili, al pari della NSSAR. L'indipendenza stocastica di NSSAR e $(a_j - \alpha_j)^2$ viene dunque stabilita mediante il teorema di Craig.

A questo scopo si esplicitano le somme nella relazione matriciale (33.1)

$$a_j - \alpha_j \,=\, [F^{-1}\Phi z]_j \,=\, \sum_{kl} (F^{-1})_{jk}\Phi_{kl} z_l$$

e si scrive quindi

$$\begin{aligned}
(a_j - \alpha_j)^2 \,&=\, \sum_{kl} (F^{-1})_{jk}\Phi_{kl} z_l \sum_{mq} (F^{-1})_{jm}\Phi_{mq} z_q \,= \\
&=\, \sum_{lq} z_l z_q \sum_{km} (F^{-1})_{jk}\Phi_{kl}(F^{-1})_{jm}\Phi_{mq}
\end{aligned}$$

per cui la matrice rappresentativa della variabile $(a_j - \alpha_j)^2$ diventa

$$A^{(j)}_{lq} \,=\, \sum_{km} (F^{-1})_{jk}\Phi_{kl}(F^{-1})_{jm}\Phi_{mq} \qquad l, q = 1, \ldots, n\,.$$

La matrice rappresentativa B di NSSAR è stata invece già ricavata in (32.3)

$$B_{qr} \,=\, (\mathbb{I} - \Phi^T F^{-1}\Phi)_{qr} \,=\, \delta_{qr} - \sum_{st} \Phi_{sq}(F^{-1})_{st}\Phi_{tr}\,.$$

Il prodotto delle matrici A e B ha elementi

$$\begin{aligned}
(A^{(j)}B)_{lr} \,&=\, \sum_q A^{(j)}_{lq} B_{qr} \,=\, \sum_{kmq} (F^{-1})_{jk}\Phi_{kl}(F^{-1})_{jm}\Phi_{mq}\Big(\delta_{qr} - \sum_{st}\Phi_{sq}(F^{-1})_{st}\Phi_{tr}\Big) \,= \\
&=\, \sum_{km} (F^{-1})_{jk}\Phi_{kl}(F^{-1})_{jm}\Phi_{mr} - \sum_{kmq} (F^{-1})_{jk}\Phi_{kl}(F^{-1})_{jm}\Phi_{mq} \sum_{st}\Phi_{sq}(F^{-1})_{st}\Phi_{tr}
\end{aligned}$$

e poiché l'ultimo termine si semplifica in

$$
\begin{aligned}
&\sum_{kmq}(F^{-1})_{jk}\Phi_{kl}(F^{-1})_{jm}\Phi_{mq}\sum_{st}\Phi_{sq}(F^{-1})_{st}\Phi_{tr} = \\
&= \sum_{kmst}(F^{-1})_{jk}\Phi_{kl}(F^{-1})_{jm}(F^{-1})_{st}\Phi_{tr}\sum_{q}\Phi_{mq}\Phi_{sq} = \\
&= \sum_{kmst}(F^{-1})_{jk}\Phi_{kl}(F^{-1})_{jm}(F^{-1})_{st}\Phi_{tr}\sum_{q}\Phi_{mq}(\Phi^T)_{qs} = \\
&= \sum_{kmst}(F^{-1})_{jk}\Phi_{kl}(F^{-1})_{jm}(F^{-1})_{st}\Phi_{tr}(\Phi\Phi^T)_{ms} = \\
&= \sum_{kmst}(F^{-1})_{jk}\Phi_{kl}(F^{-1})_{jm}(F^{-1})_{st}\Phi_{tr}F_{ms} = \\
&= \sum_{kmt}(F^{-1})_{jk}\Phi_{kl}(F^{-1})_{jm}\Phi_{tr}\sum_{s}F_{ms}(F^{-1})_{st} = \\
&= \sum_{kmt}(F^{-1})_{jk}\Phi_{kl}(F^{-1})_{jm}\Phi_{tr}\delta_{mt} = \\
&= \sum_{km}(F^{-1})_{jk}\Phi_{kl}(F^{-1})_{jm}\Phi_{mr}
\end{aligned}
$$

si perviene al risultato richiesto

$$
A^{(j)}B)_{lr} = \sum_{km}(F^{-1})_{jk}\Phi_{kl}(F^{-1})_{jm}\Phi_{mr} - \sum_{km}(F^{-1})_{jk}\Phi_{kl}(F^{-1})_{jm}\Phi_{mr} = 0\,.
$$

Il teorema di Craig garantisce l'indipendenza stocastica di NSSAR e $(a_j - \alpha_j)^2$, dunque anche di NSSAR e a_j. $\square$

34. Intervalli di confidenza per i coefficienti della regressione lineare

Sia Φ di rango massimo p. Allora gli intervalli di confidenza per i parametri di regressione lineare $\alpha_1, \ldots, \alpha_p$ sono dati da

$$
a_j - t_{[1-\frac{\alpha}{2}],(n-p)}\sqrt{(F^{-1})_{jj}\frac{\text{NSSAR}}{n-p}} \le \alpha_j \le a_j + t_{[1-\frac{\alpha}{2}],(n-p)}\sqrt{(F^{-1})_{jj}\frac{\text{NSSAR}}{n-p}} \qquad (34.1)
$$

$\forall\, j = 1, \ldots, p$, ciascuno con livello di confidenza $1-\alpha \in (0,1)$. Nel caso di eguali varianze σ_i, gli intervalli di confidenza sono indipendenti dal comune valore della varianza.

Dimostrazione

Il teorema 33 stabilisce che per ogni $j = 1, \ldots, p$ le variabili casuali a_j e NSSAR sono stocasticamente indipendenti. Di queste — teorema 32 — la NSSAR costituisce una variabile di $\mathcal{X}^2$ a $n-p$ gradi di libertà, mentre in virtù del teorema 31 la stima a_j del parametro di regressione α_j è una variabile gaussiana di media

$$
\mathbb{E}(a_j) = \alpha_j
$$

e varianza

$$\mathbb{E}[(a_j - \alpha_j)^2] \,=\, (C_a)_{jj} \,=\, (F^{-1})_{jj} \,.$$

Ne deriva che

$$\frac{a_j - \alpha_j}{\sqrt{(F^{-1})_{jj}}}$$

è una variabile normale e che di conseguenza

$$t \,=\, \sqrt{n-p}\,\frac{1}{\sqrt{\mathrm{NSSAR}}}\,\frac{a_j - \alpha_j}{\sqrt{(F^{-1})_{jj}}}$$

segue una distribuzione di Student a $n-p$ gradi di libertà, quale quoziente di due variabili casuali indipendenti, l'una normale e l'altra radice quadrata di una variabile di chi-quadrato ridotta. Con livello di confidenza $1-\alpha \in (0,1)$ si ha perciò

$$t_{[\frac{\alpha}{2}],(n-p)} \le t \le t_{[1-\frac{\alpha}{2}],(n-p)}$$

condizione che causa la simmetria della distribuzione di Student può riscriversi come

$$-t_{[1-\frac{\alpha}{2}],(n-p)} \le t \le t_{[1-\frac{\alpha}{2}],(n-p)}$$

ossia

$$-t_{[1-\frac{\alpha}{2}],(n-p)} \le \sqrt{n-p}\,\frac{1}{\sqrt{\mathrm{NSSAR}}}\,\frac{a_j - \alpha_j}{\sqrt{(F^{-1})_{jj}}} \le t_{[1-\frac{\alpha}{2}],(n-p)}$$

ed equivale infine all'intervallo di confidenza cercato

$$a_j - t_{[1-\frac{\alpha}{2}],(n-p)}\sqrt{(F^{-1})_{jj}\,\frac{\mathrm{NSSAR}}{n-p}} \le \alpha_j \le a_j + t_{[1-\frac{\alpha}{2}],(n-p)}\sqrt{(F^{-1})_{jj}\,\frac{\mathrm{NSSAR}}{n-p}} \,.$$

Nel caso si abbia $\sigma_i = \sigma \;\forall\, i = 1,\ldots,n$, i coefficienti $(F^{-1})_{jk}$ sono proporzionali a σ^2, data la proporzionalità fra F e σ^{-2}

$$F_{jk} \,=\, \sum_{i=1}^{n} \frac{\phi_j(x_i)\phi_k(x_i)}{\sigma_i^2} \,=\, \frac{1}{\sigma^2}\sum_{i=1}^{n}\phi_j(x_i)\phi_k(x_i) \qquad \Longrightarrow \qquad (F^{-1})_{jk} \sim \sigma^2 \,,$$

mentre gli elementi di Φ sono proporzionali a σ^{-1}

$$\Phi_{ji} \,=\, \frac{\phi_j(x_i)}{\sigma_i} \,=\, \frac{\phi_j(x_i)}{\sigma} \,.$$

Di qui segue che le stime di chi-quadrato dei parametri di regressione non presentano nessuna dipendenza da σ

$$a \,=\, F^{-1}\Phi\sigma^{-1}y \sim \sigma^2\sigma^{-1}\sigma^{-1} \,=\, 1 \,.$$

In modo analogo, la somma normalizzata dei quadrati degli scarti attorno alla regressione risulta proporzionale a σ^{-2}

$$\mathrm{NSSAR} = \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \Big[-y_i + \sum_{j=1}^{p} a_j \phi_j(x_i)\Big]^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} \Big[-y_i + \sum_{j=1}^{p} a_j \phi_j(x_i)\Big]^2$$

e di conseguenza anche le ampiezze degli intervalli di confidenza sono indipendenti da σ

$$(F^{-1})_{jj}\mathrm{NSSAR} \sim \sigma^2 \mathrm{NSSAR} = \sum_{i=1}^{n} \Big[-y_i + \sum_{j=1}^{p} a_j \phi_j(x_i)\Big]^2 . \qquad \square$$

Osservazione. Intervalli di confidenza simultanei dei parametri di regressione
Si sottolinea che la probabilità che i valori veri α_j dei parametri di regressione siano **tutti** compresi contemporaneamente entro i rispettivi intervalli confidenza

$$a_j - t_{[1-\frac{\alpha}{2}],(n-p)} \sqrt{(F^{-1})_{jj} \frac{\mathrm{NSSAR}}{n-p}} \le \alpha_j \le a_j + t_{[1-\frac{\alpha}{2}],(n-p)} \sqrt{(F^{-1})_{jj} \frac{\mathrm{NSSAR}}{n-p}}$$

non è in generale eguale a $(1-\alpha)^p$. Ciò in conseguenza del fatto che le variabili $a_1, \ldots, a_p$ non sono indipendenti e che quindi gli eventi descritti dagli intervalli di confidenza precedenti non risultano a loro volta stocasticamente indipendenti: la probabilità dell'evento composto non si identifica con il prodotto delle probabilità dei singoli eventi componenti. Tale circostanza ricorre soltanto nell'ipotesi di indipendenza stocastica delle variabili a_j, che a sua volta richiede la condizione di ortogonalità dei vettori $\Phi_1, \ldots, \Phi_p$, come specificato dal teorema 31.

35. Intervallo di confidenza per le previsioni

Sia Φ di rango massimo p. Dato $x_0 \in \mathbb{R}$ si assuma che il valore della variabile y corrispondente all'ascissa $x = x_0$ sia descritto dalla variabile casuale

$$\zeta_0 = \sum_{j=1}^{p} \alpha_j \phi_j(x_0) + \varepsilon_0$$

con ε_0 variabile gaussiana di media nulla e varianza σ_0^2, indipendente dalle ε_i del modello di regressione lineare. Si definisca la **previsione** *del valore di y in x_0 in base alla regressione lineare come*

$$y_0 = \sum_{j=1}^{p} a_j \phi_j(x_0) .$$

Allora la previsione y_0 è una variabile gaussiana di media ζ_0 e varianza

$$\mathbb{E}[(y_0 - \zeta_0)^2] = \sum_{j,k=1}^{p} (F^{-1})_{jk}\, \phi_j(x_0) \phi_k(x_0) + \sigma_0^2$$

e l'intervallo di confidenza di ζ_0*, con livello di confidenza* $1-\alpha \in (0,1)$*, assume la forma*

$$\begin{aligned} & y_0 - t_{[1-\frac{\alpha}{2}],(n-p)} \sqrt{\mathbb{E}[(y_0-\zeta_0)^2]} \sqrt{\frac{\text{NSSAR}}{n-p}} \le \zeta_0 \le \\ & \le y_0 + t_{[1-\frac{\alpha}{2}],(n-p)} \sqrt{\mathbb{E}[(y_0-\zeta_0)^2]} \sqrt{\frac{\text{NSSAR}}{n-p}} \,. \end{aligned} \tag{35.1}$$

Qualora la varianza sia indipendente dall'ascissa — $\sigma_i = \sigma \ \forall\, i = 1,\ldots,n$ *e* $\sigma_0 = \sigma$ *— il precedente intervallo di confidenza è a sua volta indipendente da* σ*.*

Dimostrazione

La previsione ad $x = x_0$ si definisce come

$$y_0 \;=\; \sum_{j=1}^{p} a_j \phi_j(x_0)$$

e stima il valore medio della variabile casuale

$$\zeta_0 \;=\; \sum_{j=1}^{p} \alpha_j \phi_j(x_0) + \varepsilon_0$$

con ε_0 variabile gaussiana di media nulla e varianza σ_0^2, indipendente da tutte le variabili ε_i, $i = 1,\ldots,n$, del modello. Ne segue che la differenza

$$y_0 - \zeta_0 \;=\; \sum_{j=1}^{p} (a_j - \alpha_j)\phi_j(x_0) - \varepsilon_0$$

è una combinazione lineare di variabili gaussiane indipendenti, di media nulla

$$\mathbb{E}(a_j - \alpha_j) \;=\; 0 \quad \forall\, j = 1,\ldots,p \qquad \mathbb{E}(\varepsilon_0) \;=\; 0$$

e varianze rispettive

$$\mathbb{E}[(a_j - \alpha_j)^2] \;=\; (F^{-1})_{jj} \quad \forall\, j = 1,\ldots,p \qquad \mathbb{E}(\varepsilon_0^2) \;=\; \sigma_0^2 \,.$$

Anche $y_0 - \zeta_0$ è dunque una variabile gaussiana, di media nulla

$$\mathbb{E}(y_0 - \zeta_0) \;=\; \sum_{j=1}^{p} \mathbb{E}(a_j - \alpha_j)\phi_j(x_0) + \mathbb{E}(\varepsilon_0) \;=\; 0$$

e varianza

$$\mathbb{E}[(y_0 - \zeta_0)^2] \;=\; \mathbb{E}\Big[\Big[\sum_{j=1}^{p} (a_j - \alpha_j)\phi_j(x_0) - \varepsilon_0\Big]^2\Big] \;=$$

$$
\begin{aligned}
&= \mathbb{E}\Big[\Big[\sum_{j=1}^{p}(a_j-\alpha_j)\phi_j(x_0)-\varepsilon_0\Big]\Big[\sum_{k=1}^{p}(a_k-\alpha_k)\phi_k(x_0)-\varepsilon_0\Big]\Big] = \\
&= \sum_{j,k=1}^{p}\mathbb{E}[(a_j-\alpha_j)(a_k-\alpha_k)]\,\phi_j(x_0)\phi_k(x_0) - 2\sum_{j=1}^{p}\mathbb{E}[(a_j-\alpha_j)\varepsilon_0]\,\phi_j(x_0)+\mathbb{E}(\varepsilon_0^2) = \\
&= \sum_{j,k=1}^{p}\mathbb{E}[(a_j-\alpha_j)(a_k-\alpha_k)]\,\phi_j(x_0)\phi_k(x_0)+\mathbb{E}(\varepsilon_0^2) = \\
&= \sum_{j,k=1}^{p}(F^{-1})_{jk}\,\phi_j(x_0)\phi_k(x_0)+\sigma_0^2\,.
\end{aligned}
$$

Si noti poi che tutte le variabili $a_j-\alpha_j$ e la ε_0 sono stocasticamente indipendenti dalla NSSAR, per cui anche

$$\frac{y_0-\zeta_0}{\sqrt{\mathbb{E}[(y_0-\zeta_0)^2]}} \qquad \text{e} \qquad \text{NSSAR}$$

risultano indipendenti, normale la prima e $\mathcal{X}^2$ a $n-p$ gradi di libertà la seconda. La variabile casuale

$$t = \sqrt{n-p}\,\frac{1}{\sqrt{\text{NSSAR}}}\,\frac{y_0-\zeta_0}{\sqrt{\mathbb{E}[(y_0-\zeta_0)^2]}}$$

si distribuisce perciò come una t di Student a $n-p$ gradi di libertà. L'intervallo di confidenza di ζ_0 è allora esprimibile nella forma (35.1), con livello di confidenza $1-\alpha\in(0,1)$.
Se $\sigma_i=\sigma\ \forall\, i=1,\ldots,n$ e $\sigma_0=\sigma$ l'intervallo di confidenza di ζ_0 è indipendente da σ. In tal caso infatti i coefficienti a_j non dipendono da σ e di conseguenza neppure y_0; basta allora considerare che

$$\mathbb{E}[(y_0-\zeta_0)^2]\sim\sigma^2 \qquad \text{e} \qquad \text{NSSAR}\sim\sigma^{-2}$$

per concludere che l'ampiezza dell'intervallo di confidenza non è funzione di σ. □

36. Regressione lineare e operatori di proiezione

• Formulazione del problema della regressione lineare
Il problema della regressione lineare con il metodo del chi quadrato si riconduce al calcolo dei punti critici del funzionale quadratico

$$\mathfrak{L}(\xi_1,\ldots,\xi_p) = \sum_{i=1}^{n}\frac{1}{\sigma_i^2}\Big[-y_i+\sum_{j=1}^{p}\xi_j\phi_j(x_i)\Big]^2 = \sum_{i=1}^{n}\Big[-\frac{1}{\sigma_i}y_i+\sum_{j=1}^{p}\xi_j\frac{1}{\sigma_i}\phi_j(x_i)\Big]^2,$$

in cui si assume che per ogni $j=1,\ldots,p$

$$\phi_j(x_i)\neq 0 \quad \text{almeno per un } i=1,\ldots,n\,, \tag{36.1}$$

in modo da assicurare l'effettiva dipendenza di $\mathfrak{L}$ da tutti gli argomenti $\xi_1,\ldots,\xi_p$.

