Indice

Elenco delle figure

Capitolo 1

avvertenza

Gli appunti di questa sezione relativi alla normale bivariata e, in particolare, quelli sulla normale multivariata, sono molto abbondanti rispetto a quanto effettivamente riportato sulla scheda trasparenza e a quanto fatto in aula (*e a quanto sarà utile ai fini dell'esame*).

Tuttavia, almeno ad oggi , ho preferito lasciare tutto il materiale, anche quello non finalizzato al nostro corso. Chiariremo in aula, e in una successiva redazione di questi appunti, le parti essenziali.

Eliminare la Normale Multivariata? NOOO!!!!

Aggiungo con sincerità che ogni anno mi chiedo se è il caso di eliminare del tutto questa parte sulla normale bivariata e multivariata da questo corso e ogni anno, fino ad ora, mi sono sempre risposto di no!

Al termine ho anche rimesso una piccola parte relativa a distribuzioni multivariate non normali (piccolissimi cenni).

Capitolo 2

La distribuzione normale multivariata

La distribuzione normale multivariata può essere introdotta in numerosi modi, ed espressa con diverse caratterizzazioni.

Qui viene presentata come la distribuzione congiunta di *combinazioni lineari di variabili normali indipendenti e standardizzate.*

Uso adesso una **notazione matriciale compatta** perché non sarebbe conveniente lavorare con una distribuzione normale trivariata, poi con una normale a quattro variabili, etc. etc.; alcuni risultati della normale trivariata saranno poi ricavati come casi particolari per qualche esercizio perchè è ancora possibile qualche rappresentazione grafica (in termini di ellissoidi 3d di equidensità) che consente di cogliere bene concetti come la differenza fra dipendenza condizionata e marginale

2.1 Distribuzione di variabili normali indipendenti

Sia $\mathbf{X}$ un vettore di variabili casuali a p componenti indipendenti:

$$\mathbf{X} = \{X_1, X_2, \ldots, X_i, \ldots, X_p\}^{\mathsf{T}}$$

ciascuna distribuita secondo una normale standardizzata.

La densità congiunta di tali variabili casuali, data l'indipendenza, è data da:

Densità congiunta di p variabili normali standardizzate e indipendenti.

$$
\begin{aligned}
f_{\mathbf{X}}(\mathbf{x}) &= \prod_{i=1}^{p} f(x_i) = (2\pi)^{-p/2} \exp\left[-\sum_{i=1}^{p} \frac{x_i^2}{2}\right] = \\
&= \frac{1}{\sqrt{(2\pi)^p}} \exp[-\frac{\mathbf{x}^\mathsf{T}\mathbf{x}}{2}] \qquad (2.1)
\end{aligned}
$$

La funzione caratteristica è:

$$\phi_{\mathbf{X}}(\mathbf{t}) = \exp\left(-\frac{1}{2}\mathbf{t}^\mathsf{T}\mathbf{t}\right)$$

Ovviamente i primi due momenti di $\mathbf{X}$, per le ipotesi fatte, sono:

$$
\begin{aligned}
\mathrm{E}\left[\mathbf{X}\right] &= \mathbf{0}_p \qquad \text{vettore nullo di } p \text{ componenti} \\
\mathrm{V}\left[\mathbf{X}\right] &= \mathbf{I}_p \qquad \text{matrice identità } p \times p
\end{aligned}
$$

E' noto, ed è facile comunque vederlo attraverso la funzione caratteristica, che una singola combinazione lineare Z del vettore aleatorio $\mathbf{X}$ si distribuisce secondo una normale univariata, con speranza matematica e varianza ricavabili dalle relazioni per i momenti di combinazioni lineari di vettori aleatori qualsiasi.

Infatti se: $Z = \mathbf{b}^\mathsf{T}\mathbf{X} + c$, allora i primi due momenti di Z sono dati da:

$$
\begin{aligned}
\mathrm{E}\left[Z\right] &= \mathbf{b}^\mathsf{T}\mathrm{E}\left[\mathbf{X}\right] + c = c \\
\mathrm{V}\left[Z\right] &= \mathbf{b}^\mathsf{T}\mathrm{V}\left[\mathbf{X}\right]\mathbf{b} = \mathbf{b}^\mathsf{T}\mathbf{b} = b_1^2 + b_2^2 + \ldots + b_i^2 + \ldots + b_p^2
\end{aligned}
$$

e si ha anche:

$$Z \sim N(\mathrm{E}\left[Z\right], \mathrm{V}\left[Z\right]).$$

Adesso occorre però studiare la distribuzione congiunta di p combinazioni lineari di variabili normali indipendenti, che ci porterà alla distribuzione normale multivariata.