• Formulazione matriciale/vettoriale del problema della regressione lineare
Si introducano — come già in (30.3) — un generico vettore colonna $y \in \mathbb{R}^n$ e la matrice $n \times n$ σ^{-1}

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \qquad \sigma^{-1} = \begin{pmatrix} 1/\sigma_1 & & & \\ & 1/\sigma_2 & & \mathbb{O} \\ & & \ddots & \\ & \mathbb{O} & & 1/\sigma_n \end{pmatrix}$$

nonché la matrice $p \times n$ Φ di elementi

$$\Phi_{ji} = \frac{1}{\sigma_i}\phi_j(x_i) \qquad 1 \le j \le p \quad 1 \le i \le n$$

e i vettori colonna $\Phi_1, \ldots, \Phi_p \in \mathbb{R}^n$ definiti da

$$\Phi_j \in \mathbb{R}^n \;:\quad (\Phi_j)_i = \Phi_{ji} = \frac{1}{\sigma_i}\phi_j(x_i) \quad 1 \le i \le n \quad 1 \le j \le p \tag{36.2}$$

nessuno dei quali è nullo in virtù della (36.1). Allora l'espressione entro parentesi quadre nella definizione di $\mathfrak{L}$ si identifica con la componente i-esima del vettore

$$-\sigma^{-1}y + \sum_{j=1}^{p} \xi_j \Phi_j$$

e il funzionale si può riscrivere nella forma

$$\begin{aligned}\mathfrak{L}(\xi_1, \ldots, \xi_p) &= \Big(-\sigma^{-1}y + \sum_{j=1}^{p} \xi_j \Phi_j\Big)^T \Big(-\sigma^{-1}y + \sum_{k=1}^{p} \xi_k \Phi_k\Big) = \\ &= \Big(-y^T\sigma^{-1} + \sum_{j=1}^{p} \xi_j \Phi_j^T\Big)\Big(-\sigma^{-1}y + \sum_{k=1}^{p} \xi_k \Phi_k\Big) .\end{aligned} \tag{36.3}$$

Si osservi che se si considera lo spazio $\mathbb{R}^n$ munito dell'usuale prodotto scalare

$$\langle x|y\rangle = \sum_{i=1}^{n} x_i y_i \qquad \forall\, x^T = (x_1, \ldots, x_n)\,,\; y^T = (y_1, \ldots, y_n)$$

e della relativa norma indotta

$$|x|_2 = \langle x|x\rangle^{1/2}\,,$$

al funzionale $\mathfrak{L}$ si può attribuire l'espressione equivalente

$$\mathfrak{L}(\xi_1, \ldots, \xi_p) = \Big\langle -\sigma^{-1}y + \sum_{j=1}^{p} \xi_j \Phi_j \Big| -\sigma^{-1}y + \sum_{k=1}^{p} \xi_k \Phi_k \Big\rangle$$

ovvero

$$\mathfrak{L}(\xi_1,\ldots,\xi_p) \;=\; \Big|-\sigma^{-1}y+\sum_{j=1}^{p}\xi_j\Phi_j\Big|_2^2 .$$

• **Esistenza dei punti critici**

Trattandosi di funzionale continuo e non negativo in $\mathbb{R}^p$, con

$$\lim_{|(\xi_1,\ldots,\xi_p)|\to+\infty}\mathfrak{L}(\xi_1,\ldots,\xi_p) \;=\; +\infty\,, \tag{36.4}$$

l'esistenza di un minimo assoluto di $\mathfrak{L}$ è assicurata dal teorema di Weierstrass, mentre la differenziabilità del funzionale in tutto $\mathbb{R}^p$ consente di identificare tale minimo con un punto critico — non necessariamente unico.
In effetti, la condizione (36.4) implica che comunque si assegni un $H>0$ prefissato a piacere sia possibile determinare un raggio $R>0$ tale che $\forall\,(\xi_1,\ldots,\xi_p)\in\mathbb{R}^p$, $|(\xi_1,\ldots,\xi_p)|_2<R$ si abbia

$$\mathfrak{L}(\xi_1,\ldots,\xi_p)>H\,.$$

Il problema della ricerca dei minimi assoluti di $\mathfrak{L}$ viene quindi limitato alla sola sfera compatta con centro l'origine e di raggio R

$$\{(\xi_1,\ldots,\xi_p)\in\mathbb{R}^p\;:\;|(\xi_1,\ldots,\xi_p)|\le R\}$$

sulla quale l'esistenza di almeno un minimo segue immediatamente dal teorema di Weierstrass, data l'evidente continuità del funzionale quadratico.

• **Equazione dei punti critici**

I punti critici $(\xi_1,\ldots,\xi_p)=(a_1,\ldots,a_p)$ del funzionale (36.3) si ricavano annullando le derivate parziali prime

$$\begin{aligned}\frac{\partial\mathfrak{L}}{\partial\xi_h}(\xi_1,\ldots,\xi_p) \;&=\; \Phi_h^T\Big(-\sigma^{-1}y+\sum_{k=1}^{p}\xi_k\Phi_k\Big)+\Big(-y^T\sigma^{-1}+\sum_{j=1}^{p}\xi_j\Phi_j^T\Big)\Phi_h \;=\\ &=\; \Phi_h^T\Big(-\sigma^{-1}y+\sum_{k=1}^{p}\xi_k\Phi_k\Big)+\Phi_h^T\Big(-\sigma^{-1}y+\sum_{k=1}^{p}\xi_k\Phi_k\Big) \;=\\ &=\; 2\Phi_h^T\Big(-\sigma^{-1}y+\sum_{k=1}^{p}\xi_k\Phi_k\Big)\end{aligned}$$

e sono quindi tutte e sole le soluzioni del sistema lineare

$$\Phi_h^T\Big(-\sigma^{-1}y+\sum_{k=1}^{p}a_k\Phi_k\Big) \;=\; 0 \qquad 1\le h\le p$$

ossia, equivalentemente, di

$$\Big\langle\Phi_h\Big|-\sigma^{-1}y+\sum_{k=1}^{p}a_k\Phi_k\Big\rangle \;=\; 0 \qquad 1\le h\le p\,, \tag{36.5}$$

che può anche porsi nella forma

$$\sum_{k=1}^{p} \langle \Phi_h | \Phi_k \rangle a_k \;=\; \langle \Phi_h | \sigma^{-1} y \rangle \qquad 1 \le h \le p\,. \tag{36.6}$$

• Unicità dei punti critici

L'esistenza di almeno una soluzione è garantita — come già sottolineato — dalle proprietà generali del funzionale $\mathfrak{L}$; per l'unicità della soluzione è tuttavia necessario e sufficiente assumere che la matrice Φ abbia rango massimo p, fatto questo che equivale a richiedere la lineare indipendenza dei vettori Φ_j, $j = 1, \ldots, p$. In effetti, la soluzione di (36.6) risulta unica se e soltanto se il sistema omogeneo associato

$$\sum_{k=1}^{p} \Phi_h^T \Phi_k z_k \;=\; 0 \qquad 1 \le h \le p \tag{36.7}$$

ammette come unica soluzione quella banale $(z_1, \ldots, z_p) = (0, \ldots, 0)$, circostanza questa che ricorre se e soltanto se la matrice incompleta F del sistema, di elementi

$$F_{hk} \;=\; \Phi_h^T \Phi_k \qquad 1 \le h \le p \quad 1 \le k \le p\,,$$

è non singolare. Per verificare questa condizione basta osservare che la matrice F è simmetrica e che $\forall\, u^T = (u_1, \ldots, u_p) \in \mathbb{R}^p$ la forma quadratica associata vale

$$\sum_{h,k=1}^{p} u_h F_{hk} u_k \;=\; \sum_{h,k=1}^{p} u_h \Phi_h^T \Phi_k u_k \;=\; \Big(\sum_{h=1}^{p} u_h \Phi_h\Big)^T \Big(\sum_{k=1}^{p} u_k \Phi_k\Big) \;=$$

$$=\; \Big\langle \sum_{h=1}^{p} u_h \Phi_h \Big| \sum_{k=1}^{p} u_k \Phi_k \Big\rangle \;=\; \Big|\sum_{h=1}^{p} u_h \Phi_h\Big|_2^2 \ge 0$$

con

$$\sum_{h,k=1}^{p} u_h F_{hk} u_k \;=\; \Big|\sum_{h=1}^{p} u_h \Phi_h\Big|_2^2 \;=\; 0 \qquad \Longleftrightarrow \qquad \sum_{h=1}^{p} u_h \Phi_h \;=\; 0\,.$$

F risulta pertanto definita positiva se e soltanto se i vettori colonna $\Phi_1, \ldots, \Phi_p$ sono linearmente indipendenti, allorché si ha

$$\sum_{h,k=1}^{p} u_h F_{hk} u_k \;=\; 0 \qquad \Longrightarrow \qquad (u_1, \ldots, u_p) \;=\; (0, \ldots, 0)$$

mentre in caso contrario F è soltanto semidefinita non definita positiva — e dunque singolare.

• Unicità del vettore di regressione

Si è stabilito che la soluzione $(\xi_1, \ldots, \xi_p) = (a_1, \ldots, a_p)$ del problema di regressione non è in generale unica, ma risulta definita a meno di un'arbitraria soluzione $(z_1, \ldots, z_p)$ dell'equazione lineare omogenea (36.7)

$$\sum_{k=1}^{p} \Phi_h^T \Phi_k z_k = 0 \qquad 1 \leq h \leq p\,.$$

Si osservi che per una qualsiasi soluzione del problema omogeneo devono valere le relazioni

$$z_h \sum_{k=1}^{p} \Phi_h^T \Phi_k z_k = 0 \qquad 1 \leq h \leq p$$

e quindi, sommando sull'indice h

$$\sum_{h=1}^{p} z_h \sum_{k=1}^{p} \Phi_h^T \Phi_k z_k = 0$$

equazione dalla quale si deduce che

$$\Big(\sum_{h=1}^{p} z_h \Phi_h\Big)^T \sum_{k=1}^{p} z_k \Phi_k = \Big\langle \sum_{h=1}^{p} z_h \Phi_h \Big| \sum_{k=1}^{p} z_k \Phi_k \Big\rangle = \Big|\sum_{h=1}^{p} z_h \Phi_h\Big|_2^2 = 0$$

ossia

$$\sum_{h=1}^{p} z_h \Phi_h = 0\,.$$

Ne deriva che la soluzione generale del problema di regressione

$$(\xi_1, \ldots, \xi_p) = (a_1, \ldots, a_p) + (z_1, \ldots, z_p)$$

individua un unico **vettore di regressione**

$$\sum_{h=1}^{p} (a_h + z_h)\Phi_h = \sum_{h=1}^{p} a_h \Phi_h\,.$$

Introdotto il sottospazio lineare generato dai vettori colonna $\Phi_1, \ldots, \Phi_p$

$$V_p = \mathrm{span}\{\Phi_1, \ldots, \Phi_p\}$$

ed il relativo complemento ortogonale in $\mathbb{R}^n$

$$V_p^\perp = \{x \in \mathbb{R}^n \,:\, \langle \Phi_h | x\rangle = 0 \;\forall\, h = 1, \ldots, p\}\,,$$

allo stesso risultato si può pervenire osservando che $\mathbb{R}^n$ si decompone in somma diretta dei precedenti sottospazi ortogonali

$$\mathbb{R}^n = V_p \oplus V_p^\perp$$

e che di conseguenza ogni vettore $\sigma^{-1}y \in \mathbb{R}^n$ deve potersi scrivere in un solo modo come somma di un vettore in V_p e di un vettore di $V_p^\perp$. Nella fattispecie si ha la decomposizione in termini del vettore di regressione **calcolato per** $\sigma^{-1}y$

$$\sigma^{-1}y = \sum_{h=1}^{p} a_h\Phi_h + \Big(\sigma^{-1}y - \sum_{h=1}^{p} a_h\Phi_h\Big)$$

nella quale ovviamente il primo vettore appartiene allo spazio V_p — ciò avverrebbe per qualsiasi scelta dei coefficienti $a_1, \ldots, a_p$ —

$$\sum_{h=1}^{p} a_h\Phi_h \in V_p$$

mentre

$$\sigma^{-1}y - \sum_{h=1}^{p} a_h\Phi_h \in V_p^\perp \,,$$

per effetto delle (36.5) — che invece valgono se e soltanto se $\sum_{h=1}^{p} a_h\Phi_h$ è il vettore di regressione calcolato per $\sigma^{-1}y$. Dall'unicità della decomposizione segue pertanto che il vettore

$$\sum_{h=1}^{p} a_h\Phi_h$$

con le $a_1, \ldots, a_p$ determinate secondo le (36.5) è **il solo vettore di regressione** corrispondente a $\sigma^{-1}y$.

• **Operatore di regressione**
Poiché il risultato precedente fa corrispondere ad ogni vettore $w \in \mathbb{R}^n$ un unico vettore di regressione $\sum_{h=1}^{p} a_h\Phi_h$, è possibile considerare l'applicazione di $\mathbb{R}^n$ in sé

$$P_p \;:\; w \in \mathbb{R}^n \longrightarrow P_p(w) = \sum_{h=1}^{p} a_h\Phi_h \in V_p \subset \mathbb{R}^n \tag{36.8}$$

la cui linearità si deduce immediatamente dalle (36.5). Comunque si fissino w', w'' in $\mathbb{R}^n$ e λ, $\lambda'' \in \mathbb{R}$ si ha infatti

$$\big\langle \Phi_h \big| -w' + P_p(w')\big\rangle = 0 \qquad 1 \le h \le p$$

$$\big\langle \Phi_h \big| -w'' + P_p(w'')\big\rangle = 0 \qquad 1 \le h \le p$$

per cui

$$\lambda'\big\langle \Phi_h \big| -w' + P_p(w')\big\rangle + \lambda''\big\langle \Phi_h \big| -w'' + P_p(w'')\big\rangle = 0 \qquad 1 \le h \le p$$

vale a dire

$$\big\langle \Phi_h \big| -(\lambda' w' + \lambda'' w'') + \lambda' P_p(w') + \lambda'' P_p(w'')\big\rangle = 0 \qquad 1 \le h \le p$$

e dunque, per definizione dell'applicazione P_p,

$$P_p(\lambda' w' + \lambda'' w'') = \lambda' P_p(w') + \lambda'' P_p(w'') .$$

L'applicazione lineare P_p è definita come **operatore di regressione lineare** e può interpretarsi come l'**operatore di proiezione ortogonale** sul sottospazio V_p di $\mathbb{R}^n$, che ad ogni $w \in \mathbb{R}^n$ fa corrispondere l'unico vettore $P_p(w)$ per il quale risulta

$$w = P_p(w) + y - P_p(w)$$

con $P_p(w) \in V_p$ e $w - P_p(\sigma^{-1}y) \in V_p^\perp$.

• Proprietà dell'operatore di regressione

Le proprietà più notevole dell'operatore P_p seguono dalla sua natura di operatore di proiezione ortogonale.

Il fatto che P_p sia un operatore di proiezione ne assicura l'**idempotenza**

$$P_p^2 = P_p$$

in quanto

$$P_p^2 w = P_p(P_p w) = P_p(P_p w + 0) = P_p w \qquad \forall\, w \in \mathbb{R}^n$$

essendo $P_p w \in V_p$ per definizione.

L'ortogonalità degli spazi V_p e $V_p^\perp$ implica inoltre la **simmetria** dell'operatore di proiezione, dal momento che se $\forall\, v, w \in \mathbb{R}^n$ si pone

$$v = v' + v'' \quad , \qquad v' \in V_p \quad v'' \in V_p^\perp$$

$$w = w' + w'' \quad , \qquad w' \in V_p \quad w'' \in V_p^\perp$$

si ricava

$$\langle v|P_p w\rangle = \langle v|w'\rangle = \langle v' + v''|w'\rangle = \langle v'|w'\rangle = \langle v'|w' + w''\rangle = \langle v'|w\rangle = \langle P_p v|w\rangle .$$

Di qui è poi facile stabilire il carattere **semidefinito positivo** dell'operatore

$$\langle w|P_p w\rangle = \langle w|P_p^2 w\rangle = \langle P_p w|P_p w\rangle = |P_p w|_2^2 \geq 0 \qquad \forall\, w \in \mathbb{R}^n .$$

Dalla simmetria e idempotenza di P_p si deduce infine che gli autovalori dell'operatore sono 0 e 1

$$P_p w = \lambda w \quad \Longrightarrow \quad 0 = (P_p^2 - P_p)w = (\lambda^2 - \lambda)w \quad \Longrightarrow \quad \lambda(\lambda - 1) = 0$$

• Traccia e matrice rappresentativa dell'operatore di regressione

La traccia $\mathrm{tr}(P_p)$ dell'operatore di regressione coincide, in generale, con la somma dei suoi autovalori; nella fattispecie, siccome gli autovalori di P_p sono soltanto 0 e 1, la traccia si identifica con la moltemplicità dell'autovalore 1. Tale molteplicità coincide peraltro con la dimensione del range di P_p

$$\mathrm{tr}(P_p) \;=\; \text{molteplicità dell'autovalore } 1 \;=\; \dim \mathrm{range}(P_p)$$

dal momento che presa una qualsiasi base di autovettori ortonormali quelli associati all'autovalore 1 costituiscono una base del range, mentre quelli relativi a 0 forniscono una base del nucleo

$$\text{molteplicità dell'autovalore } 0 \;=\; \dim \ker(P_p)\,.$$

D'altra parte, per un operatore di proiezione quale è P_p il range coincide con l'intero sottospazio di proiezione, ossia

$$\mathrm{range}(P_p) \;=\; V_p \;=\; \mathrm{span}\{\Phi_1, \ldots, \Phi_p\}$$

per cui

$$\dim \mathrm{range}(P_p) \;=\; \dim V_p \;=\; \dim \mathrm{span}\{\Phi_1, \ldots, \Phi_p\} \;=\; \mathrm{rank}\Phi\,.$$

In definitiva la traccia — e rango — dell'operatore di regressione è data dalla relazione

$$\mathrm{tr}(P_p) \;=\; \mathrm{rank}\Phi$$

e coincide quindi con il numero massimo di vettori colonna $\Phi_1, \ldots, \Phi_p$ linearmente indipendenti.

In particolare, nell'ipotesi consueta che i vettori $\Phi_1, \ldots, \Phi_p$ siano linearmente indipendenti — ovvero Φ di rango massimo p — si conclude che

$$\mathrm{tr}(P_p) \;=\; p\,.$$

Osservazione. Ancora sulla traccia di P_p
Allo stesso risultato si può pervenire attraverso il calcolo diretto della matrice rappresentativa L'operatore di regressione P_p è definito da

$$P_p(w) \;=\; \sum_{k=1}^{p} a_k \Phi_k$$

con gli $a_1, \ldots, a_p$ verificanti il sistema (36.5)

$$\sum_{k=1}^{p} \Phi_h^T \Phi_k a_k \;=\; \Phi_h^T w \qquad 1 \le h \le p$$

ovvero

$$\sum_{k=1}^{p} F_{hk} a_k \;=\; \Phi_h^T w \qquad 1 \le h \le p\,.$$

Qualora si assuma la lineare indipendenza dei vettori colonna $\Phi_1, \ldots, \Phi_p$ — vale a dire il rango massimo p della matrice Φ —, la matrice F risulta invertibile e i coefficienti $a_1, \ldots, a_p$ sono determinati univocamente come

$$a_k = \sum_{h=1}^{p} (F^{-1})_{kh} \Phi_h^T w \qquad 1 \leq k \leq p \,.$$

Ne deriva la seguente espressione per l'operatore di regressione

$$P_p(w) = \sum_{k=1}^{p} a_k \, \Phi_k = \sum_{h,k=1}^{p} (F^{-1})_{kh} \Phi_h^T w \; \Phi_k$$

e la componente i-esima dell'immagine $P_p(w) \in \mathbb{R}^n$ si scriverà

$$\begin{aligned} [P_p(w)]_i &= \sum_{h,k=1}^{p} (F^{-1})_{kh} \Phi_h^T w \; (\Phi_k)_i = \\ &= \sum_{h,k=1}^{p} (F^{-1})_{kh} \sum_{j=1}^{n} (\Phi_h)_j w_j \; (\Phi_k)_i = \\ &= \sum_{j=1}^{n} \sum_{h,k=1}^{p} (F^{-1})_{kh} (\Phi_h)_j (\Phi_k)_i \; w_j \end{aligned}$$

e conseguentemente la matrice rappresentativa $[P_p]$ dell'operatore rispetto alla base canonica avrà elementi

$$[P_p]_{ij} = \sum_{h,k=1}^{p} (F^{-1})_{kh} (\Phi_h)_j (\Phi_k)_i \qquad 1 \leq i \leq n \quad 1 \leq j \leq n \,.$$

La traccia dell'operatore coincide con la traccia di una sua qualsiasi rappresentazione matriciale, perciò

$$\begin{aligned} \mathrm{tr}(P_p) &= \mathrm{tr}[P_p] = \sum_{i=1}^{n} [P_p]_{ii} = \sum_{i=1}^{n} \sum_{h,k=1}^{p} (F^{-1})_{kh} (\Phi_h)_i (\Phi_k)_i = \\ &= \sum_{h,k=1}^{p} (F^{-1})_{kh} \Phi_h^T \Phi_k = \sum_{h,k=1}^{p} (F^{-1})_{kh} F_{hk} = \sum_{k=1}^{p} (F^{-1}F)_{kk} = \sum_{k=1}^{p} \delta_{kk} = p \,, \end{aligned}$$

a conferma del risultato già ottenuto.

• **Proprietà dell'operatore $\mathbb{I} - P_p$**
È evidente che anche $\mathbb{I} - P_p$ costituisce un operatore di proiezione ortogonale

$$(\mathbb{I} - P_p) \; : \; w \in \mathbb{R}^n \longrightarrow w - P_p w \in V_p^{\perp}$$

simmetrico, idempotente e semidefinito positivo. Infatti

$$\langle v|(\mathbb{I}-P_p)w\rangle \;=\; \langle v|w\rangle - \langle v|P_p w\rangle \;=\; \langle v|w\rangle - \langle P_p v|w\rangle \;=\; \langle(\mathbb{I}-P_p)v|w\rangle \qquad \forall\, v,w\in\mathbb{R}^n$$

$$(\mathbb{I}-P_p)^2 \;=\; \mathbb{I}-2P_p+P_2^2 \;=\; \mathbb{I}-2P_p+P_p \;=\; \mathbb{I}-P_p$$

$$\langle w|(\mathbb{I}-P_p)w\rangle \;=\; \langle w|(\mathbb{I}-P_p)^2 w\rangle \;=\; \langle(\mathbb{I}-P_p)w|(\mathbb{I}-P_p)w\rangle \geq 0 \qquad \forall\, w\in\mathbb{R}^n\,.$$

La composizione con il proiettore P_p è ovviamente nulla

$$P_p(\mathbb{I}-P_p) \;=\; P_p - P_p^2 \;=\; P_p - P_p \;=\; 0$$

mentre la traccia è legata in modo ovvio a quella di P_p

$$\mathrm{tr}(\mathbb{I}-P_p) \;=\; \mathrm{tr}(\mathbb{I}) - \mathrm{tr}(P_p) \;=\; n - \mathrm{tr}(P_p)\,.$$

• La somma normalizzata dei quadrati degli scarti attorno alla regressione come variabile di $\mathcal{X}^2$

La somma normalizzata dei quadrati degli scarti attorno alla regressione — definita dalla (32.1) — è data dal funzionale quadratico della regressione lineare $\mathfrak{L}(\xi_1,\ldots,\xi_p)$ calcolato nella soluzione $(\xi_1,\ldots,\xi_p)=(a_1,\ldots,a_p)$

$$\mathrm{NSSAR} \;=\; \mathfrak{L}(a_1,\ldots,a_p) \;=\; \sum_{i=1}^{n}\Big[-\frac{1}{\sigma_i}y_i + \sum_{j=1}^{p} a_j\frac{1}{\sigma_i}\phi_j(x_i)\Big]^2\,.$$

In accordo con il modello statistico illustrato in (29.1), le variabili casuali y_i sono stocasticamente indipendenti e gaussiane, con varianza σ_i^2 e valor medio

$$\mathbb{E}(y_i) \;=\; \sum_{j=1}^{p}\alpha_j\phi_j(x_i) \qquad 1\leq i\leq n$$

per cui risulta

$$\frac{1}{\sigma_i}\mathbb{E}(y_i) \;=\; \sum_{j=1}^{p}\alpha_j\frac{1}{\sigma_i}\phi_j(x_i) \;=\; \sum_{j=1}^{p}\alpha_j(\Phi_j)_i \;=\; \Big(\sum_{j=1}^{p}\alpha_j\Phi_j\Big)_i \qquad 1\leq i\leq n$$

e dunque

$$\sigma^{-1}\mathbb{E}(y) \;=\; \sum_{j=1}^{p}\alpha_j\Phi_j\,.$$

Considerata la definizione (36.8) dell'operatore di regressione P_p

$$P_p(\sigma^{-1}y) \;=\; \sum_{j=1}^{p}a_j\Phi_j \qquad \forall\, y\in\mathbb{R}^n$$

con gli $a_1, \ldots, a_p$ calcolati secondo le (36.6), l'essere $\sum_{j=1}^{p} \alpha_j \Phi_j \in V_p$ implica che si abbia

$$P_p(\sigma^{-1}\mathbb{E}(y)) \;=\; P_p\Big(\sum_{j=1}^{p} \alpha_j \Phi_j\Big) \;=\; \sum_{j=1}^{p} \alpha_j \Phi_j$$

da cui si deduce

$$P_p[\sigma^{-1}(y - \mathbb{E}(y))] \;=\; P_p(\sigma^{-1}y) - P_p(\sigma^{-1}\mathbb{E}(y)) \;=\; \sum_{j=1}^{p} (a_j - \alpha_j)\Phi_j$$

e conseguentemente

$$\begin{aligned}(\mathbb{I} - P_p)[\sigma^{-1}(y - \mathbb{E}(y))] \;&=\; \sigma^{-1}(y - \mathbb{E}(y)) - P_p[\sigma^{-1}(y - \mathbb{E}(y))] \;= \\ &=\; \sigma^{-1}(y - \mathbb{E}(y)) - \sum_{j=1}^{p} (a_j - \alpha_j)\Phi_j \;=\; \sigma^{-1}y - \sum_{j=1}^{p} a_j \Phi_j \,.\end{aligned}$$

È allora facile esprimere il valore minimo NSSAR $= \mathfrak{L}(a_1, \ldots, a_p)$ del funzionale quadratico in termini dell'operatore di regressione P_p

$$\begin{aligned}\mathfrak{L}(a_1, \ldots, a_p) \;&=\; \Big|\sigma^{-1}y - \sum_{j=1}^{p} a_j \Phi_j\Big|_2^2 \;=\; \Big\langle \sigma^{-1}y - \sum_{k=1}^{p} a_j \Phi_j \Big| \sigma^{-1}y - \sum_{j=1}^{p} a_j \Phi_j \Big\rangle \;= \\ &=\; \big\langle (\mathbb{I} - P_p)[\sigma^{-1}(y - \mathbb{E}(y))] \big| (\mathbb{I} - P_p)[\sigma^{-1}(y - \mathbb{E}(y))] \big\rangle \;= \\ &=\; \big\langle \sigma^{-1}(y - \mathbb{E}(y)) \big| (\mathbb{I} - P_p)^2[\sigma^{-1}(y - \mathbb{E}(y))] \big\rangle \;= \\ &=\; \big\langle \sigma^{-1}(y - \mathbb{E}(y)) \big| (\mathbb{I} - P_p)[\sigma^{-1}(y - \mathbb{E}(y))] \big\rangle\end{aligned}$$

e siccome le variabili

$$z_i \;=\; [\sigma^{-1}(y - \mathbb{E}(y))]_i \;=\; \frac{y_i - \mathbb{E}(y_i)}{\sigma_i} \qquad 1 \le i \le n$$

sono normali e indipendenti, si conclude che

$$\mathrm{NSSAR} \;=\; \mathfrak{L}(a_1, \ldots, a_p) \;=\; \langle z | (\mathbb{I} - P_p) z \rangle \tag{36.9}$$

è una variabile di $\mathcal{X}^2$ a $\mathrm{tr}(\mathbb{I} - P_p) = n - \mathrm{tr}(P_p)$ gradi di libertà. In particolare, nell'ipotesi che il rango della matrice Φ sia massimo, ossia che i vettori $\Phi_1, \ldots, \Phi_p$ siano linearmente indipendenti, il numero di gradi di libertà è pari a $n - p$.

• Relazione fra l'operatore di regressione P_p e l'operatore P_{p+s} relativo ad un modello di regressione con s parametri addizionali

Si considerino il sistema di vettori non nulli di $\mathbb{R}^n$

$$\Phi_1 \,, \quad \Phi_2 \,, \quad \ldots \,, \quad \Phi_{p+s}$$

i sottospazi lineari

$$V_p = \operatorname{span}\{\Phi_1, \ldots, \Phi_p\}$$

$$V_{p+s} = \operatorname{span}\{\Phi_1, \ldots, \Phi_p, \Phi_{p+1}, \ldots, \Phi_{p+s}\}$$

e i relativi operatori di proiezione ortogonale

$$P_p : \mathbb{R}^n \longrightarrow V_p$$

$$P_{p+s} : \mathbb{R}^n \longrightarrow V_{p+s} .$$

I sottospazi lineari soddisfano le ovvie relazioni di inclusione

$$V_p \subseteq V_{p+s} \qquad V_p^{\perp} \supseteq V_{p+s}^{\perp}$$

mentre per gli operatori di proiezione vale

$$P_p P_{p+s} = P_{p+s} P_p = P_p .$$

Quest'ultima relazione risulta ovvia nel caso si abbia $V_p = V_{p+s}$, circostanza che non è dato di escludere in quanto i vettori $\Phi_1, \ldots, \Phi_{p+s}$ possono essere linearmente dipendenti. Qualora sia $V_p \subset V_{p+s}$, si ha ovviamente che

$$a := \dim V_p < \dim V_{p+s} := b$$

e che data una base ortonormale di V_p

$$e_1 , \quad e_2 , \quad \ldots , \quad e_a$$

è sempre possibile prolungarla in una base ortonormale di V_{p+s}

$$e_1 , \quad e_2 , \quad \ldots , \quad e_a , \quad e_{a+1} , \quad \ldots , \quad e_b$$

a sua volta prolungabile in una base ortonormale dell'intero $\mathbb{R}^n$

$$e_1 , \quad e_2 , \quad \ldots , \quad e_a , \quad e_{a+1} , \quad \ldots , \quad e_b , \quad e_{b+1} , \quad e_n$$

— a questo scopo basta osservare che qualsiasi base di V_p può prolungarsi in una base di $V_{p+s} \supset V_p$, ed applicare quindi il teorema di Gram-Schmidt. In modo analogo si procede per prolungare la base di V_{p+s} così ottenuta in una base ortonormale di $\mathbb{R}^n$. — I vettori

$$e_{a+1} , \quad \ldots , \quad e_b , \quad e_{b+1} , \quad \ldots , \quad e_n$$

costituiscono una base ortonormale dello spazio ortogonale $V_p^{\perp}$, mentre

$$e_{b+1} , \quad \ldots , \quad e_n$$

forniscono una base di $V_{p+s}^\perp$. Per ogni $x = \sum_{i=1}^n x_i e_i \in \mathbb{R}^n$ vale allora

$$x = \sum_{i=1}^a x_i e_i + \sum_{i=a+1}^n x_i e_i \qquad \sum_{i=1}^a x_i e_i \in V_p \qquad \sum_{i=a+1}^n x_i e_i \in V_p^\perp$$

come pure

$$x = \sum_{i=1}^b x_i e_i + \sum_{i=b+1}^n x_i e_i \qquad \sum_{i=1}^b x_i e_i \in V_{p+s} \qquad \sum_{i=b+1}^n x_i e_i \in V_{p+s}^\perp$$

e pertanto

$$P_p x = P_p\Big(\sum_{i=1}^n x_i e_i\Big) = \sum_{i=1}^a x_i e_i$$

$$P_{p+s} x = P_{p+s}\Big(\sum_{i=1}^n x_i e_i\Big) = \sum_{i=1}^b x_i e_i\,.$$

Da queste relazioni si deduce che

$$P_{p+s}P_p x = P_{p+s}\Big(\sum_{i=1}^a x_i e_i\Big) = P_{p+s}\Big(\sum_{i=1}^a x_i e_i + \sum_{i=a+1}^b 0 e_i\Big) = \sum_{i=1}^a x_i e_i = P_p x \qquad \forall\, x \in \mathbb{R}^n$$

e quindi

$$P_{p+s}P_p = P_p\,.$$

In modo analogo si ottiene la relazione $P_p P_{p+s} = P_p$

$$P_p P_{p+s} x = P_p\Big(\sum_{i=1}^b x_i e_i\Big) = P_p\Big(\sum_{i=1}^a x_i e_i + \sum_{i=a+1}^b x_i e_i\Big) = \sum_{i=1}^a x_i e_i = P_p x\,.$$

• **Proprietà dell'operatore** $P_{p+s} - P_p$

L'operatore $P_{p+s} - P_p$ è simmetrico, semidefinito positivo e idempotente. La sua composizione con P_p è nulla. La simmetria è immediata

$$(P_{p+s} - P_p)^T = P_{p+s}^T - P_p^T = P_{p+s} - P_p$$

mentre il carattere semidefinito positivo segue dalla definizione degli operatori di regressione, in quanto

$$\begin{aligned}\langle x|(P_{p+s} - P_p)x\rangle &= \langle x|P_{p+s}x\rangle - \langle x|P_p x\rangle = \langle x|P_{p+s}^2 x\rangle - \langle x|P_p^2 x\rangle = \\ &= |P_{p+s}x|_2^2 - |P_p x|_2^2 \geq 0 \qquad \forall\, x \in \mathbb{R}^n\end{aligned}$$

mentre $\langle x|(P_{p+s} - P_p)x\rangle = 0$ per ogni $x \in V_p$. Per verificare l'idempotenza è sufficiente applicare le proprietà algebriche dimostrate al punto precedente

$$(P_{p+s} - P_p)^2 = P_{p+s}^2 + P_p^2 - P_{p+s}P_p - P_p P_{p+s} = P_{p+s} + P_p - P_p - P_p = P_{p+s} - P_p$$

le stesse che consentono di provare l'annullarsi della composizione di $P_{p+s} - P_p$ con P_p

$$(P_{p+s} - P_p)P_p = P_{p+s}P_p - P_p^2 = P_p - P_p = 0\,.$$

37. Parametro addizionale nella regressione lineare

Si considerino i modelli di regressione lineare

$$\zeta_i \;=\; \sum_{j=1}^{p} \alpha_j \phi_j(x_i) + \varepsilon_i \qquad i = 1, \ldots, n \tag{37.1}$$

e

$$\zeta_i \;=\; \sum_{j=1}^{p+1} \alpha_j \phi_j(x_i) + \varepsilon_i \qquad i = 1, \ldots, n \tag{37.2}$$

che differiscono l'uno dall'altro per l'introduzione di un parametro addizionale α_{p+1} *corrispondente alla componente* ϕ_{p+1} *della funzione di regressione. Le matrici — definite secondo la* (36.2) *—*

$$\Phi \;=\; (\Phi_1, \ldots, \Phi_p) \qquad \text{e} \qquad \Phi' \;=\; (\Phi_1, \ldots, \Phi_p, \Phi_{p+1})$$

si assumano di rango massimo, rispettivamente p *e* $p+1$.
Se (37.1) *è il modello di regressione corretto, allora le somme normalizzate dei quadrati degli scarti attorno alle due regressioni*

$$\mathrm{NSSAR}_p \;=\; \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \left[-y_i + \sum_{j=1}^{p} \phi_j(x_i) \right]^2$$

$$\mathrm{NSSAR}_{p+1} \;=\; \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \left[-y_i + \sum_{j=1}^{p+1} \phi_j(x_i) \right]^2$$

sono variabili di $\mathcal{X}^2$ *rispettivamente a* $n-p$ *e* $n-p-1$ *gradi di libertà, la loro differenza*

$$\mathrm{NSSAR}_p - \mathrm{NSSAR}_{p+1}$$

costituisce una variabile di $\mathcal{X}^2$ *a* 1 *grado di libertà ed infine*

$$\mathrm{NSSAR}_p - \mathrm{NSSAR}_{p+1} \qquad e \qquad \mathrm{NSSAR}_{p+1}$$

risultano stocasticamente indipendenti.