2.2 Densità della distribuzione congiunta di p combinazioni lineari di p variabili normali indipendenti

Consideriamo allora il vettore aleatorio $\mathbf{Y}$, trasformazione lineare del vettore aleatorio $\mathbf{X}$, definito dalla relazione:

$$\mathbf{Y} = \mathbf{A}^{\mathsf{T}}\mathbf{X} + \boldsymbol{\mu}$$

essendo:

$\mathbf{A}$ una matrice quadrata di dimensione p e rango pieno;

$\boldsymbol{\mu}$ un vettore di p elementi;

Per ora abbiamo posto la condizione che $\mathbf{A}$ sia a rango pieno p, sarà poi possibile generalizzare a trasformazioni $\mathbf{X} \Rightarrow \mathbf{Y}$ anche singolari, ossia a rango non pieno.

Rispetto alla notazione ordinaria ho indicata la trasformazione mediante una matrice trasposta, perché di solito si dà un significato geometrico alle colonne di $\mathbf{A}$, ed ogni componente di $\mathbf{Y}$ corrisponde ad una colonna di $\mathbf{A}$; inoltre è irrilevante ai fini del risultato partire da p variabili standardizzate $\mathbf{X}_i$ oppure a varianza qualsiasi: l'importante è che siano indipendenti.

Per trovare la densità di $\mathbf{Y}$ seguono una serie di passaggi che il lettore non interessato può anche saltare per andare direttamente al risultato finale che segue in (2.3).

Per le proprietà sui momenti di trasformate lineari di v.a., i momenti di $\mathbf{Y}$ sono dati da:

$$\begin{aligned} \mathrm{E}[\mathbf{Y}] &= \mathbf{A}^{\mathsf{T}}\mathrm{E}[\mathbf{X}] + \boldsymbol{\mu} = \boldsymbol{\mu} \\ \mathrm{V}[\mathbf{Y}] &= \mathbf{A}^{\mathsf{T}}\mathrm{V}[\mathbf{X}]\,\mathbf{A} = \mathbf{A}^{\mathsf{T}}\mathbf{A} \end{aligned}$$

ricordando che per ipotesi $\mathrm{E}[\mathbf{X}] = \mathbf{0}_p$ e $\mathrm{V}[\mathbf{X}] = \mathbf{I}_p$. Per ricavare la densità di $\mathbf{Y}$ è conveniente esplicitare la trasformazione inversa; dalla relazione diretta:

$$\mathbf{Y} = \mathbf{A}^{\mathsf{T}}\mathbf{X} + \boldsymbol{\mu},$$

si ottiene subito la relazione inversa:

$$\mathbf{X} = \mathbf{B}^{\mathsf{T}}(\mathbf{Y} - \boldsymbol{\mu}), \quad \text{avendo posto:} \quad \mathbf{B} = (\mathbf{A})^{-1}$$

Pertanto, applicando la regola per le densità di trasformazioni di variabili aleatorie, la densità di $\mathbf{Y}$ è data da:

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}) &= f_{\mathbf{X}}(\mathbf{B}^{\mathsf{T}}[\mathbf{y} - \boldsymbol{\mu}]) \left|\mathbf{J}_{(\mathbf{X},\mathbf{Y})}\right| \\ &= \left|\mathbf{J}_{(\mathbf{X},\mathbf{Y})}\right| (2\pi)^{-p/2} \exp\left(-\frac{1}{2}[\mathbf{y} - \boldsymbol{\mu}]^{\mathsf{T}}\mathbf{B}\mathbf{B}^{\mathsf{T}}[\mathbf{y} - \boldsymbol{\mu}]\right) \end{aligned}$$

essendo $\left|\mathbf{J}_{(\mathbf{X},\mathbf{Y})}\right|$ lo Jacobiano della trasformazione da $\mathbf{Y}$ a $\mathbf{X}$, ossia il determinante della matrice $\frac{\partial \mathbf{X}}{\partial \mathbf{Y}}$, che ovviamente è dato da $\mathbf{J}_{(\mathbf{X},\mathbf{Y})} = |\mathbf{B}|$, per cui si ha:

$$f_{\mathbf{y}}(\mathbf{y}) = |B|(2\pi)^{-p/2} \exp\left(-\frac{1}{2}[\mathbf{y} - \boldsymbol{\mu}]^{\mathsf{T}}\mathbf{B}\mathbf{B}^{\mathsf{T}}[\mathbf{y} - \boldsymbol{\mu}]\right) \tag{2.2}$$

```
statistica3_2020provetrinorm1.R
```

Questa è la densità richiesta, tuttavia è meglio parametrizzare questa distribuzione in modo che sia esplicito, se possibile, il legame con i momenti di $\mathbf{Y}$.