Dimostrazione

Si osservi preliminarmente che se il modello (37.1) è corretto, tale può intendersi anche il modello (37.2), a condizione di assumere $\alpha_{p+1} = 0$. Grazie alle condizioni assunte sul rango delle matrici Φ e Φ', il teorema 32 assicura che

$$\mathrm{NSSAR}_p \;=\; \big\langle \sigma^{-1}\big(y - \mathbb{E}(y)\big) \big| (\mathbb{I} - P_p)\sigma^{-1}\big(y - \mathbb{E}(y)\big) \big\rangle$$

sia una variabile di $\mathcal{X}^2$ a $n-p$ gradi di libertà e, allo stesso modo, che

$$\mathrm{NSSAR}_{p+1} \;=\; \big\langle \sigma^{-1}\big(y - \mathbb{E}(y)\big) \big| (\mathbb{I} - P_{p+1})\sigma^{-1}\big(y - \mathbb{E}(y)\big) \big\rangle$$

costituisca a propria volta una variabile di $\mathcal{X}^2$ a $n-(p+1)=n-p-1$ gradi di libertà. Come evidenziato, dette proprietà possono anche essere dedotte esprimendo le somme normalizzate in termini degli operatori di proiezione P_p e P_{p+1}, per mezzo della relazione (36.9). La stessa relazione (36.9) consente inoltre di scrivere la differenza $\mathrm{NSSAR}_p - \mathrm{NSSAR}_{p+1}$ nella forma

$$\begin{aligned}\mathrm{NSSAR}_p - \mathrm{NSSAR}_{p+1} \,&=\, \big\langle \sigma^{-1}\big(y-\mathbb{E}(y)\big)\big|(\mathbb{I}-P_p)\sigma^{-1}\big(y-\mathbb{E}(y)\big)\big\rangle - \\ &- \big\langle \sigma^{-1}\big(y-\mathbb{E}(y)\big)\big|(\mathbb{I}-P_{p+1})\sigma^{-1}\big(y-\mathbb{E}(y)\big)\big\rangle \,= \\ &= \big\langle \sigma^{-1}\big(y-\mathbb{E}(y)\big)\big|(P_{p+1}-P_p)\sigma^{-1}\big(y-\mathbb{E}(y)\big)\big\rangle\end{aligned}$$

e dall'idempotenza e simmetria dell'operatore $P_{p+1}-P_p$ si deduce che la variabile è di $\mathcal{X}^2$ con

$$\mathrm{tr}(P_{p+1}-P_p) \,=\, \mathrm{tr}(P_{p+1}) - \mathrm{tr}(P_p) \,=\, p+1-p \,=\, 1$$

gradi di libertà, come affermato. L'indipendenza stocastica di $\mathrm{NSSAR}_p - \mathrm{NSSAR}_{p+1}$ e NSSAR_{p+1} segue infine dalle proprietà algebriche degli operatori di proiezione e dal teorema di Craig

$$\begin{aligned}(P_{p+1}-P_p)(\mathbb{I}-P_{p+1}) \,&=\, P_{p+1}-P_p-P_{p+1}^2+P_pP_{p+1} \,= \\ &=\, P_{p+1}-P_p-P_{p+1}+P_p \,=\, 0\,. \qquad \square\end{aligned}$$

Osservazione
Le variabili casuali $\mathrm{NSSAR}_p - \mathrm{NSSAR}_{p+1}$ e NSSAR_p **non** sono stocasticamente indipendenti. Le variabili sono infatti forme quadratiche semidefinite positive delle variabili normali $\sigma^{-1}(y-\mathbb{E}(y))$, ma il teorema di Craig non è infatti verificato in quanto la composizione degli operatori di rappresentazione non è identicamente nullo

$$\begin{aligned}(P_{p+1}-P_p)(\mathbb{I}-P_p) \,&=\, P_{p+1}-P_p-P_{p+1}P_p+P_p^2 \,= \\ &=\, P_{p+1}-P_p-P_p+P_p \,=\, P_{p+1}-P_p \neq 0\end{aligned}$$

Corollario. F-test per il parametro addizionale nella regressione lineare
Dal teorema precedente si deduce immediatamente che se il modello di regressione corretto è (37.2) con $\alpha_{p+1}=0$, il quoziente

$$F \,=\, (n-p-1)\,\frac{\mathrm{NSSAR}_p - \mathrm{NSSAR}_{p+1}}{\mathrm{NSSAR}_{p+1}}$$

segue una distribuzione di Fisher a $(1, n-p-1)$ gradi di libertà. È chiaro dalla definizione di NSSAR_{p+1} che per $\alpha_{p+1}=0$ deve aversi verosimilmente $a_{p+1}\simeq 0$ e quindi

$$\mathrm{NSSAR}_{p+1} \lesssim \mathrm{NSSAR}_p$$

mentre qualora risultasse $\alpha_{p+1}\neq 0$ la somma normalizzata dei quadrati degli scarti attorno alla regressione dovrà diminuire sensibilmente nell'incrementare il numero dei parametri da p a $p+1$

$$\mathrm{NSSAR}_{p+1} \ll \mathrm{NSSAR}_p$$

Nel primo caso ci si aspetta pertanto che la variabile F sia prossima a zero

$$F = (n-p-1)\,\frac{\mathrm{NSSAR}_p - \mathrm{NSSAR}_{p+1}}{\mathrm{NSSAR}_{p+1}} \gtrsim 0$$

mentre nel secondo è del tutto ragionevole presumere un valore di F sensibilmente maggiore

$$F = (n-p-1)\,\frac{\mathrm{NSSAR}_p - \mathrm{NSSAR}_{p+1}}{\mathrm{NSSAR}_{p+1}} \gg 0\,.$$

È allora possibile testare l'ipotesi nulla

$$H_0 \;:\; \text{modello di regressione} \quad \sum_{j=1}^{p+1} \alpha_j \phi_j \quad \text{corretto con} \quad \alpha_{p+1} = 0$$

contro l'ipotesi alternativa

$$H_1 \;:\; \text{modello di regressione} \quad \sum_{j=1}^{p+1} \alpha_j \phi_j \quad \text{corretto con} \quad \alpha_{p+1} \neq 0$$

facendo uso della F come variabile del test e rigettando H_0 se

$$F \geq F_{[1-\alpha],(1,n-p-1)}$$

con livello di significatività $\alpha \in (0,1)$. Qualora l'ipotesi nulla venga rifiutata, sarà vantaggioso includere l'ulteriore termine $a_{p+1}\phi_{p+1}(x)$ nella funzione di regressione; in caso contrario la regressione a p termini dovrà essere considerata soddisfacente e l'introduzione del parametro addizionale non necessaria.

38. Retta di regressione

Dato il modello di regressione lineare

$$\zeta_i = \alpha + \beta x_i + \varepsilon_i \qquad i = 1, \ldots, n$$

e posto:

$$S = \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \qquad S_x = \sum_{i=1}^{n} \frac{x_i}{\sigma_i^2} \qquad S_y = \sum_{i=1}^{n} \frac{y_i}{\sigma_i^2}$$

$$S_{xx} = \sum_{i=1}^{n} \frac{x_i^2}{\sigma_i^2} \qquad S_{xy} = \sum_{i=1}^{n} \frac{x_i y_i}{\sigma_i^2} \qquad \Delta = S S_{xx} - S_x^2 > 0\,,$$

le stime di chi-quadrato dei parametri α e β sono date da

$$a = \frac{S_{xx}S_y - S_x S_{xy}}{\Delta} \qquad b = \frac{S S_{xy} - S_x S_y}{\Delta}$$

e costituiscono una coppia di variabili gaussiane multivariate con matrice di covarianza

$$C = \frac{1}{\Delta}\begin{pmatrix} S_{xx} & -S_x \\ -S_x & S \end{pmatrix}$$

e coefficiente di correlazione

$$\mathrm{corr}(a,b) = -\frac{S_x}{\sqrt{SS_{xx}}} .$$

La somma normalizzata dei quadrati degli scarti attorno alla regressione si scrive

$$\mathrm{NSSAR} = \sum_{i=1}^{n} \frac{1}{\sigma_i^2}(a + bx_i - y_i)^2$$

e risulta una variabile di $\mathcal{X}^2$ a $n-2$ gradi di libertà, stocasticamente indipendente tanto da a quanto da b.

Dimostrazione

La soluzione del problema di regressione lineare con il metodo del $\mathcal{X}^2$ è individuata dall'equazione normale (30.4)

$$Fa = \Phi\sigma^{-1}y$$

nella quale la scelta delle funzioni di base

$$\phi_1(x) = 1 \qquad \phi_2(x) = x$$

e le definizioni (30.2) porgono per la matrice Φ le relazioni

$$\Phi_{1i} = \frac{1}{\sigma_i} \qquad \Phi_{2i} = \frac{x_i}{\sigma_i} \qquad \forall\, i = 1,\ldots,n$$

e per $F = \Phi\Phi^T$ — matrice 2×2 —

$$\begin{aligned}
F_{11} &= \sum_{i=1}^{n} \Phi_{1i}\Phi_{1i} = \sum_{i=1}^{n} \frac{1}{\sigma_i^2} = S \\
F_{12} = F_{21} &= \sum_{i=1}^{n} \Phi_{1i}\Phi_{2i} = \sum_{i=1}^{n} \frac{1}{\sigma_i}\frac{x_i}{\sigma_i} = \sum_{i=1}^{n} \frac{x_i}{\sigma_i^2} = S_x \\
F_{22} &= \sum_{i=1}^{n} \Phi_{2i}\Phi_{2i} = \sum_{i=1}^{n} \frac{x_i}{\sigma_i}\frac{x_i}{\sigma_i} = \sum_{i=1}^{n} \frac{x_i^2}{\sigma_i^2} = S_{xx} .
\end{aligned}$$

La matrice F vale dunque

$$F = \begin{pmatrix} S & S_x \\ S_x & S_{xx} \end{pmatrix}$$

con determinante

$$\Delta = SS_{xx} - S_x^2$$

sempre positivo per via della diseguaglianza di Cauchy-Schwarz

$$S_x^2 = \Big(\sum_{i=1}^n \frac{x_i}{\sigma_i^2}\Big)^2 = \Big(\sum_{i=1}^n \frac{x_i}{\sigma_i}\frac{1}{\sigma_i}\Big)^2 \le \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2}\sum_{i=1}^n \frac{1}{\sigma_i^2} = S_{xx}S$$

nella quale il segno di eguaglianza vale se e soltanto se

$$\frac{x_i}{\sigma_i} = \lambda\frac{1}{\sigma_i} \qquad \forall\, i = 1,\ldots,n \qquad (\lambda \text{ costante})$$

e pertanto non può ricorrere nell'ipotesi di x_i distinti. La soluzione dell'equazione normale è individuata univocamente dall'inversa di F

$$F^{-1} = \frac{1}{\Delta}\begin{pmatrix} S_{xx} & -S_x \\ -S_x & S \end{pmatrix}$$

e risulta

$$\begin{pmatrix} a \\ b \end{pmatrix} = F^{-1}\Phi\sigma^{-1}y = \frac{1}{\Delta}\begin{pmatrix} S_{xx} & -S_x \\ -S_x & S \end{pmatrix}\begin{pmatrix} \displaystyle\sum_{i=1}^n \frac{1}{\sigma_i}\frac{y_i}{\sigma_i} \\ \displaystyle\sum_{i=1}^n \frac{x_i}{\sigma_i}\frac{y_i}{\sigma_i} \end{pmatrix} =$$

$$= \frac{1}{\Delta}\begin{pmatrix} S_{xx} & -S_x \\ -S_x & S \end{pmatrix}\begin{pmatrix} S_y \\ S_{xy} \end{pmatrix} = \frac{1}{\Delta}\begin{pmatrix} S_{xx}S_y - S_xS_{xy} \\ SS_{xy} - S_xS_y \end{pmatrix}$$

come affermato. Il teorema 31 stabilisce che le stime a e b costituiscono una coppia di variabili gaussiane bivariate con matrice di covarianza

$$C = F^{-1} = \frac{1}{\Delta}\begin{pmatrix} S_{xx} & -S_x \\ -S_x & S \end{pmatrix}$$

e conseguentemente varianze e covarianza assumono la forma

$$\mathrm{var}(a) = \frac{S_{xx}}{\Delta} \qquad \mathrm{var}(b) = \frac{S}{\Delta} \qquad \mathrm{cov}(a,b) = -\frac{S_x}{\Delta}$$

per cui il coefficiente di correlazione si riduce a

$$\mathrm{corr}(a,b) = \frac{\mathrm{cov}(a,b)}{\sqrt{\mathrm{var}(a)\,\mathrm{var}(b)}} = -\frac{S_x}{\sqrt{S_{xx}}}\,.$$

Questo risultato implica che per $S_x \neq 0$ le variabili casuali a e b non sono stocasticamente indipendenti. L'espressione per la NSSAR non è altro che la specializzazione della (32.1) al modello di regressione qui considerato

$$\mathrm{NSSAR} = \sum_{i=1}^n \frac{1}{\sigma_i^2}(a + bx_i - y_i)^2$$

e che si tratti di una variabile di χ^2 a $n-2$ gradi di libertà segue dal teorema 32 con $p = \mathrm{rank}(F) = \mathrm{rank}(\Phi) = 2$. Il teorema 33 se assicura infine l'indipendenza stocastica sia da a che da b.

39. Retta di regressione, modello alternativo

Dato il modello di regressione lineare

$$\zeta_i = \mu + \kappa(x_i - \overline{x}) + \varepsilon_i \qquad i = 1, \ldots, n$$

con $\overline{x}$ *media pesata delle ascisse*

$$\overline{x} = \Big(\sum_{i=1}^{n} \frac{1}{\sigma_i^2}\Big)^{-1} \sum_{i=1}^{n} \frac{1}{\sigma_i^2} x_i \,,$$

le stime di chi-quadrato dei parametri μ *e* κ *sono date da*

$$m = \Big(\sum_{i=1}^{n} \frac{1}{\sigma_i^2}\Big)^{-1} \sum_{i=1}^{n} \frac{1}{\sigma_i^2} y_i \qquad q = \Big(\sum_{i=1}^{n} \frac{1}{\sigma_i^2}(x_i - \overline{x})^2\Big)^{-1} \sum_{i=1}^{n} \frac{1}{\sigma_i^2}(x_i - \overline{x}) y_i$$

e costituiscono una coppia di variabili gaussiane multivariate con matrice di covarianza

$$C = \begin{pmatrix} \Big(\sum_{i=1}^{n} \frac{1}{\sigma_i^2}\Big)^{-1} & 0 \\ 0 & \Big(\sum_{i=1}^{n} \frac{1}{\sigma_i^2}(x_i - \overline{x})^2\Big)^{-1} \end{pmatrix},$$

risultando perciò stocasticamente indipendenti. La somma normalizzata dei quadrati degli scarti attorno alla regressione assume la forma

$$\mathrm{NSSAR} = \sum_{i=1}^{n} \frac{1}{\sigma_i^2} [m + q(x_i - \overline{x}) - y_i]^2$$

e risulta una variabile di $\mathcal{X}^2$ *a* $n-2$ *gradi di libertà, stocasticamente indipendente tanto da* m *quanto da* q.

Dimostrazione

Le funzioni di base sono in questo caso

$$\phi_1(x) = 1 \qquad \phi_2(x) = x - \overline{x}$$

in modo che le definizioni (30.2) porgono per la matrice Φ le relazioni

$$\Phi_{1i} = \frac{\Phi_1(x_i)}{\sigma_i} = \frac{1}{\sigma_i} \qquad \Phi_{2i} = \frac{\Phi_2(x_i)}{\sigma_i} = \frac{x_i - \overline{x}}{\sigma_i} \qquad \forall\, i = 1, \ldots, n$$

e per la matrice F

$$F_{11} = \sum_{i=1}^{n} \Phi_{1i}\Phi_{1i} = \sum_{i=1}^{n} \frac{1}{\sigma_i^2}$$

$$F_{12} = F_{21} = \sum_{i=1}^{n} \Phi_{1i}\Phi_{2i} = \sum_{i=1}^{n} \frac{x_i - \overline{x}}{\sigma_i^2} = \sum_{i=1}^{n} \frac{x_i}{\sigma_i^2} - \sum_{i=1}^{n} \frac{1}{\sigma_i^2}\,\overline{x} = 0$$

$$F_{22} = \sum_{i=1}^{n} \Phi_{2i}\Phi_{2i} = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{\sigma_i^2} \,.$$

La matrice F assume dunque la forma diagonale

$$F = \begin{pmatrix} \displaystyle\sum_{i=1}^{n} \frac{1}{\sigma_i^2} & 0 \\ 0 & \displaystyle\sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{\sigma_i^2} \end{pmatrix}$$

e siccome gli elementi diagonali sono sempre strettamente positivi, l'uno per la positività delle varianze σ_i^2 e l'altro per l'essere gli x_i distinti, ne deriva che la F è sempre invertibile, con inversa

$$F^{-1} = \begin{pmatrix} \displaystyle\Big(\sum_{i=1}^{n} \frac{1}{\sigma_i^2}\Big)^{-1} & 0 \\ 0 & \displaystyle\Big(\sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{\sigma_i^2}\Big)^{-1} \end{pmatrix}.$$

L'equazione normale (30.4) individua univocamente la soluzione di chi-quadrato

$$\begin{pmatrix} m \\ q \end{pmatrix} = F^{-1}\Phi\sigma^{-1}y = \begin{pmatrix} \displaystyle\Big(\sum_{i=1}^{n} \frac{1}{\sigma_i^2}\Big)^{-1} & 0 \\ 0 & \displaystyle\Big(\sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{\sigma_i^2}\Big)^{-1} \end{pmatrix} \begin{pmatrix} \displaystyle\sum_{i=1}^{n} \frac{1}{\sigma_i^2} y_i \\ \displaystyle\sum_{i=1}^{n} \frac{(x_i - \overline{x}) y_i}{\sigma_i^2} \end{pmatrix} =$$

$$= \begin{pmatrix} \displaystyle\Big(\sum_{i=1}^{n} \frac{1}{\sigma_i^2}\Big)^{-1} \sum_{i=1}^{n} \frac{1}{\sigma_i^2} y_i \\ \displaystyle\Big(\sum_{i=1}^{n} \frac{1}{\sigma_i^2}(x_i - \overline{x})^2\Big)^{-1} \sum_{i=1}^{n} \frac{1}{\sigma_i^2}(x_i - \overline{x}) y_i \end{pmatrix}.$$

La matrice F^{-1} coincide anche con la matrice di covarianza delle variabili m, q e la sua struttura diagonale implica l'indipendenza stocastica delle stesse variabili. La sostituzione in (32.1) del modello statistico qui considerato conduce a

$$\text{NSSAR} = \sum_{i=1}^{n} \frac{1}{\sigma_i^2} [m + q(x_i - \overline{x}) - y_i]^2 .$$

L'indipendenza stocastica di NSSAR, m e q, come pure il fatto che NSSAR sia una variabile di $\mathcal{X}^2$ a $n-2$ gradi di libertà, si dimostra come nel teorema precedente. □

40. Retta di regressione. Intervalli di confidenza per i coefficienti a e b

Dato il modello di regressione lineare a due parametri

$$\zeta_i = \alpha + \beta x_i + \varepsilon_i \qquad i = 1, \ldots, n$$

e introdotte le notazioni di cui al teorema 38, *gli intervalli di confidenza per le stime a e b dei parametri α e β sono dati da*

$$a - t_{[1-\frac{\varepsilon}{2}],(n-2)} \sqrt{\frac{S_{xx}}{\Delta} \frac{\text{NSSAR}}{n-2}} \leq \alpha \leq a + t_{[1-\frac{\varepsilon}{2}],(n-2)} \sqrt{\frac{S_{xx}}{\Delta} \frac{\text{NSSAR}}{n-2}}$$

$$b - t_{[1-\frac{\varepsilon}{2}],(n-2)} \sqrt{\frac{S}{\Delta} \frac{\mathrm{NSSAR}}{n-2}} \leq \beta \leq b + t_{[1-\frac{\varepsilon}{2}],(n-2)} \sqrt{\frac{S}{\Delta} \frac{\mathrm{NSSAR}}{n-2}}$$

con livello di confidenza $1-\varepsilon \in (0,1)$. Nel caso di eguali varianze $\sigma_i^2 = \sigma^2$ gli intervalli di confidenza sono indipendenti da σ.