Indichiamo con $\boldsymbol{\Sigma}$ la matrice di varianza e covarianza di $\mathbf{Y}$, ossia $\mathrm{V}[\mathbf{Y}]$, che abbiamo già visto essere uguale a $\mathbf{A}^\mathsf{T}\mathbf{A}$. Se vogliamo esprimere $\mathrm{V}[\mathbf{X}]$ in funzione di $\mathrm{V}[\mathbf{Y}]$ si ha:

$$\mathrm{V}[\mathbf{X}] = \mathbf{B}^\mathsf{T}\mathrm{V}[\mathbf{Y}]\,\mathbf{B} = \mathbf{B}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{B}.$$

Per ipotesi abbiamo però che $\mathrm{V}[\mathbf{X}] = \mathbf{I}_p$, per cui:

$$\mathbf{B}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{B} = \mathbf{I}_p$$

e quindi la matrice $\mathbf{B}$ diagonalizza $\boldsymbol{\Sigma}$, per cui ha colonne proporzionali agli autovettori di $\boldsymbol{\Sigma}$ divisi per le radici dei rispettivi autovalori (si riveda eventualmente la diagonalizzazione di matrici semidefinite positive $\boldsymbol{\Gamma}^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{\Gamma} = \boldsymbol{\Lambda}$, basate su autovalori ed autovettori).

verificare se la matrice è unica

Inoltre, prendendo in esame la relazione $\mathbf{B}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{B} = \mathbf{I}$, premoltiplicando ambo i membri per $\mathbf{B}$ e postmoltiplicando per $\mathbf{B}^\mathsf{T}$, si ottiene:

$$\mathbf{B}\mathbf{B}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{B}\mathbf{B}^\mathsf{T} = \mathbf{B}\mathbf{B}^\mathsf{T} \qquad \text{e quindi,}$$

postmoltiplicando o premoltiplicando ora ambo i membri per $(\mathbf{B}\mathbf{B}^\mathsf{T})^{-1}$ (che esiste sempre essendo $\mathbf{B}$, e quindi anche $\mathbf{B}\mathbf{B}^\mathsf{T}$, a rango pieno p), si ha:

$$\mathbf{B}\mathbf{B}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{B}\mathbf{B}^\mathsf{T}(\mathbf{B}\mathbf{B}^\mathsf{T})^{-1} = \mathbf{B}\mathbf{B}^\mathsf{T}(\mathbf{B}\mathbf{B}^\mathsf{T})^{-1} \qquad \text{e quindi:} \mathbf{B}\mathbf{B}^\mathsf{T}\boldsymbol{\Sigma} = \mathbf{I}.$$

e in definitiva:

$$\mathbf{B}\mathbf{B}^\mathsf{T} = \boldsymbol{\Sigma}^{-1}$$

e quindi nella forma quadratica ad esponente nell'espressione (2.2) di $f_{\mathbf{Y}}(\mathbf{y})$ potremo sostituire $\mathbf{B}\mathbf{B}^\mathsf{T}$ con $\boldsymbol{\Sigma}^{-1}$.

Più semplicemente in effetti, senza riferirsi all'ortogonalizzazione, si poteva ottenere:

$$\mathbf{B} = (\mathbf{A})^{-1}; \qquad \mathbf{B}^\mathsf{T} = \left(\mathbf{A}^{-1}\right)^\mathsf{T} \qquad \mathbf{A} = (\mathbf{B})^{-1} \qquad \mathbf{A}^\mathsf{T}\mathbf{A} = \left(\mathbf{B}^\mathsf{T}\mathbf{B}\right)^{-1}$$

e siccome $\mathbf{A}^\mathsf{T}\mathbf{A} = \boldsymbol{\Sigma}$, in definitiva si ha:

$$\left(\mathbf{B}^\mathsf{T}\mathbf{B}\right)^{-1} = \boldsymbol{\Sigma} \qquad \text{e infine:} \qquad \mathbf{B}^\mathsf{T}\mathbf{B} = \boldsymbol{\Sigma}^{-1}$$

Per potere ottenere il determinante di $\mathbf{B}$ che compare in $f_{\mathbf{Y}}(\mathbf{y})$, basta applicare le note regole sui determinanti delle trasposte, dei prodotti e delle inverse, per vedere che:

$$|\mathbf{B}| = |\mathbf{B}^\mathsf{T}| = \left(|\mathbf{B}||\mathbf{B}^\mathsf{T}|\right)^{\frac{1}{2}} = |\mathbf{B}\mathbf{B}^\mathsf{T}|^{\frac{1}{2}} = |\boldsymbol{\Sigma}^{-1}|^{\frac{1}{2}} = |\boldsymbol{\Sigma}|^{-\frac{1}{2}}$$

Inoltre essendo $\boldsymbol{\Sigma}$ definita positiva, il suo determinante è certamente positivo.

2.2.1 Densità della distribuzione normale multivariata

In conclusione, sostituendo nella densità di $\mathbf{y}$:

$$f_{\mathbf{Y}}(\mathbf{y}) = |\mathbf{B}|(2\pi)^{-p/2}\exp[-\frac{1}{2}[\mathbf{y}-\boldsymbol{\mu}]^\mathsf{T}\mathbf{B}\mathbf{B}^\mathsf{T}[\mathbf{y}-\boldsymbol{\mu}]]$$

abbiamo:

Densità della distribuzione normale multivariata *non singolare* di parametri $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$:

$$f_{\mathbf{Y}}(\mathbf{y}) = |\boldsymbol{\Sigma}|^{-\frac{1}{2}}(2\pi)^{-p/2}\exp\left\{-\frac{1}{2}[\mathbf{y}-\boldsymbol{\mu}]^\mathsf{T}\boldsymbol{\Sigma}^{-1}[\mathbf{y}-\boldsymbol{\mu}]\right\} \tag{2.3}$$

o anche:

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{\sqrt{|\boldsymbol{\Sigma}|\,(2\pi)^p}} e^{-\frac{1}{2}[\mathbf{y}-\boldsymbol{\mu}]^\mathsf{T}\boldsymbol{\Sigma}^{-1}[\mathbf{y}-\boldsymbol{\mu}]}$$

Dalle espressioni della densità riportate sopra, è evidente l'analogia con l'espres– sione della densità della distribuzione normale univariata.

I primi due momenti multivariati sono (come già visto prima e senza alcun bisogno di effettuare integrazioni p-dimensionali):

$$\mathrm{E}\left[\mathbf{Y}\right] = \boldsymbol{\mu}$$

$$\mathrm{V}\left[\mathbf{Y}\right] = \boldsymbol{\Sigma}$$

Si vede quindi, come nel caso univariato, che la distribuzione normale multivariata dipende soltanto dai primi due momenti (multivariati) di $\mathbf{Y}$.

La funzione caratteristica (applicando la regola per le trasformazioni lineari di variabili aleatorie) è data da:

$$\phi_{\mathbf{Y}}(\mathbf{t}) = \exp[i\mathbf{t}^\mathsf{T}\boldsymbol{\mu} - \frac{1}{2}\mathbf{t}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{t}] \tag{2.4}$$

Ricordo che i momenti possono essere eventualmente ricavati dalle opportune derivate di $\phi_{\mathbf{Y}}(\mathbf{t})$, valutate in $\mathbf{t} = \mathbf{0}$.

Sottolineo ancora che se $\mathbf{A}$ non ha rango p, la distribuzione di $\mathbf{Y}$ è sempre una normale multivariata, ma non è dotata di densità. Si intuisce facilmente questa proprietà, senza darne una dimostrazione, perchè in effetti alla funzione caratteristica (2.4) si arriva anche senza alcuna restrizione sul rango di $\mathbf{A}$.