Dimostrazione
È sufficiente inserire le espressioni esplicite delle varianze $(F^{-1})_{jj}$

$$(F^{-1})_{11} = \frac{S_{xx}}{\Delta} \qquad (F^{-1})_{22} = \frac{S}{\Delta}$$

e porre $p = 2$ nelle relazioni generali (34.1)

$$a_j - t_{[1-\frac{\varepsilon}{2}],(n-p)} \sqrt{(F^{-1})_{jj} \frac{\mathrm{NSSAR}}{n-p}} \leq \alpha_j \leq a_j + t_{[1-\frac{\varepsilon}{2}],(n-p)} \sqrt{(F^{-1})_{jj} \frac{\mathrm{NSSAR}}{n-p}}$$

per ottenere il risultato richiesto. Il teorema 34 garantisce che in caso di eguali varianze $\sigma_i^2 = \sigma^2$ gli intervalli di confidenza non dipendano esplicitamente da σ. Si ricorda che le variabili gaussiane a e b non sono stocasticamente indipendenti, per cui la probabilità che i valori medi α e β appartengano simultaneamente ai rispettivi intervalli di confidenza **non coincide** con il prodotto $(1-\varepsilon)^2$ dei singoli livelli di confidenza. □

41. Retta di regressione. Intervalli di confidenza per i coefficienti m e q

Dato il modello di regressione lineare

$$\zeta_i = \mu + \kappa(x_i - \overline{x}) + \varepsilon_i \qquad i = 1, \ldots, n$$

e introdotte le notazioni di cui al teorema 39, gli intervalli di confidenza per le stime m e q dei parametri μ e κ si scrivono

$$m - t_{[1-\frac{\alpha}{2}],(n-2)} \sqrt{\Big(\sum_{i=1}^{n} \frac{1}{\sigma_i^2}\Big)^{-1} \frac{\mathrm{NSSAR}}{n-2}} \leq \mu \leq$$
$$\leq m + t_{[1-\frac{\alpha}{2}],(n-2)} \sqrt{\Big(\sum_{i=1}^{n} \frac{1}{\sigma_i^2}\Big)^{-1} \frac{\mathrm{NSSAR}}{n-2}}$$

$$q - t_{[1-\frac{\alpha}{2}],(n-2)} \sqrt{\Big(\sum_{i=1}^{n} \frac{(x_i-\overline{x})^2}{\sigma_i^2}\Big)^{-1} \frac{\mathrm{NSSAR}}{n-2}} \leq \kappa \leq$$
$$\leq q + t_{[1-\frac{\alpha}{2}],(n-2)} \sqrt{\Big(\sum_{i=1}^{n} \frac{(x_i-\overline{x})^2}{\sigma_i^2}\Big)^{-1} \frac{\mathrm{NSSAR}}{n-2}}$$

con livello di confidenza $1-\alpha \in (0,1)$. Nel caso di eguali varianze $\sigma_i^2 = \sigma^2$ gli intervalli di confidenza sono indipendenti da σ.

Dimostrazione
Al risultato si perviene come nel caso precedente, ricordando che che

$$(F^{-1})_{11} = \Big(\sum_{i=1}^{n} \frac{1}{\sigma_i^2}\Big)^{-1} \qquad \text{e} \qquad (F^{-1})_{22} = \Big(\sum_{i=1}^{n} \frac{(x_i-\overline{x})^2}{\sigma_i^2}\Big)^{-1}$$

e utilizzando la (34.1). L'indipendenza da σ degli intervalli di confidenza nel caso di eguali varianze $\sigma_i^2 = \sigma^2$ si deduce immediatamente dal teorema 34. È importante sottolineare che in questo caso le variabili m e q sono stocasticamente indipendenti e che pertanto i valori veri μ e κ appartengono simultaneamente ai rispettivi intervalli di confidenza con probabilità $(1-\alpha)^2$. $\square$

42. Retta di regressione. Intervallo di confidenza per le previsioni

Dato il modello di regressione lineare

$$\zeta_i = \mu + \kappa(x_i - \overline{x}) + \varepsilon_i \qquad i = 1, \ldots, n$$

e $x_0 \in \mathbb{R}$, si assuma che il valore della variabile y corrispondente all'ascissa $x = x_0$ sia descritto dalla variabile casuale

$$\zeta_0 = \mu + \kappa(x_0 - \overline{x}) + \varepsilon_0$$

con ε_0 variabile gaussiana di media nulla e varianza σ_0^2, indipendente dalle ε_i del modello. Si definisca la **previsione** *del valore di y in x_0 in base alla regressione lineare come*

$$y_0 = m + q(x_0 - \overline{x}) .$$

Allora la previsione y_0 è una variabile gaussiana di media ζ_0 e varianza

$$\mathbb{E}[(y_0-\zeta_0)^2] = \Big(\sum_{i=1}^{n} \frac{1}{\sigma_i^2}\Big)^{-1} + \Big(\sum_{i=1}^{n} \frac{1}{\sigma_i^2}(x_i-\overline{x})^2\Big)^{-1}(x_0-\overline{x})^2 + \sigma_0^2$$

e l'intervallo di confidenza di ζ_0, con livello di confidenza $1-\alpha \in (0,1)$, assume la forma

$$\begin{aligned} y_0 - t_{[1-\frac{\alpha}{2}],(n-2)} \sqrt{\mathbb{E}[(y_0-\zeta_0)^2]} \sqrt{\frac{\text{NSSAR}}{n-2}} \le \zeta_0 \le \\ \le y_0 + t_{[1-\frac{\alpha}{2}],(n-2)} \sqrt{\mathbb{E}[(y_0-\zeta_0)^2]} \sqrt{\frac{\text{NSSAR}}{n-2}} . \end{aligned} \tag{42.1}$$

Qualora la varianza sia indipendente dall'ascissa — $\sigma_i = \sigma \ \forall\, i = 1, \ldots, n$ e $\sigma_0 = \sigma$ — il precedente intervallo di confidenza è a sua volta indipendente da σ

$$y_0 - t_{[1-\frac{\alpha}{2}],(n-2)} \sqrt{V} \sqrt{\frac{SS}{n-2}} \le \zeta_0 \le y_0 + t_{[1-\frac{\alpha}{2}],(n-2)} \sqrt{V} \sqrt{\frac{SS}{n-2}} \tag{42.2}$$

con

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad SS = \sum_{i=1}^{n}\big[-y_i + m + q(x_i - \overline{x})\big]^2$$

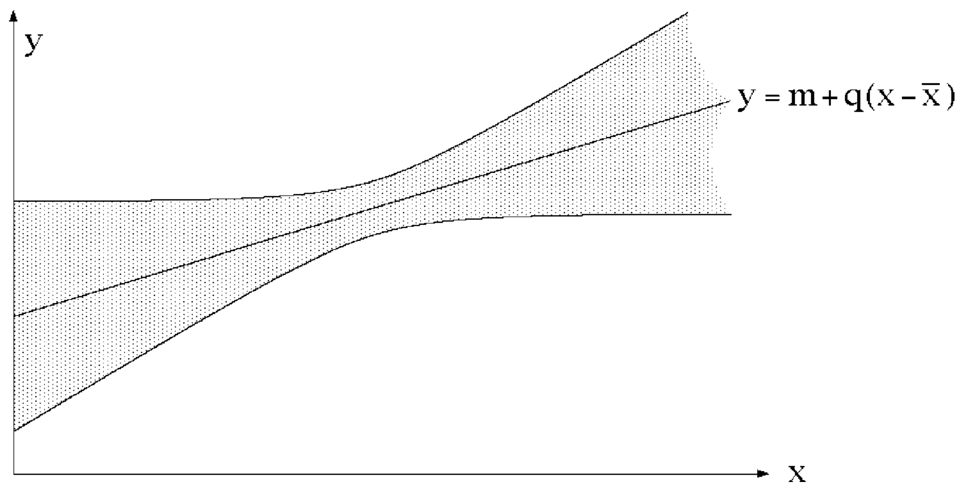
$$V = 1 + \frac{1}{n} + \frac{1}{\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})^2}(x_0 - \overline{x})^2 .$$

Dimostrazione
Si tratta di specializzare al modello qui considerato i risultati del teorema 35. □

Osservazione
Giova sottolineare che al variare di $x_0 \in \mathbb{R}$ l'intervallo di confidenza (42.2) descrive una regione del piano (x, y) del tipo

$$\Big\{(x,y) \in \mathbb{R}^2 \;:\; m + q(x-\overline{x}) - c\sqrt{1+g(x-\overline{x})^2} \le y \le m + q(x-\overline{x}) - c\sqrt{1+g(x-\overline{x})^2}\Big\}$$

con c e g costanti positive opportune. Questa **regione di confidenza** costituisce un intorno della retta di regressione e il suo aspetto qualitativo è illustrato col tratteggio nella figura seguente



Con l'aumentare della dispersione delle x_i attorno al loro valore medio $\overline{x}$, ovvero di

$$\sum_{i=1}^{n}(x_i - \overline{x})^2$$

si verifica una riduzione del coefficiente V e quindi l'estensione della regione di confidenza diminuisce. Ciò riflette l'idea intuitiva che a parità di ogni altra condizione l'incertezza nel posizionamento della retta di regressione debba essere maggiore per valori x_i molto ravvicinati e minore nel caso di elevata dispersione dei dati in x.

43. Singular Value Decomposition di una matrice $m \times n$

La SVD di una matrice A è una riscrittura di A come prodotto

$$\begin{array}{ccccc} A & = & V & \begin{pmatrix} \Sigma & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{pmatrix} & U^T \\ m \times n & & m \times m & m \times n & n \times n \end{array} \tag{43.1}$$

dove V e U sono matrici ortogonali e Σ è una matrice diagonale con elementi positivi (i **valori singolari** positivi della matrice A).

Precisamente, l'esistenza della SVD per una matrice $A \neq 0$ arbitraria è assicurata nei termini seguenti:

- U è una matrice $n \times n$ le cui colonne $u_1, \ldots, u_n$ costituiscono una base ortonormale di autovettori della matrice $n \times n$, simmetrica e semidefinita positiva, $A^T A$

$$U = (u_1 \; \ldots \; u_n) \qquad A^T A u_i = \sigma_i^2 u_i \quad \forall \, i = 1, \ldots, n$$

$$\sigma_1^2 \geq \sigma_2^2 \geq \ldots \geq \sigma_r^2 > \sigma_{r+1}^2 = \ldots = \sigma_n^2 = 0$$

- Σ è una matrice diagonale $r \times r$, con $r = \mathrm{rank}(A) \leq \min(m,n)$,

$$\Sigma = \begin{pmatrix} \sigma_1 & & & \mathbb{O} \\ & \sigma_2 & & \\ & & \ddots & \\ \mathbb{O} & & & \sigma_r \end{pmatrix} \tag{43.2}$$

  i cui elementi diagonali sono le radici quadrate aritmetiche degli autovalori positivi di $A^T A$ (**valori singolari** positivi della matrice A)

- V è una matrice $m \times m$ le cui prime r colonne sono date dai vettori ortonormali

$$v_i = \frac{1}{|Au_i|_2} A u_i \qquad \forall i = 1, \ldots, r \tag{43.3}$$

  mentre le eventuali $m - r$ colonne residue si scelgono in modo da generare una base ortonormale di $\mathbb{R}^m$

Il caso $A = 0$ è di scarso interesse, in quanto la SVD può essere scritta ponendo $\Sigma = 0$ e facendo uso di matrici ortogonali U e V arbitrarie.

Base ortonormale di $A^T A$

L'esistenza di una base ortonormale $u_1, \ldots, u_n$ di autovettori di $A^T A$ segue immediatamente dal carattere simmetrico della matrice. Trattandosi poi di matrice semidefinita positiva

$$u^T A^T A u = (Au)^T A u = |Au|_2^2 \geq 0 \qquad \forall \, u \in \mathbb{R}^n ,$$

i relativi autovalori sono tutti reali e non negativi, rappresentabili come quadrati di quantità reali non negative

$$A^T A u_i = \sigma_i^2 u_i \quad \forall\, i = 1, \ldots, n$$

e ordinabili in senso decrescente

$$\sigma_1^2 \geq \sigma_2^2 \geq \ldots \geq \sigma_r^2 > \sigma_{r+1}^2 = \ldots = \sigma_n^2 = 0 \,.$$

Nella precedente espressione il parametro r individua il numero di autovalori non nulli della matrice $A^T A$ e di conseguenza $n - r$ rappresenta la molteplicità dell'autovalore zero. Si ha perciò

$$\dim \ker(A^T A) = n - r$$

e dunque

$$\mathrm{rank}(A^T A) = \dim \mathrm{range}(A^T A) = n - \dim \ker(A^T A) = r \,.$$

Il paramtro r si identifica pertanto con il rango della matrice simmetrica $A^T A$.

Il rango di $A^T A$ coincide con quello di A

Conviene dimostrare l'affermazione provando l'identità dei nuclei di A e $A^T A$. Se $x \in \ker(A^T A)$, allora

$$A^T A x = 0$$

per cui

$$x^T A^T A x = 0 \qquad \Longleftrightarrow \qquad |Ax|_2 = 0 \qquad \Longleftrightarrow \qquad Ax = 0 \qquad \Longleftrightarrow \qquad x \in \ker(A)$$

e quindi

$$\ker(A^T A) \subseteq \ker(A) \,.$$

Viceversa, per ogni $x \in \ker(A)$ vale

$$Ax = 0 \qquad \Longrightarrow \qquad A^T A x = A^T 0 = 0 \qquad \Longrightarrow \qquad x \in \ker(A^T A)$$

e dunque

$$\ker(A) \subseteq \ker(A^T A) \,.$$

Dalla reciproca inclusione dei due nuclei si conclude che

$$\ker(A) = \ker(A^T A) \subseteq \mathbb{R}^n$$

per cui

$$\dim \ker(A) = \dim \ker(A^T A)$$

ed infine

$$\begin{aligned} \mathrm{rank}(A) &= \dim \mathrm{range}(A) = n - \dim \ker(A) = \\ &= n - \dim \ker(A^T A) = \dim \mathrm{range}(A^T A) = \mathrm{rank}(A^T A) = r \,. \end{aligned}$$

Di qui segue, in particolare, che

$$r \leq \min(m, n) \,.$$

Le prime r colonne di V sono ortonormali
È immediato verificare che le prime r colonne della matrice V, definite secondo l'equazione (43.3), costituiscono un sistema ortonormale. Per ogni $i, j = 1, \ldots, r$ si ha infatti

$$\begin{aligned} v_j^T v_i \,&=\, \frac{1}{|Au_j|_2}(Au_j)^T \frac{1}{|Au_i|_2} Au_i \,=\, \frac{1}{|Au_j|_2}\,\frac{1}{|Au_i|_2}\, u_j^T A^T A u_i \,= \\ &=\, \frac{1}{|Au_j|_2}\,\frac{1}{|Au_i|_2}\,\sigma_i^2\, u_j^T u_i \,=\, \frac{1}{|Au_j|_2}\,\frac{1}{|Au_i|_2}\,\sigma_i^2\,\delta_{ij} \,=\, \frac{1}{|Au_i|_2^2}\,\sigma_i^2\,\delta_{ij} \end{aligned}$$

e in forza della relazione

$$|Au_i|_2^2 \,=\, u_i^T A^T A u_i \,=\, u_i^T \sigma_i^2 u_i \,=\, \sigma_i^2$$

si conclude quanto affermato

$$v_j^T v_i \,=\, \delta_{ij} \qquad \forall\, i, j = 1, \ldots, r\,.$$

Le colonne di V come base ortonormale di $\mathbb{R}^m$
Si è verificato che le prime $r < \min(m, n)$ colonne della matrice V costituiscono un sistema ortonormale, sebbene in generale non completo, di $\mathbb{R}^m$. Ai vettori v_i, $i = 1, \ldots, r$ è sempre possibile aggiungere $m - r$ vettori residui w_i, $i = r + 1, \ldots, m$, in modo che il sistema di vettori

$$\{v_1,\, \ldots, v_r,\, w_{r+1} \ldots, w_m\}$$

rappresenti una base di $\mathbb{R}^m$. Applicando a questo set di vettori l'algoritmo di ortonormalizzazione di Gram-Schmidt, si perviene infine ad una base ortonormale di $\mathbb{R}^m$ i cui primi r elementi coincidono con quelli forniti dalla (43.3).

Per ogni matrice A vale la SVD (43.1)
Basta verificare che le matrici a primo e secondo membro nella (43.1) conducono allo stesso risultato se moltiplicate a sinistra per un qualsiasi vettore della base ortonormale $\{u_i,\, i = 1, \ldots, n\}$. Per ogni $i = 1, \ldots, n$ fissato, risulta in effetti

$$V \begin{pmatrix} \Sigma & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{pmatrix} U^T u_i \,=\, V \begin{pmatrix} \Sigma & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{pmatrix} \begin{pmatrix} u_1^T \\ \hline \vdots \\ \hline u_n^T \end{pmatrix} u_i \,=\, V \begin{pmatrix} \Sigma & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \leftarrow i\text{-esima riga}$$

e quindi

$$V \begin{pmatrix} \Sigma & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{pmatrix} U^T u_i \,=\, V \begin{pmatrix} 0 \\ \vdots \\ \sigma_i \\ \vdots \\ 0 \end{pmatrix} \leftarrow i\text{-esima riga}$$

sicché per $i = 1, \ldots, r$ vale

$$V \begin{pmatrix} \Sigma & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{pmatrix} U^T u_i \,=\, \sigma_i v_i \,=\, \sigma_i \frac{1}{|Au_i|_2} Au_i \,=\, Au_i$$

mentre per $i = r+1, \ldots, n$ si ottiene

$$V \begin{pmatrix} \Sigma & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{pmatrix} U^T u_i \,=\, V0 \,=\, 0 \,=\, Au_i$$

in quanto $u_i \in \ker(A^T A) \,=\, \ker(A) \ \ \forall\, i = r+1, \ldots, n$. In definitiva

$$V \begin{pmatrix} \Sigma & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{pmatrix} U^T u_i \,=\, Au_i \qquad \forall\, i = 1, \ldots, n$$

e l'asserto è dimostrato.