Inoltre è possibile far vedere, rifacendo a ritroso i passaggi precedenti, che, se $\mathbf{C}$ è una matrice simmetrica definita positiva di rango p, qualsiasi vettore aleatorio $\mathbf{Y}$ la cui densità è data da:

$$f_{\mathbf{Y}}(\mathbf{y}) = |\mathbf{C}|^{\frac{1}{2}}(2\pi)^{-p/2}\exp\left(-\frac{1}{2}[\mathbf{y}-\boldsymbol{\mu}]^{\mathsf{T}}\mathbf{C}[\mathbf{y}-\boldsymbol{\mu}]\right) \tag{2.5}$$

è distribuito secondo una normale multivariata di parametri $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma} = \mathbf{C}^{-1}$. Esiste inoltre una trasformazione lineare di $\mathbf{Y}$ che conduce ad un vettore aleatorio $\mathbf{X}$ a componenti standardizzate e indipendenti:

$$\mathbf{X} = \mathbf{B}^{\mathsf{T}}[\mathbf{Y}-\boldsymbol{\mu}], \quad \text{in cui } \mathbf{B} \text{ è tale che:} \quad \mathbf{B}^{\mathsf{T}}\boldsymbol{\Sigma}\mathbf{B} = \mathbf{I}$$

2.2.2 Distribuzioni marginali

La distribuzione marginale di un qualsiasi sottoinsieme di componenti di un vettore aleatorio distribuito secondo una normale multivariata è ancora distribuito secondo una normale multivariata con parametri uguali ai corrispondenti sottoinsiemi di $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$: il risultato si dimostra facilmente, sfruttando la funzione caratteristica.[1].

Infatti se il vettore $\mathbf{Y}$ è suddiviso in due sottovettori $[\mathbf{Y}_A, \mathbf{Y}_B]$, corrispondentemente suddividiamo il vettore delle medie e la matrice di varianza e covarianza:

$$\begin{aligned}\boldsymbol{\mu} &= \left(\begin{array}{c}\boldsymbol{\mu}_A \\ \hline \boldsymbol{\mu}_B\end{array}\right)\\ \boldsymbol{\Sigma} &= \left(\begin{array}{c|c}\boldsymbol{\Sigma}_{AA} & \boldsymbol{\Sigma}_{AB} \\ \hline \boldsymbol{\Sigma}_{AB}^{\mathsf{T}} & \boldsymbol{\Sigma}_{BB}\end{array}\right)\end{aligned}$$

Posta ora, corrispondentemente alla partizione di $\mathbf{Y}$, una partizione $\mathbf{t} = \{\mathbf{t}_A, \mathbf{t}_B\}$, come si sa la funzione caratteristica della distribuzione marginale di $\mathbf{Y}_A$ si ottiene da quella di $\mathbf{Y}$ ponendo $\mathbf{t}_2 = \mathbf{0}$:

$$\phi_{\mathbf{Y}_A}(\mathbf{t}_A) = \phi_{\mathbf{Y}}(\mathbf{t}_A, \mathbf{0}) = \exp[i\mathbf{t}_A^{\mathsf{T}}\boldsymbol{\mu}_A - \frac{1}{2}\mathbf{t}_A^{\mathsf{T}}\boldsymbol{\Sigma}_{AA}\mathbf{t}_A]$$

che è la funzione caratteristica di una normale di parametri μ_A e $\boldsymbol{\Sigma}_{AA}$.

In particolare tutte le distribuzioni marginali delle singole componenti sono normali univariate.

Come corollario è facile vedere che $\mathbf{Y}_A$ e $\mathbf{Y}_B$ (vettori aleatori normali) sono indipendenti se e solo se $\boldsymbol{\Sigma}_{AB} = \mathbf{0}$.

2.2.3 Condizioni di indipendenza

E' evidente che l'indipendenza fra tutte le componenti di $\mathbf{Y}$ si può avere solo quando la $f_{\mathbf{Y}}(\mathbf{y})$ è fattorizzabile nelle rispettive densità marginali, il che può avvenire se (*e solo se*) $\boldsymbol{\Sigma}$ è diagonale, ossia con covarianze nulle, e quindi correlazioni lineari semplici nulle, il che porta un'altra fondamentale proprietà della normale multivariata:

[1] ottenere lo stesso risultato integrando la densità (2.3) rispetto alle componenti di $\mathbf{Y}_B$ sarebbe molto più complicato

Indipendenza e assenza di correlazione nella normale multivariata

Assenza di correlazione $\iff$ Indipendenza (nella normale multivariata)

Un vettore aleatorio $\mathbf{Y}$ con distribuzione normale multivariata, è a componenti indipendenti se *(e solo se)* le correlazioni lineari fra le sue componenti prese a due a due sono nulle, ossia se la matrice di varianza e covarianza è diagonale. Quindi, se due variabili sono congiuntamente normali, l'assenza di correlazione implica l'indipendenza stocastica.