44. Pseudo inversa di Moore-Penrose di una matrice $m \times n$

Siano dati la matrice A, con SVD nota

$$A \,=\, V \begin{pmatrix} \Sigma & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{pmatrix} U^T ,$$

ed un vettore $b \in \mathbb{R}^m$. Allora il sistema lineare

$$Ax \,=\, b$$

ammette la soluzione nel senso dei minimi quadrati

$$x \,=\, U \begin{pmatrix} \Sigma^{-1} & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{pmatrix} V^T b \,=\, A^+ b \tag{44.1}$$

essendo

$$A^+ = U \begin{pmatrix} \Sigma^{-1} & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{pmatrix} V^T \tag{44.2}$$

la **pseudo-inversa di Moore-Pensore** *della matrice A. Qualora la soluzione secondo i minimi quadrati non sia unica, la relazione (44.1) individua quella con il minimo modulo in norma 2.*

Determinare una soluzione x di $Ax = b$ nel senso dei minimi quadrati significa individuare x come minimo assoluto del funzionale

$$\mathfrak{L}(x) \,=\, |Ax - b|_2^2$$

che sostituendo la SVD di A diventa

$$\mathfrak{L}(x) \,=\, \left| V \begin{pmatrix} \Sigma & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{pmatrix} U^T x - b \right|_2^2$$

e ricordando che V^T è una matrice ortonormale si riduce a

$$\mathfrak{L}(x) = \left| \begin{pmatrix} \Sigma & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{pmatrix} U^T x - V^T b \right|_2^2 .$$

Conviene porre, per brevità

$$U^T x = \begin{pmatrix} z \\ z' \end{pmatrix} \qquad z \in \mathbb{R}^r \ , \quad z' \in \mathbb{R}^{n-r}$$

e

$$V^T b = \begin{pmatrix} c \\ c' \end{pmatrix} \qquad c \in \mathbb{R}^r \ , \quad c' \in \mathbb{R}^{m-r}$$

con i vettori z' e c' rispettivamente definiti per $\mathrm{rank}(A) = r < n$ e $\mathrm{rank}(A) = r < m$. In tal modo il funzionale $\mathfrak{L}$ assume la forma

$$\mathfrak{L}(x) = \left| \begin{pmatrix} \Sigma & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{pmatrix} \begin{pmatrix} z \\ z' \end{pmatrix} - \begin{pmatrix} c \\ c' \end{pmatrix} \right|_2^2 = \left| \begin{pmatrix} \Sigma z \\ 0 \end{pmatrix} - \begin{pmatrix} c \\ c' \end{pmatrix} \right|_2^2 = \left| \begin{pmatrix} \Sigma z - c \\ -c' \end{pmatrix} \right|_2^2 =$$

$$= \sum_{i=1}^{r} (\sigma_i z_i - c_i)^2 + \sum_{i=r+1}^{m} (c_i')^2 .$$

I minimi assoluti si hanno per z' arbitrario e

$$\sigma_i z_i - c_i = 0 \quad \forall\, i = 1, \ldots, r \qquad \Longleftrightarrow \qquad \Sigma z = c \qquad \Longleftrightarrow \qquad z = \Sigma^{-1} c$$

ossia per

$$U^T x = \begin{pmatrix} \Sigma^{-1} c \\ z' \end{pmatrix} \qquad \forall\, z' \in \mathbb{R}^{n-r} .$$

I punti di minimo assoluto del funzionale $\mathfrak{L}$ sono quindi individuati tutti e soltanto dalla condizione

$$x = U \begin{pmatrix} \Sigma^{-1} c \\ z' \end{pmatrix} \qquad \forall\, z' \in \mathbb{R}^{n-r} , \tag{44.3}$$

con

$$x^T x = \big((\Sigma^{-1} c)^T \quad z'^T \big)\, U^T U \begin{pmatrix} \Sigma^{-1} c \\ z' \end{pmatrix} = (\Sigma^{-1} c)^T \Sigma^{-1} c + z'^T z' .$$

Di tali soluzioni quella di minimo modulo (in norma 2) si ricava per $z' = 0$

$$x = U \begin{pmatrix} \Sigma^{-1} c \\ 0 \end{pmatrix} = U \begin{pmatrix} \Sigma^{-1} & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{pmatrix} \begin{pmatrix} c \\ c' \end{pmatrix} = U \begin{pmatrix} \Sigma^{-1} & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{pmatrix} V^T b = A^+ b .$$

A causa della (44.3) la soluzione è inoltre unica se e soltanto se $r = n$, ossia per

$$m \geq n \qquad \text{e} \qquad \mathrm{rank}(A) = n \quad \text{(rango massimo)} .$$

45. Principal Component Analysis (PCA)

Organizzazione dei dati per la PCA

La Principal Component Analysis si applica allo studio di un set di dati organizzato secondo una matrice X a n righe e p colonne. Tipicamente, le n righe della matrice corrispondono a diversi campioni, mentre i p valori di ogni riga rappresentano i valori di p grandezze assegnate relativamente al campione considerato.

Per esempio, si può immaginare di sottoporre un materiale a vari trattamenti superficiali e di analizzare la susseguente presenza di impurezze alla sua superficie mediante spettrometria XPS (X-ray Photoelectron Spectroscopy). La matrice X dei dati avrà elementi

$$X_{ij} \qquad i = 1, \ldots, n \quad j = 1, \ldots, p$$

l'indice i individuerà i diversi trattamenti (e/o le eventuali prove ripetute con lo stesso tipo di trattamento), mentre j potrà specificare un particolare picco (o bin) fra i p caratteristici dello spettro XPS. L'elemento di matrice X_{ij} potrà allora intendersi come l'intensità misurata sul campione i (superficie trattata) del picco j.

Un altro esempio di applicazione della PCA può essere lo studio della frequenza di particolari sequenze genetiche nel DNA o di specifici caratteri (gruppo sanguigno, fattore Rh, ecc.) nelle popolazioni di determinate aree geografiche. Nella matrice X dei dati, l'indice i specifica l'area geografica rispetto ad una mappa assegnata, mentre j corrisponde al carattere esaminato: X_{ij} esprimerà la frequenza osservata del carattere j presso la popolazione ubicata nell'area i.

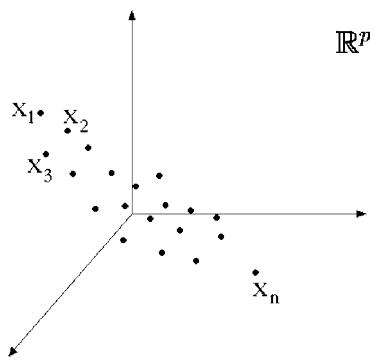
Una ulteriore applicazione può essere costituita dall'analisi dei dati spettroscopici (nell'ottico o nell'infrarosso, ad esempio) di diverse soluzioni. L'indice i individua la soluzione, j specifica un picco caratteristico dello spettro di assorbimento e X_{ij} è l'intensità del picco di assorbimento j per la soluzione i.

Si può assumere che ogni riga di X corrisponda ad un punto nello spazio $\mathbb{R}^p$ a p componenti, una per ogni colonna. La matrice X individua allora un set di n punti in $\mathbb{R}^p$

$$X = \begin{pmatrix} x_1^T \\ \hline x_2^T \\ \hline \vdots \\ \hline x_n^T \end{pmatrix} \qquad x_i \in \mathbb{R}^p \quad \forall\, i = 1, \ldots, n\,.$$

Sottospazi vettoriali di $\mathbb{R}^p$

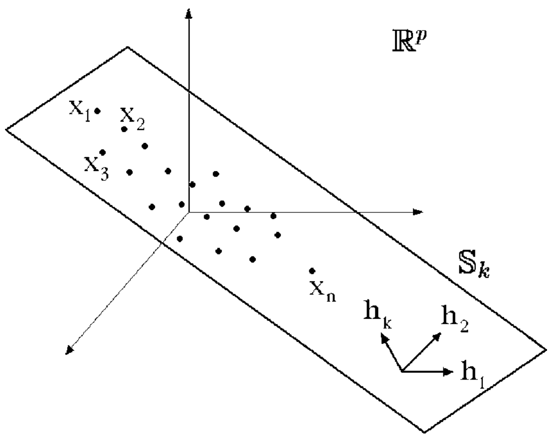
Può verificarsi che i punti x_i non siano distribuiti a caso in $\mathbb{R}^p$, ma si localizzino lungo un particolare sottoinsieme di $\mathbb{R}^p$. Ciò comporta il sussistere di una correlazione fra i dati x_i, $i = 1, \ldots, n$. Il sottoinsieme in questione potrebbe presentare una struttura più o meno complessa, ma l'ipotesi più semplice consiste nel ritenere che si tratti di un sottospazio vettoriale di $\mathbb{R}^p$.



In particolare, si assuma che i dati siano descritti da uno spazio vettoriale $\mathbb{S}_k$, di dimensione $k \leq p$, generato da un sistema ortonormale di vettori di $\mathbb{R}^p$

$$\{h_1 \ , \ h_2 \ , \ \ldots \ , \ h_k\}$$

per i quali risulta quindi $h_j^T h_{j'} = \delta_{jj'} \ \forall j, j' = 1, \ldots, k$.



Tali vettori si potranno identificare con le colonne di una matrice $p \times k$

$$H = \Big(h_1 \Big| h_2 \Big| \ldots \Big| h_k\Big)$$

per la quale si avrà $H^T H = \mathbb{I}$.

Proiezioni ortogonali sul sottospazio
Le proiezioni ortogonali dei vettori x_i su $\mathbb{S}_k$ sono date dai vettori colonna

$$HH^T x_i \qquad \forall i = 1, \ldots, n$$

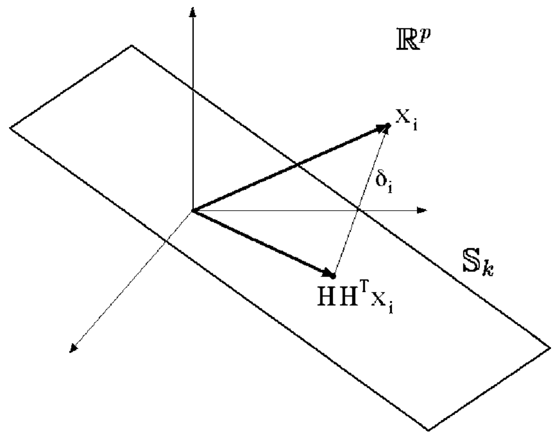
ovvero dai vettori riga

$$x_i^T HH^T \qquad \forall i = 1, \ldots, n$$

per cui la matrice delle proiezioni ortogonali vale

$$XHH^T \, .$$

Le proiezioni ortogonali su $\mathbb{S}_k$ dei vettori x_i si identificano dunque con le righe della matrice XHH^T.



La verifica di queste affermazioni è immediata. Poiché lo spazio $\mathbb{R}^p$ è dato dalla somma diretta di $\mathbb{S}_k$ e del suo complemento ortogonale

$$\mathbb{R}^p \,=\, \mathbb{S}_k \oplus \mathbb{S}_k^{\perp} \, ,$$

un generico vettore $x \in \mathbb{R}^p$ si scriverà nella forma

$$x \,=\, \sum_{j=1}^{k} \alpha_j h_j \,+\, x - \sum_{j=1}^{k} \alpha_j h_j$$

con $\sum_{j=1}^{k} \alpha_j h_j$ certamente compreso in $\mathbb{S}_k$ e $x - \sum_{j=1}^{k} \alpha_j h_j$ incluso in $\mathbb{S}_k^{\perp}$ a patto di determinare i coefficienti α_j in modo che si abbia

$$h_{j'} \Big(x - \sum_{j=1}^{k} \alpha_j h_j \Big) \,=\, 0 \qquad \forall\, j' = 1, \ldots, k$$

condizione che assicura l'ortogonalità del vettore rispetto ad ogni $h_{j'}$ e dunque rispetto a $\mathbb{S}_k$, che dagli $h_{j'}$ è generato. Per l'ortonormalità dei vettori h_j, detta condizione si traduce in

$$h_{j'}^T x \,=\, \sum_{j=1}^{k} \alpha_j h_{j'}^T h_j \,=\, \sum_{j=1}^{k} \alpha_j \delta_{j'j} \,=\, \alpha_{j'} \qquad \forall\, j' = 1, \ldots, k$$

in modo che la proiezione ortogonale di x su $\mathbb{S}_k$ è data dal vettore colonna

$$\sum_{j=1}^{k} \alpha_j h_j \,=\, \sum_{j=1}^{k} h_j^T x h_j \,=\, \sum_{j=1}^{k} x^T h_j h_j$$

ovvero dal vettore riga

$$\Big(\sum_{j=1}^{k} \alpha_j h_j\Big)^T = \sum_{j=1}^{k} x^T h_j h_j^T = x^T \sum_{j=1}^{k} h_j h_j^T = x^T HH^T .$$

Le proiezioni ortogonali su $\mathbb{S}_k$ dei vettori x_i si identificano perciò con le righe della matrice

$$XHH^T .$$

Somma dei moduli quadrati degli scarti

Le distanze dei punti x_i da $\mathbb{S}_k$ — "scarti" dei dati rispetto al modello — si scrivono

$$\delta_i^T = x_i^T - x_i^T HH^T \qquad i = 1, \ldots, n$$

e si identificano pertanto con le righe di

$$X - XHH^T = X(\mathbb{I} - HH^T) .$$

La somma dei moduli quadrati degli scarti risulta dunque

$$\begin{aligned}
SS &= \sum_{i=1}^{n} \delta_i^T \delta_i = \sum_{i=1}^{n}\sum_{j=1}^{p} (\delta_i)_j (\delta_i)_j = \\
&= \sum_{i=1}^{n}\sum_{j=1}^{p} [X(\mathbb{I} - HH^T)]_{ij} [X(\mathbb{I} - HH^T)]_{ij} = \\
&= \sum_{i=1}^{n}\sum_{j=1}^{p} [X(\mathbb{I} - HH^T)]_{ij} [(\mathbb{I} - HH^T)^T X^T]_{ji} = \\
&= \sum_{i=1}^{n} [X(\mathbb{I} - HH^T)(\mathbb{I} - HH^T)^T X^T]_{ii} = \\
&= \sum_{i=1}^{n} [X(\mathbb{I} - HH^T)(\mathbb{I} - HH^T) X^T]_{ii} = \\
&= \mathrm{tr}[X(\mathbb{I} - HH^T)(\mathbb{I} - HH^T) X^T]
\end{aligned}$$

e poiché

$$X(\mathbb{I} - HH^T)(\mathbb{I} - HH^T)X^T = X(\mathbb{I} - HH^T)^2 X^T = X(\mathbb{I} - HH^T)X^T$$

l'espressione si riduce a

$$\begin{aligned}
SS &= \mathrm{tr}[X(\mathbb{I} - HH^T)X^T] = \\
&= \mathrm{tr}[XX^T] - \mathrm{tr}[XHH^T X^T] = \mathrm{tr}[XX^T] - \mathrm{tr}[H^T X^T XH]
\end{aligned} \tag{45.1}$$

con $X^T X$ matrice $p \times p$ reale simmetrica e semidefinita positiva.

Determinazione del sottospazio vettoriale $\mathbb{S}_k$. Minimi quadrati
Per individuare il sottospazio $\mathbb{S}_k$, ovvero i vettori ortonormali u_j, $j = 1, \ldots, k$ che lo generano, o ancora la matrice H di cui detti vettori costituiscono le colonne, si fa uso del **metodo dei minimi quadrati**: si impone che la matrice H — e quindi il sottospazio $\mathbb{S}_k$, di dimensione k **fissata** — sia scelta in modo da rendere minima la somma dei moduli quadrati degli scarti. Dal momento che la traccia di XX^T è una costante fissata, l'equazione (45.1) implica che il minimo della somma SS si abbia per $\mathrm{tr}[H^T X^T X H]$ **massima**

$$SS \quad \text{minima} \qquad \Longleftrightarrow \qquad \mathrm{tr}[H^T X^T X H] \quad \text{massima}$$

sotto la condizione che le colonne $h_1, \ldots, h_k$ di H siano ortonormali. Si tratta pertanto di un problema di minimi quadrati molto particolare. Vale il seguente

Teorema (Minimo di SS)
Sia data una base ortonormale di autovettori $v_1 \ldots, v_p$, della matrice $p \times p$ reale simmetrica e semidefinita positiva $X^T X$, con

$$X^T X v_j \;=\; \lambda_j v_j \quad \forall\, j = 1, \ldots, p$$

e autovalori ordinati in senso decrescente

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0\,.$$

Il massimo di $\mathrm{tr}[H^T X^T X H]$, con H matrice $p \times k$ a colonne ortonormali, ricorre allora per

$$H \;=\; \big(v_1 \big| v_2 \big| \ldots \big| v_k\big)$$

ovvero

$$\mathbb{S}_k \;=\; \mathrm{span}\{v_1\,,\; v_2\,, \ldots,\; v_k\}$$

e la corrispondente somma dei moduli quadrati degli scarti vale

$$\mathrm{tr}[X(\mathbb{I} - HH^T)X^T] \;=\; \sum_{j=k+1}^{p} \lambda_j\,.$$

Dimostrazione
Poiché la matrice $X^T X$ è reale simmetrica e semidefinita positiva, esiste una matrice ortogonale U che la riduce alla forma diagonale

$$X^T X \;=\; U^T D U \qquad D \;=\; \mathrm{diag}\{\lambda_1, \ldots, \lambda_p\} \;=\; \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_p \end{pmatrix}$$

con elementi diagonali non negativi

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0\,.$$

La traccia di $H^T X^T XH$ diventa allora

$$\mathrm{tr}(H^T X^T XH) \,=\, \mathrm{tr}(H^T U^T DUH) \,=\, \mathrm{tr}[(UH)^T D(UH)] \,=\, \mathrm{tr}(G^T DG) \tag{45.2}$$

in cui la matrice $p \times k$

$$G \,=\, UH \,=\, \big(g_1 \big| g_2 \big| \ldots \big| g_k\big)$$

è composta di k colonne ortonormali in $\mathbb{R}^p$

$$G^T G \,=\, (UH)^T UH \,=\, H^T U^T UH \,=\, H^T U^{-1} UH \,=\, H^T H \,=\, \mathbb{I}\,,$$

sicché la precedente relazione (45.2) si può porre nella forma più esplicita

$$\mathrm{tr}(H^T X^T XH) \,=\, \sum_{j=1}^{k} g_j^T D g_j \,=\, \sum_{j=1}^{k}\sum_{i=1}^{p} (g_j)_i \lambda_i (g_j)_i \,=\, \sum_{j=1}^{k}\sum_{i=1}^{p} \lambda_i [(g_j)_i]^2$$

con $g_j^T g_{j'} \,=\, \sum_{i=1}^{p} (g_j)_i (g_{j'})_i \,=\, \delta_{jj'}$, $\forall\, j, j' \,=\, 1, \ldots, k$. La traccia può quindi intendersi come una media pesata degli autovalori non negativi di $X^T X$

$$\mathrm{tr}(H^T X^T XH) \,=\, \sum_{i=1}^{p} \lambda_i w_i$$

con pesi

$$w_i \,=\, \sum_{j=1}^{k} [(g_j)_i]^2 \ge 0 \qquad \forall\, i = 1, \ldots, p$$

soddisfacenti le condizioni

$$0 \le w_i \,=\, \sum_{j=1}^{k} [(g_j)_i]^2 \le \sum_{j=1}^{p} [(g_j)_i]^2 \,=\, 1 \qquad \forall\, i = 1, \ldots, p$$

$$\sum_{i=1}^{p} w_i \,=\, \sum_{j=1}^{k}\sum_{i=1}^{p} [(g_j)_i]^2 \,=\, \sum_{j=1}^{k} 1 \,=\, k\,.$$

Il massimo di $\mathrm{tr}(H^T X^T XH) \,=\, \mathrm{tr}(G^T DG)$ sull'insieme delle matrici $p \times k$ G a colonne ortonormali ammette la stima

$$\max_{\substack{G\ p\times k \text{ a colonne} \\ \text{ortonormali}}} \mathrm{tr}(G^T DG) \le \max_{\substack{w_i \in [0,1] \\ \sum_{i=1}^{p} w_i = k}} \sum_{i=1}^{p} \lambda_i w_i\,.$$

Ma la determinazione di

$$\max_{\substack{w_i \in [0,1] \\ \sum_{i=1}^{p} w_i = k}} \sum_{i=1}^{p} \lambda_i w_i$$

costituisce un problema di ottimizzazione di una funzione lineare su un simplesso di $\mathbb{R}^p$

$$\Big\{(w_1,\ldots,w_p)\in\mathbb{R}^p\,:\,w_i\in[0,1]\,,\;i=1,\ldots,p\,,\;\sum_{i=1}^p w_i\,=\,k\Big\}\,,$$

per cui il relativo massimo si ottiene fissando $w_i=1\;\forall\,i=1,\ldots,k$ e $w_i=0\;\forall\,i=k{+}1\ldots,p$. Il valore del massimo è dato pertanto da $\sum_{i=1}^k\lambda_i$ e quindi

$$\max_{\substack{G\;p\times k\text{ a colonne}\\ \text{ortonormali}}}\operatorname{tr}(G^TDG)\le\sum_{i=1}^k\lambda_i\,.$$

Inversamente, una possibile scelta della matrice G si ottiene considerando i primi k autovettori ortonormali di D

$$g_1\,=\,\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix}\,=\,e_1\qquad g_2\,=\,\begin{pmatrix}0\\1\\\vdots\\0\end{pmatrix}\,=\,e_2\qquad\ldots\qquad g_k\,=\,\begin{pmatrix}0\\0\\\vdots\\0\end{pmatrix}\,=\,e_k$$

per cui posto $G=\big(g_1\big|g_2\big|\ldots\big|g_k\big)$ e ricordato che $Dg_j=\lambda_jg_j$, $\forall\,j=1,\ldots,k$, si ha

$$\operatorname{tr}(G^TDG)\,=\,\sum_{j=1}^k g_j^TDg_j\,=\,\sum_{j=1}^k g_j^T\lambda_jg_j\,=\,\sum_{j=1}^k\lambda_jg_j^Tg_j\,=\,\sum_{j=1}^k\lambda_j$$

con la conseguente conclusione che $\sum_{j=1}^k\lambda_j$ rappresenta effettivamente il valore del massimo

$$\sum_{j=1}^k\lambda_j\le\max_{\substack{G\;p\times k\text{ a colonne}\\ \text{ortonormali}}}\operatorname{tr}(G^TDG)\le\max_{\substack{w_i\,\in\,[0,1]\\ \sum_{i=1}^p w_i=k}}\sum_{i=1}^p\lambda_iw_i\,=\,\sum_{j=1}^k\lambda_j\,.$$

Vale inoltre

$$UH\,=\,G\,=\,\big(g_1\big|g_2\big|\ldots\big|g_k\big)\,=\,\big(e_1\big|e_2\big|\ldots\big|e_k\big)$$

in modo che

$$H\,=\,U^T\big(e_1\big|e_2\big|\ldots\big|e_k\big)\,=\,\big(U^Te_1\big|U^Te_2\big|\ldots\big|U^Te_k\big)$$

con

$$X^TX(U^Te_j)\,=\,U^TDU(U^Te_j)\,=\,U^TD(UU^T)e_j\,=\,U^TDe_j\,=\,U^T\lambda_je_j\,=\,\lambda_jU^Te_j\,.$$

Si conclude che per ogni $j=1,\ldots,k$ il vettore U^Te_j è autovettore di X^TX relativo all'autovalore λ_j e che di conseguenza le colonne della matrice H sono costituite da un set di autovettori ortonormali associati ai k più grandi autovalori di $X^T X$, come affermato. Ricordando infine la proprietà di circolarità della traccia, in virtù della quale $\mathrm{tr}(XX^T) = \mathrm{tr}(X^T X)$, la minima somma dei quadrati degli scarti si scrive

$$SS = \mathrm{tr}[X(\mathbb{I} - HH^T)X^T] = \mathrm{tr}(XX^T) - \mathrm{tr}(XHH^T X^T) = \mathrm{tr}(X^T X) - \sum_{j=1}^{k} \lambda_j$$

e siccome la traccia della matrice simmetrica $X^T X$ è la somma dei suoi autovalori (reali)

$$\mathrm{tr}(X^T X) = \sum_{j=1}^{p} \lambda_j$$

ne deriva che

$$SS = \sum_{j=1}^{p} \lambda_j - \sum_{j=1}^{k} \lambda_j = \sum_{j=k+1}^{p} \lambda_j \qquad \square$$

Modello a k componenti principali

Se si ritiene che i dati x_i siano prevalentemente ubicati lungo $\mathbb{S}_k$, e che gli eventuali scarti si possano interpretare come dovuti a fluttuazioni puramente casuali, si sostituiscono i punti x_i con le loro proiezioni ortogonali su $\mathbb{S}_k$

$$x_i \longrightarrow HH^T x_i$$

in modo che in luogo della matrice X dei dati viene considerato il **modello a k componenti principali**

$$XHH^T = \sum_{j=1}^{k} (Xh_j)h_j^T \tag{45.3}$$

nel quale si definisce:

- h_i come i-esima componente principale (**loading**);
- Xh_i come i-esimo **score**.

In forza del teorema precedente, la somma dei moduli quadrati dei dati (interpretabile formalmente come "devianza del campione")

$$\sum_{i=1}^{n} \sum_{j=1}^{p} x_{ij}^2 = \mathrm{tr}(X^T X) = \sum_{j=1}^{p} \lambda_j$$

si compone di una "devianza spiegata dal modello"

$$\mathrm{tr}[XHH^T X^T] = \mathrm{tr}[H^T X^T XH] = \sum_{j=1}^{k} \lambda_j$$

e della somma (minima) dei quadrati degli scarti o "devianza non spiegata" dal modello

$$SS \,=\, \mathrm{tr}[X(\mathbb{I} - HH^T)X^T] \,=\, \sum_{j=k+1}^{p} \lambda_j\,,$$

attribuibile alle fluttuazioni casuali e pertanto non interpretabile alla luce del modello a componenti principali.

Confronto con la SVD

La Singular Value Decomposition della matrice X si scrive come

$$X \,=\, U \begin{pmatrix} \Sigma & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{pmatrix} V^T \,=\, \sum_{j=1}^{r} \sigma_j u_j v_j^T$$

espressione nella quale:

- r indica il rango della matrice X — ovvero di $X^T X$ —

  $$r \,=\, \mathrm{rank}(X) \,=\, \mathrm{rank}(X^T X) \leq \min(n,p)\,;$$

- V è la matrice ortogonale $p \times p$

  $$V \,=\, \Big(v_1 \Big| v_2 \Big| \ldots \Big| v_p\Big)$$

  le cui colonne $v_1, v_2, \ldots, v_p$ costituiscono una base ortonormale di autovettori di $X^T X$

  $$X^T X v_i \,=\, \sigma_i^2 v_i \qquad i = 1, \ldots, p$$

  secondo gli autovalori di $X^T X$ ordinati in senso decrescente

  $$\sigma_1^2 \geq \sigma_2^2 \geq \ldots \geq \sigma_r^2 \geq \sigma_{r+1}^2 \,=\, \ldots \,=\, \sigma_p^2 \,=\, 0\,;$$

- Σ è la matrice diagonale $r \times r$ costituita dagli r valori singolari positivi di X

  $$\Sigma \,=\, \mathrm{diag}(\sigma_1, \ldots, \sigma_r) \,=\, \begin{pmatrix} \sigma_1 & & \mathbb{O} \\ & \ddots & \\ \mathbb{O} & & \sigma_r \end{pmatrix}$$

  radici quadrate aritmetiche degli r autovalori positivi di $X^T X$;

- U è la matrice $n \times n$

  $$U \,=\, \Big(u_1 \Big| u_2 \Big| \ldots \Big| u_n\Big)$$

  le cui colonne $u_1, u_2, \ldots, u_n$ formano una base ortonormale di $\mathbb{R}^n$ tale che

  $$u_i \,=\, \frac{1}{\sigma_i} X v_i \qquad \forall\, i = 1, \ldots, r\,.$$

Se $k \leq r$ — unico caso interessante —, si ha pertanto

$$\sum_{j=1}^{k} \sigma_j u_j v_j^T \;=\; \sum_{j=1}^{k} \sigma_j \frac{1}{\sigma_j}(Xv_j)v_j^T \;=\; \sum_{j=1}^{k}(Xv_j)v_j^T$$

e per ottenere il modello a k componenti principali basta porre $h_j = v_j \;\forall\, j = 1,\ldots,k$. Vale inoltre $\lambda_j = \sigma_j^2 \;\forall\, j = 1,\ldots,k$. Per confronto, si conclude che il modello a k componenti principali non è altro che il troncamento ai k maggiori valori singolari della SVD di X. Ne segue che gli algoritmi numerici per la PCA sulla matrice X dei dati sono gli stessi utilizzati per il calcolo della SVD.

Generalizzazione. k-varietà lineari di $\mathbb{R}^p$
Un modello più generale si ottiene approssimando la matrice X dei dati per mezzo di una varietà lineare k dimensionale di $\mathbb{R}^p$, $\Im_k$, che diversamente da un sottospazio vettoriale non necessariamente contiene lo zero e quindi non passa in generale per l'origine. Questo tipo di modello, in realtà, si riconduce facilmente a quello precedente. Basta osservare che i punti di una k-varietà lineare si ottengono sommando un vettore assegnato a ai vettori di un sottospazio vettoriale di dimensione k

$$\Im_k \;=\; a + S_k \;=\; \{a + x \;;\; x \in S_k\}\,.$$

Se S_k è individuato da una matrice H di k colonne ortonormali in $\mathbb{R}^p$, la proiezione ortogonale di un vettore $x \in \mathbb{R}^p$ su $\Im_k$ vale

$$(x-a)^T HH^T + a^T$$

e il vettore distanza di x da $\Im_k$ si riduce a

$$\delta^T \;=\; x^T - [(x-a)^T HH^T + a^T] \;=\; (x-a)^T - (x-a)^T HH^T \;=\; (x-a)^T(\mathbb{I} - HH^T)$$

con modulo quadrato

$$\delta^T\delta \;=\; (x-a)^T(\mathbb{I} - HH^T)(\mathbb{I} - HH^T)(x-a) \;=\; (x-a)^T(\mathbb{I} - HH^T)(x-a)\,.$$

Quest'ultima espressione può porsi in una forma più comoda per la successiva elaborazione, eseguendo i prodotti matriciali

$$\begin{aligned}
\delta^T\delta \;&=\; x^T(\mathbb{I} - HH^T)x - a^T(\mathbb{I} - HH^T)x - x^T(\mathbb{I} - HH^T)a + a^T(\mathbb{I} - HH^T)a \\
&=\; x^T(\mathbb{I} - HH^T)x - 2a^T(\mathbb{I} - HH^T)x + a^T(\mathbb{I} - HH^T)a\,.
\end{aligned}$$

Per il set di n dati $x_i \in \mathbb{R}^p$ la somma dei moduli quadrati delle distanze — o scarti rispetto al modello $\Im_k$ — risulta

$$\begin{aligned}
\sum_{i=1}^{n} \delta_i^T\delta_i \;&=\; \sum_{i=1}^{n}(x_i - a)^T(\mathbb{I} - HH^T)(x_i - a) \;= \\
&=\; \sum_{i=1}^{n}\Big[x_i^T(\mathbb{I} - HH^T)x_i - 2a^T(\mathbb{I} - HH^T)x_i + a^T(\mathbb{I} - HH^T)a\Big] \;= \\
&=\; \mathrm{tr}[X(\mathbb{I} - HH^T)X] - 2a^T(\mathbb{I} - HH^T)\sum_{i=1}^{n} x_i + na^T(\mathbb{I} - HH^T)a
\end{aligned}$$

e indicata con $\overline{x}$ la media aritmetica degli n vettori x_i

$$\sum_{i=1}^{n} \delta_i^T \delta_i \;=\; \mathrm{tr}[X(\mathbb{I} - HH^T)X] - 2na^T(\mathbb{I} - HH^T)\overline{x} + na^T(\mathbb{I} - HH^T)a\,.$$

Fra tutte le k-varietà lineari associate allo stesso sottospazio vettoriale $\mathbb{S}_k$ e quindi descritte dalla stessa matrice H, quella per la quale la somma dei moduli quadrati degli scarti risulta minima si ottiene fissando il vettore a in modo che si abbia

$$\frac{\partial}{\partial a_j}\sum_{i=1}^{n} \delta_i^T \delta_i \;=\; -2n[(\mathbb{I} - HH^T)\overline{x}]_j + 2n\sum_{j'=1}^{p}(\mathbb{I} - HH^T)_{jj'}a_{j'} \;=\; 0 \qquad \forall\, j = 1,\ldots,p$$

ossia

$$(\mathbb{I} - HH^T)a \;=\; (\mathbb{I} - HH^T)\overline{x}\,,$$

imponendo cioè che la distanza di a dal sottospazio $\mathbb{S}_k$ sia la stessa della media $\overline{x}$. Al valore ottimale della somma dei moduli quadrati degli scarti si perviene quindi ponendo $a = \overline{x}$

$$(x - \overline{x})^T HH^T + \overline{x}^T\,.$$

Introdotta la matrice $n \times n$ N di elementi costanti

$$N_{ij} \;=\; \frac{1}{n} \qquad \forall\, i,j = 1,\ldots,n\,,$$

il modello a k componenti principali diventa perciò

$$(X - NX)HH^T + NX$$

ottenibile con la stessa procedura già esaminata nel caso di $a = 0$ a patto però di sottrarre a tutti i dati x_i il loro valore medio

$$x_i - \overline{x} \;=\; x_i - \frac{1}{n}\sum_{q=1}^{n} x_q$$

ovvero di eseguire l'analisi sulla matrice dei dati "centrata" sul relativo valore medio

$$X - NX \;=\; (\mathbb{I} - N)X\,.$$

La somma dei moduli quadrati degli scarti assume perciò il valore minimo

$$\begin{aligned}\sum_{i=1}^{n} \delta_i^T \delta_i \;&=\; \mathrm{tr}[(X - NX)(\mathbb{I} - HH^T)(X - NX)^T] \;=\\ &=\; \mathrm{tr}[(X - NX)(X - NX)^T] - \mathrm{tr}[(X - NX)HH^T(X - NX)^T]\end{aligned}$$

per cui la **devianza dei dati**

$$\mathrm{tr}[(X - NX)(X - NX)^T] \;=\; \mathrm{tr}[(X - NX)^T(X - NX)]$$

può scomporsi in un **devianza "spiegata"** dal modello a componenti principali

$$\mathrm{tr}[(X - NX)(\mathbb{I} - HH^T)(X - NX)^T]$$

e in un termine di **devianza non spiegata**

$$\mathrm{tr}[(X - NX)HH^T(X - NX)^T] \;=\; \mathrm{tr}[H^T(X - NX)^T(X - NX)H]\,.$$

46. Modelli di regressione lineare con un parametro addittivo

@@@

47. Risultato 2

$$\mathrm{SSAR} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}\Big[y_i - \overline{y} - \frac{x_i - \overline{x}}{S_{xx}}\sum_{j=1}^{n}(x_j - \overline{x})(y_j - \overline{y})\Big]^2 =$$

$$= \sum_{i=1}^{n}\Big[\sum_{j=1}^{n}\Big(\delta_{ij} - \frac{(x_i - \overline{x})(x_j - \overline{x})}{S_{xx}}\Big)(y_j - \overline{y})\Big]^2 = \sum_{i=1}^{n}\Big[\sum_{j=1}^{n}A_{ij}(y_j - \overline{y})\Big]^2$$

con

$$A_{ij} = A_{ji} = \delta_{ij} - \frac{(x_i - \overline{x})(x_j - \overline{x})}{S_{xx}}$$

$$\mathrm{SSAR} = \sum_{i=1}^{n}\Big[\sum_{j=1}^{n}A_{ij}(y_j - \overline{y})\Big]\Big[\sum_{k=1}^{n}A_{ik}(y_k - \overline{y})\Big] = \sum_{i,j,k=1}^{n}A_{ij}A_{ik}(y_j - \overline{y})(y_k - \overline{y}) =$$

$$= \sum_{j,k=1}^{n}\Big(\sum_{i=1}^{n}A_{ij}A_{ik}\Big)(y_j - \overline{y})(y_k - \overline{y})$$

$$\sum_{i=1}^{n}A_{ij}A_{ik} = \sum_{i=1}^{n}\Big[\delta_{ij} - \frac{(x_i - \overline{x})(x_j - \overline{x})}{S_{xx}}\Big]\Big[\delta_{ik} - \frac{(x_i - \overline{x})(x_k - \overline{x})}{S_{xx}}\Big] = \delta_{jk} -$$

$$- \frac{(x_j - \overline{x})(x_k - \overline{x})}{S_{xx}} - \frac{(x_k - \overline{x})(x_j - \overline{x})}{S_{xx}} + \frac{(x_j - \overline{x})(x_k - \overline{x})}{S_{xx}} = \delta_{jk} - \frac{(x_j - \overline{x})(x_k - \overline{x})}{S_{xx}} = A_{jk}$$

per $1 \leq j,k \leq n$; la matrice A è cioè idempotente

$$A^2 = A$$

ed i suoi soli autovalori possibili sono 0 ed 1.