ATTENZIONE:

Le cose notevoli sono due (almeno):

- Due variabili con distribuzione congiunta normale bivariata sono indipendenti **se e solo se** sono non correlate
- L'indipendenza a due a due (e quindi la non correlazione a coppie) implica l'indipendenza totale delle componenti

2.3 Distribuzione di combinazioni lineari di variabili normali qualsiasi.

Qualsiasi combinazione lineare (con un numero qualsiasi di componenti) di un vettore aleatorio distribuito secondo una qualsiasi normale multivariata si distribuisce ancora secondo una distribuzione normale multivariata:

Distribuzione congiunta di combinazioni lineari di normali multivariate

Se $\mathbf{Y} = \mathbf{A}^{\mathsf{T}}\mathbf{Z}$, con $\mathbf{A}$ $(p \times k)$, e $\mathbf{Z}(\mathcal{N}_p(\boldsymbol{\mu}_{\mathbf{Z}}, \boldsymbol{\Sigma}_{\mathbf{Z}})$ allora:

$$\mathbf{Y} \sim \mathcal{N}_k(\mathbf{A}^{\mathsf{T}}\boldsymbol{\mu}_{\mathbf{Z}}, \mathbf{A}^{\mathsf{T}}\boldsymbol{\Sigma}_{\mathbf{Z}}\mathbf{A})$$

Mediante la funzione caratteristica è possibile vedere ora che qualsiasi insieme di k combinazioni lineari di un vettore aleatorio distribuito secondo una qualsiasi normale multivariata si distribuisce ancora secondo una distribuzione normale multivariata:

Infatti dal momento che se $\mathbf{Y} = \mathbf{A}^\mathsf{T}\mathbf{Z}$, si ha:

$$\phi_{\mathbf{Y}}(\mathbf{t}_k) = \phi_{\mathbf{Z}}(\mathbf{A}\mathbf{t}_k),$$

se $\mathbf{Z} \sim (N_p(\boldsymbol{\mu}_{\mathbf{Z}}, \boldsymbol{\Sigma}_{\mathbf{Z}})$ allora:

$$\phi_{\mathbf{Z}}(\mathbf{t}_p) = \exp\left[i\mathbf{t}_p^\mathsf{T}\boldsymbol{\mu}_{\mathbf{Z}} - \frac{1}{2}\mathbf{t}_p^\mathsf{T}\boldsymbol{\Sigma}_{\mathbf{Z}}\mathbf{t}_p\right]$$

e quindi:

$$\phi_{\mathbf{Y}}(\mathbf{t}_k) = \phi_{\mathbf{Z}}(\mathbf{A}\mathbf{t}_k) = \exp\left[i\mathbf{t}_k^\mathsf{T}\mathbf{A}^\mathsf{T}\boldsymbol{\mu}_{\mathbf{Z}} - \frac{1}{2}\mathbf{t}_k^\mathsf{T}\mathbf{A}^\mathsf{T}\boldsymbol{\Sigma}_{\mathbf{Z}}\mathbf{A}\mathbf{t}_k\right]$$

per cui è immediato vedere che questa è ancora la funzione caratteristica di una normale multivariata di k componenti con parametri $\mathbf{A}\boldsymbol{\mu}_{\mathbf{Z}}$ e $\mathbf{A}\boldsymbol{\Sigma}_{\mathbf{Z}}\mathbf{A}^\mathsf{T}$.

2.4 Caratterizzazione della distribuzione normale multivariata.

Le proprietà viste prima sulla distribuzione congiunta di combinazioni lineari di variabili normali costituiscono una caratterizzazione della distribuzione normale multivariata.

Infatti ricordo una importante proprietà che caratterizza la distribuzione normale multivariata (di cui non fornisco la dimostrazione) (Mardia, 1970):

Caratterizzazione della distribuzione normale multivariata.

$\mathbf{X}$, vettore aleatorio a p componenti, è distribuito secondo una normale multivariata *se e solo se* $\mathbf{b}^\mathsf{T}\mathbf{X}$ è distribuito secondo una normale (univariata) per qualsiasi vettore $\mathbf{b}$ di p componenti.