La SSAR assume allora la forma più semplice

$$\mathrm{SSAR} = \sum_{j,k=1}^{n}A_{jk}(y_j - \overline{y})(y_k - \overline{y})$$

dalla quale, osservato che

$$y_j - \overline{y} = \sum_{j=1}^{p}\delta_{jj'}y_{j'} - \sum_{j=1}^{p}\frac{1}{n}y_{j'} = \sum_{j=1}^{p}\Big(\delta_{jj'} - \frac{1}{n}\Big)y_{j'}$$

$$y_k - \overline{y} = \sum_{k=1}^{p}\Big(\delta_{kk'} - \frac{1}{n}\Big)y_{k'} \ ,$$

si deduce l'espressione

$$\mathrm{SSAR} = \sum_{j,k,j',k'=1}^{n}A_{jk}\Big(\delta_{jj'} - \frac{1}{n}\Big)y_{j'}\Big(\delta_{kk'} - \frac{1}{n}\Big)y_{j'}y_{k'} =$$

$$= \sum_{j',k'=1}^{n} y_{j'} y_{k'} \sum_{j,k=1}^{n} \Big(\delta_{jj'} - \frac{1}{n}\Big) y_{j'} A_{jk} \Big(\delta_{kk'} - \frac{1}{n}\Big) y_{j'} y_{k'}$$

in cui si riconosce una forma quadratica semidefinita positiva delle variabili y_i. La matrice rappresentativa di tale forma quadratica risulta, chiaramente,

$$B = \Big[\mathbb{I} - \frac{1}{n} Q^{(n)}\Big] A \Big[\mathbb{I} - \frac{1}{n} Q^{(n)}\Big]$$

in modo che

$$\mathrm{SSAR} = \sum_{j,k=1}^{n} B_{jk} y_j y_k \ .$$

□ $\mathbb{I} - \dfrac{1}{n} Q^{(n)}$ è reale simmetrica semidefinita positiva e idempotente:

$$\Big[\mathbb{I} - \frac{1}{n} Q^{(n)}\Big] \Big[\mathbb{I} - \frac{1}{n} Q^{(n)}\Big] = \mathbb{I} - \frac{1}{n} Q^{(n)} - \frac{1}{n} Q^{(n)} + \frac{1}{n^2} {Q^{(n)}}^2 =$$

$$= \mathbb{I} - \frac{2}{n} Q^{(n)} + \frac{1}{n^2}\, n Q^{(n)} = \mathbb{I} - \frac{1}{n} Q^{(n)}$$

essendosi fatto uso dell'ovvia identità ${Q^{(n)}}^2 = nQ^{(n)}$. Gli autovalori di $\mathbb{I} - \dfrac{1}{n} Q^{(n)}$ possono quindi essere soltanto 0 ed 1.

□ Molteplicità dell'autovalore 0

$$\Big(\mathbb{I} - \frac{1}{n} Q^{(n)}\Big) t = 0 \ , \quad t \in \mathbb{R}^n \qquad \Longleftrightarrow \qquad \Big\langle t \Big| \Big(\mathbb{I} - \frac{1}{n} Q^{(n)}\Big) t \Big\rangle = 0$$

in quanto la matrice è reale simmetrica semidefinita positiva (questo fatto non è del tutto banale e merita una dimostrazione, pur succinta). Equivalentemente:

$$\sum_{i,j=1}^{n} t_i \Big(\delta_{ij} - \frac{1}{n}\Big) t_j = \sum_{i=1}^{n} t_i^2 - \frac{1}{n} \sum_{i,j=1}^{n} t_i t_j = \frac{1}{n} \Big[n \sum_{i=1}^{n} t_i^2 - \Big(\sum_{i=1}^{n} t_i\Big)\Big(\sum_{j=1}^{n} t_j\Big)\Big] =$$

$$= \frac{1}{n} \Big[\sum_{j=1}^{n} 1^2 \sum_{i=1}^{n} t_i^2 - \Big(\sum_{i=1}^{n} 1\ t_i\Big)^2\Big] = \frac{1}{n}\Big[\|(1\ldots 1)\|^2 \|t\|^2 - \big|\langle (1\ldots 1)|t\rangle\big|^2 \Big] \geq 0 \ .$$

Per Cauchy-Schwarz si ha che $\Big\langle t \Big| \Big(\mathbb{I} - \frac{1}{n} Q^{(n)}\Big) t \Big\rangle = 0$ se e soltanto se

$$t = \alpha\,(1\ldots 1) \ , \quad \alpha \in \mathbb{R} \ .$$

e pertanto

$$\Big(\mathbb{I} - \frac{1}{n} Q^{(n)}\Big) t = \qquad \Longleftrightarrow \qquad t \in \big\{\alpha\,(1\ldots 1) \,,\ \alpha \in \mathbb{R}\big\}$$

che costituisce uno spazio vettoriale di dimensione 1. Ciò dimostra che 0 è un autovalore ed implica inoltre che anche 1 lo sia a propria volta. D'altra parte

$$\ker\Big(\mathbb{I} - \frac{1}{n}Q^{(n)}\Big) \perp \ker\Big(\mathbb{I} - \frac{1}{n}Q^{(n)} - \mathbb{I}\Big)$$

— 0 ed 1 sono i soli autovalori — e

$$\mathbb{R}^n = \ker\Big(\mathbb{I} - \frac{1}{n}Q^{(n)}\Big) \oplus \ker\Big(\mathbb{I} - \frac{1}{n}Q^{(n)} - \mathbb{I}\Big)$$

— la matrice è reale e simmetrica. Perciò

$$\dim\ker\Big(\mathbb{I} - \frac{1}{n}Q^{(n)} - \mathbb{I}\Big) = n-1$$

ed inoltre

$$\ker\Big(\mathbb{I} - \frac{1}{n}Q^{(n)} - \mathbb{I}\Big) = \Big\{t \in \mathbb{R}^n \,:\, \langle t|(1\dots1)\rangle = 0\Big\} .$$

□ A è reale rimmetrica semidefinita positiva e idempotente, in quanto $A^2 = A$. Gli autovalori di A possono perciò essere soltanto 0 ed 1.

□ Molteplicità dell'autovalore 0.

$$At = 0 , \quad t \in \mathbb{R}^n \qquad \Longleftrightarrow \qquad \langle t|At\rangle = 0$$

poiché la matrice è reale simmetrica semidefinita positiva. Pertanto

$$\sum_{i,j=1}^n t_i A_{ij} t_j = \sum_{i,j=1}^n t_i t_j \Big[\delta_{ij} - \frac{1}{S_{xx}}(x_i-\overline{x})(x_j-\overline{x})\Big] = \sum_{i=1}^n t_i^2 - \frac{1}{S_{xx}}\sum_{i,j=1}^n (x_i-\overline{x})t_i(x_j-\overline{x})t_j =$$

$$= \frac{1}{S_{xx}}\Big[\sum_{i=1}^n t_i^2 \sum_{j=1}^n (x_j-\overline{x})^2 - \Big(\sum_{k=1}^n (x_k-\overline{x})t_k\Big)^2\Big] = \frac{1}{S_{xx}}\Big[\|t\|^2\|x-\overline{x}\|^2 - \big|\langle t|x-\overline{x}\rangle\big|^2\Big] \geq 0$$

Per Cauchy-Schwarz è $\langle t|At\rangle = 0$ se e soltanto se i vettori t ed $x-\overline{x}$ sono paralleli, vale a dire se e soltanto se

$$t = \beta(x-\overline{x}) , \quad \beta \in \mathbb{R} ,$$

e quindi

$$At = 0 \qquad \Longleftrightarrow \qquad t \in \big\{\beta(x-\overline{x}) ,\ \beta \in \mathbb{R}\big\}$$

(si ricordi che per ipotesi $x-\overline{x} \neq 0$ e che le componenti $x_i - \overline{x}$ non possono essere tutte uguali).
Poiché inoltre:

$$\ker(A) \perp \ker(A-\mathbb{I}) \qquad \text{(0 ed 1 sono autovalori distinti di } A \text{ reale simmetrica)}$$

e

$$\ker(A) \oplus \ker(A - \mathbb{I}) \;=\; \mathbb{R}^n \qquad \text{(0 ed 1 sono i soli autovalori di } A \text{ diagonalizzabile)}$$

si conclude che $\dim\ker(A - \mathbb{I}) = n - 1$ e che inoltre

$$\ker(A - \mathbb{I}) \;=\; \big\{t \in \mathbb{R}^n \,:\, \langle t|x - \overline{x}\rangle = 0\big\}$$

□ Si osservi che $x - \overline{x}$ è autovettore di $\mathbb{I} - \dfrac{1}{n}Q^{(n)}$:

$$\Big(\mathbb{I} - \frac{1}{n}Q^{(n)}\Big)(x - \overline{x}) \;=\; x - \overline{x} - \frac{1}{n}Q^{(n)}(x - \overline{x}) \;=\; x - \overline{x}$$

per cui

$$x - \overline{x} \in \ker\Big(\mathbb{I} - \frac{1}{n}Q^{(n)} - \mathbb{I}\Big)$$

e di conseguenza

$$\ker(A) \subset \ker\Big(\mathbb{I} - \frac{1}{n}Q^{(n)} - \mathbb{I}\Big)$$

Si può allora scrivere

$$\begin{aligned}
\ker\Big(\mathbb{I} - \frac{1}{n}Q^{(n)} - \mathbb{I}\Big) \;&=\; \ker(A) \oplus \Big[\ker(A)^{\perp} \cap \ker\Big(\mathbb{I} - \frac{1}{n}Q^{(n)} - \mathbb{I}\Big)\Big] \;= \\
&=\; \ker(A) \oplus \Big[\ker(A - \mathbb{I}) \cap \ker\Big(\mathbb{I} - \frac{1}{n}Q^{(n)} - \mathbb{I}\Big)\Big] \;= \\
&=\; \ker(A) \oplus \Big[\big\{\beta(x - \overline{x})\,,\; \beta \in \mathbb{R}\big\}^{\perp} \cap \big\{t \,:\, \langle t|(1\ldots 1)\rangle = 0\big\}\Big] \;= \\
&=\; \ker(A) \oplus \Big[\big\{t \,:\, \langle t|x - \overline{x}\rangle = 0\big\} \cap \big\{t \,:\, \langle t|(1\ldots 1)\rangle = 0\big\}\Big] \;= \\
&=\; \ker(A) \oplus \big\{t \,:\, \langle t|x - \overline{x}\rangle = 0\,,\; \langle t|(1\ldots 1)\rangle = 0\big\}
\end{aligned}$$

In definitiva

$$\begin{aligned}
\mathbb{R}^n \;&=\; \ker(A) \oplus \Big[\ker(A - \mathbb{I}) \cap \ker\Big(\mathbb{I} - \frac{1}{n}Q^{(n)} - \mathbb{I}\Big)\Big] \oplus \ker\Big(\mathbb{I} - \frac{1}{n}Q^{(n)}\Big) \;= \\
&=\; \big\{\beta(x - \overline{x})\,,\; \beta \in \mathbb{R}\big\} \oplus \big\{t \,:\, \langle t|x - \overline{x}\rangle = \langle t|(1\ldots 1)\rangle = 0\big\} \oplus \big\{\alpha(1\ldots 1)\,,\; \alpha \in \mathbb{R}\big\}
\end{aligned}$$

dove $\big\{t \,:\, \langle t|x - \overline{x}\rangle = \langle t|(1\ldots 1)\rangle = 0\big\}$ è uno spazio vettoriale di dimensione $n-2$ in quanto $x - \overline{x}$ e $(1\ldots 1)$ risultano ortogonali:

$$\langle x - \overline{x}|(1\ldots 1)\rangle \;=\; \sum_{i=1}^{n}(x_i - \overline{x}) \;=\; 0$$

□ Autovalori ed autovettori di $B = \Big(\mathbb{I} - \frac{1}{n}Q^{(n)}\Big)A\Big(\mathbb{I} - \frac{1}{n}Q^{(n)}\Big)$
Sia $\{e_1, e_2, \ldots, e_{n-2}\}$ una base di $\big\{t \,:\, \langle t|x-\overline{x}\rangle = \langle t|(1\ldots 1)\rangle = 0\big\}$. Allora $\{e_1, e_2, \ldots, e_{n-2}, x-\overline{x}, (1\ldots 1)\}$ costituisce una base di $\mathbb{R}^n$. Per ogni $j = 1, 2, \ldots, n-2$ si ha:

$$Be_j \,=\, \Big(\mathbb{I} - \frac{1}{n}Q^{(n)}\Big)A\Big(\mathbb{I} - \frac{1}{n}Q^{(n)}\Big)e_j \,=\, \Big(\mathbb{I} - \frac{1}{n}Q^{(n)}\Big)Ae_j \,=\, \Big(\mathbb{I} - \frac{1}{n}Q^{(n)}\Big)e_j \,=\, e_j$$

e quindi

$$e_j \in \ker(B - \mathbb{I}) \qquad 1 \le j \le n-2$$

Analogamente:

$$B(x-\overline{x}) = \Big(\mathbb{I}-\frac{1}{n}Q^{(n)}\Big)A\Big(\mathbb{I}-\frac{1}{n}Q^{(n)}\Big)(x-\overline{x}) = \Big(\mathbb{I}-\frac{1}{n}Q^{(n)}\Big)A(x-\overline{x}) = \Big(\mathbb{I}-\frac{1}{n}Q^{(n)}\Big)\mathbb{O} = \mathbb{O}$$

e:

$$B(1\ldots 1) \,=\, \Big(\mathbb{I} - \frac{1}{n}Q^{(n)}\Big)A\Big(\mathbb{I} - \frac{1}{n}Q^{(n)}\Big)(1\ldots 1) \,=\, \Big(\mathbb{I} - \frac{1}{n}Q^{(n)}\Big)A\mathbb{O} \,=\, \mathbb{O}$$

per cui

$$x-\overline{x}\,,\ (1\ldots 1) \,\in\, \ker(B)\ .$$

In conclusione, gli autovalori di B sono

1 , con molteplicità $n-2$;

0 , di molteplicità 2.

□ Gradi di libertà
Poiché B è reale e simmetrica, essa è riconducibile alla forma diagonale per mezzo di una trasformazione di similitudine

$$C^{-1}BC \,=\, D\ ,$$

con C ortogonale — $C^{-1} = C^T$ — e D data da

$$D \,=\, \begin{pmatrix} 1 & & & & & \\ & 1 & & & \mathbb{O} & \\ & & \ddots & & & \\ & & & 1 & & \\ & \mathbb{O} & & & 0 & \\ & & & & & 0 \end{pmatrix}$$

in forza delle proprietà spettrali precedentemente stabilite.

Nell'espressione di SSAR si introduce il cambiamento di variabili

$$y \,=\, Cz$$

sicché

$$\mathrm{SSAR} \,=\, \sum_{i,j=1}^{n} y_i y_j B_{ij} \,=\, \sum_{i,j,i',j'=1}^{n} C_{ii'}z_{i'}B_{ij}C_{jj'}z_{j'} \,=\, \sum_{i,j,i',j'=1}^{n} z_{i'}(C^T)_{i'i}B_{ij}C_{jj'}z_{j'} \,=$$

$$= \sum_{i',j'=1}^{n} z_{i'} \sum_{i,j=1}^{n} (C^{-1})_{i'i} B_{ij} C_{jj'} z_{j'} = \sum_{i',j'=1}^{n} z_{i'} D_{i'j'} z_{j'} = \sum_{i=1}^{n-2} z_i^2$$

Le nuove variabili casuali

$$z_i = \sum_{i'=1}^{n} (C^T)_{ii'} y_{i'} = \sum_{i'=1}^{n} C_{i'i} y_{i'}$$

sono gaussiane in quanto combinazioni lineari di variabili gaussiane, con media nulla

$$\mathbb{E}(z_i) = \sum_{i'=1}^{n} C_{i'i} \mathbb{E}(y_{i'}) = \sum_{i'=1}^{n} C_{i'i} 0 = 0$$

e correlazioni

$$\mathbb{E}(z_i z_j) = \mathbb{E}\Big[\sum_{i',j'=1}^{n} C_{i'i} y_{i'} C_{j'j} y_{j'} \Big] = \sum_{i',j'=1}^{n} C_{i'i} C_{j'j} \mathbb{E}(y_{i'} y_{j'}) = \sigma^2 \sum_{i',j'=1}^{n} C_{i'i} C_{j'j} \delta_{i'j'} =$$

$$= \sigma^2 \sum_{i'=1}^{n} C_{i'i} C_{i'j} = \sigma^2 \sum_{i'=1}^{n} (C^T)_{ii'} C_{i'j} = \sigma^2 (C^T C)_{ij} = \sigma^2 \delta_{ij} \qquad 1 \le i, j \le n$$

dove si è fatto uso della relazione

$$\mathbb{E}(y_{i'} y_{j'}) = \sigma^2 \delta_{i'j'} \qquad 1 \le i', j' \le n$$

valida per via dell'ipotesi che vede le $y_1, y_2, \ldots, y_n$ variabili scorrelate di uguale varianza σ^2 (usualmente si assume che le $y_1, y_2, \ldots, y_n$ siano indipendenti e di uguale varianza σ^2, condizione che per variabili gaussiane è equivalente alla precedente).

Si osserva quindi che le $z_1, z_2, \ldots, z_n$ sono variabili gaussiane scorrelate e perciò indipendenti; tutte hanno la stessa varianza σ^2. Di conseguenza, la variabile

$$\frac{\text{SSAR}}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^{n-2} z_i^2 = \sum_{i=1}^{n-2} (z_i/\sigma)^2$$

è la somma dei quadrati di $n-2$ variabili gaussiane indipendenti di media nulla e varianza unitaria. *Si tratta, per definizione, di una variabile di $\mathcal{X}^2$ a $n-2$ gradi di libertà.*

□ Un'utile osservazione.

$$\mathbb{E}(\text{SSAR}) = \mathbb{E}\Big[\sum_{i=1}^{n-2} z_i^2 \Big] = \sum_{i=1}^{n-2} \mathbb{E}(z_i^2) = \sum_{i=1}^{n-2} \sigma^2 = (n-2)\sigma^2$$

per cui $\sigma^2 = \mathbb{E}(\text{SSAR}/(n-2))$.