2.5 Assi principali degli ellissoidi di equiprobabilità

E' immediato vedere che le superfici ($p-$dimensionali) con densità $f(\mathbf{y})$ costante per la normale multivariata di parametri $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ sono, in uno spazio p -dimensionale, degli ellissoidi di centro in $\boldsymbol{\mu}$; infatti le curve a densità costante hanno equazione:

$$|\boldsymbol{\Sigma}|^{-\frac{1}{2}}(2\pi)^{-p/2}\exp(-[\mathbf{y}-\boldsymbol{\mu}]^\mathsf{T}\boldsymbol{\Sigma}^{-1}[\mathbf{y}-\boldsymbol{\mu}]/2) = k_0$$

essendo k_0 una costante non negativa ($0 \leq k_0 \leq |\boldsymbol{\Sigma}|^{-\frac{1}{2}}(2\pi)^{-p/2}$), e quindi:

$$[\mathbf{y}-\boldsymbol{\mu}]^\mathsf{T}\boldsymbol{\Sigma}^{-1}[\mathbf{y}-\boldsymbol{\mu}] = k_1$$

che è palesemente l'equazione di un ellissoide di centro in $\boldsymbol{\mu}$ e in cui k_1 è un termine costante (rispetto a $\mathbf{x}$), dato da:

$$k_1 = 2\log\left(|\boldsymbol{\Sigma}|^{-\frac{1}{2}}(2\pi)^{-\frac{p}{2}}\right) - 2\log k_0$$

- E' facile vedere quindi che al variare del livello costante k_1, cambia solo il volume dell'ellissoide, ma le proporzioni fra gli assi e le coordinate del centro dell'ellissoide restano inalterate;
- le equazioni degli assi principali di tali ellissoidi sono date dagli autovettori di $\boldsymbol{\Sigma}$, che forniscono l'orientamento degli assi.
- i quadrati delle lunghezze degli assi principali di tali ellissoidi sono proporzionali agli autovalori di $\boldsymbol{\Sigma}$.
- Se $\boldsymbol{\Sigma}$ è diagonale, gli ellissoidi hanno assi paralleli agli assi coordinati e lunghezza proporzionale agli scarti quadratici medi delle singole componenti.
- Si può fare vedere che gli autovettori danno le direzioni degli assi principali impostando ancora un problema di massimo, ossia cercando i due punti sulla superficie dell'ellissoide che hanno distanza massima.

Ellissoidi e Normali multivariate

Fissato un qualsiasi valore positivo di k_1, esiste una corrispondenza biunivoca fra ellissoidi in $\Re_p$ e distribuzioni normali multivariate non singolari.

Esempi e grafici sulla normale trivariata
Normale trivariata a media nulla con Matrice di varianze e covarianze:
parametric_ellissoide1_gr_3.gif
Con autovalori:
parametric_ellissoide1_gr_5.gif
Ellissoidi di equidensità
(sono due sezioni tridimensionali della densità (a 4D)
Scalato in modo tale che la probabilità che un punto risulti interno all'ellissoide risulti del 90%.
parametric_ellissoide1_gr_11.gif
Scalato in modo tale che la probabilità che un punto risulti interno all'ellissoide è del 50%
parametric_ellissoide1_gr_14.gif
Normale trivariata a media nulla con Matrice di varianze e covarianze:
parametric_ellissoide1_gr_17.gif
Con autovalori:
parametric_ellissoide1_gr_19.gif

La varianza misurata lungo gli assi originari è 1 per tutte e tre le componenti.
Ellissoidi di equidensità (sono due sezioni tridimensionali della densità (a 4D)
Scalato in modo tale che la probabilità che un punto risulti interno all'ellissoide sia del 90%.
parametric_ellissoide1_gr_25.gif
Scalato in modo tale che la probabilità che un punto risulti interno all'ellissoide sia del 50%.
parametric_ellissoide1_gr_27.gif
parametric_ellissoide1_gr_34.gif
la varianza massima $2,6$, su un totale di tre, è dovuta al primo asse; gli autovalori sono infatti: **trovare e finire esempio**
Dalla figura a fianco si vedono le caratteristiche della distribuzioni condizionate.
Normale trivariata a media nulla con Matrice di varianze e covarianze:
parametric_ellissoide1_gr_41.gif
Ellissoide di equidensità (è una sezione tridimensionale della densità (a 4D)
Scalato in modo tale che la probabilità che un punto risulti interno all'ellissoide è del 50parametric_ellissoide1_gr_49.gif

2.5.1 Probabilità integrali e adattamento a dati reali

E' appena il caso di dire che il calcolo delle probabilità integrali su domini rettangolari della normale multivariata è estremamente complesso, e comunque non riconducibile a trasformazioni semplici di integrali unidimensionali, se le variabili sono correlate.

Ancora va chiarito, sulla genesi della normale multivariata utilizzata in queste pagine, che questa è una impostazione utile per ricavare la distribuzione di combinazioni lineari di variabili normali indipendenti: nell'analisi di fenomeni reali, ovviamente, non è quasi mai ragionevole pensare che delle variabili osservate correlate siano state effettivamente ottenute come combinazioni lineari di fattori o variabili non correlate, anche se ovviamente è possibile, come si vede nell'analisi delle componenti principali, operare una rotazione per ricavare variabili non correlte, che non necessariamente corrispondono però a variabili osservabili o dotate di significato, ma che tuttavia forniscono utili strumenti operativi almeno nell'ambito di un'analisi esplorativa dei dati multivariati `esempio da fare con file PCA Rmd`

2.6 Distribuzioni condizionate nella normale multivariata

Una proprietà fondamentale della normale, che oltretutto la caratterizza, riguarda le distribuzioni di un gruppo di componenti condizionatamente ai valori di un altro gruppo di componenti. Questo argomento viene trattato adesso, esponendo i risultati fondamentali per tre ordini di ragioni:

1. la peculiarità delle caratteristiche delle distribuzioni condizionate nella normale multivariata, che ne rappresentano un aspetto fondamentale;

2. come premessa *quasi* indispensabile ai modelli lineari, che costituiscono il punto centrale dei miei appunti;

3. la possibilità di dare un significato statistico e probabilistico autonomo agli elementi della inversa della matrice di correlazione di una variabile con distribuzione normale multivariata.

Si è già visto molto semplicemente che nella normale bivariata la distribuzione condizionata di una variabile è normale, con varianza costante e speranza matematica che varia linearmente in funzione dei valori assunti dalla variabile condizionante.

Prevedibilmente, il risultato per la normale multivariata generalizza quello della normale bivariata, fatte le opportune modifiche per il passaggio ad una situazione multivariata; come si vedrà nelle pagine successive, la distribuzione di un gruppo di variabili $\mathbf{Y_A}$ condizionata ad un particolare valore $\mathbf{y_B}$ assunto da un altro gruppo di variabili $\mathbf{Y_B}$ è:

1. ancora normale multivariata;

2. con una matrice di varianze e covarianze costante per qualsiasi $\mathbf{y_B}$, ossia che non dipende dai valori della componente condizionante (omoscedasticità);

3. con valore atteso di una componente $\mathbf{y}_A$ rispetto alle altre componenti lineare (e quindi funzioni di regressione lineari).

I risultati esposti in queste pagine generalizzano le proprietà note per distribuzioni normali bivariate, viste nella (4).

2.6.1 Distribuzione condizionata nel caso generale di un gruppo di componenti rispetto ad un altro gruppo di componenti.

Supponiamo di avere un vettore $\mathbf{Y}$ di p componenti, con distribuzione normale multivariata, suddiviso nel caso più generale in due sottovettori $[\mathbf{Y_A}, \mathbf{Y_B}]$, con corrispondente suddivisione del vettore delle medie e della matrice di varianze e covarianze:

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_A \\ \mathbf{Y}_B \end{pmatrix} \qquad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_B \end{pmatrix} \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{AA} & \boldsymbol{\Sigma}_{AB} \\ \boldsymbol{\Sigma}_{AB}^{\mathsf{T}} & \boldsymbol{\Sigma}_{BB} \end{pmatrix}$$

I due insiemi di indici A e B costuituiscono una partizione dell'insieme di indici $I = 1, 2, \ldots, p$ così che:

$$A \cup B = I \qquad A \cap B = \emptyset \qquad A \neq \emptyset \qquad B \neq \emptyset$$

per il resto A e B sono costituiti da sottoinsiemi di indici qualsiasi (con la restrizione che esistano le inverse delle matrici di varianze e covarianze che si richiederanno nel seguito).

In effetti i casi più rilevanti, che poi dettaglieremo e che sono più importanti, sono quelli in cui:

- **$A = \{i\}$, per lo studio della distribuzione della variabile i-esima condizionatamente alle altre $p-1$;**
- **$A = \{i, j\}$, per lo studio della distribuzione di due variabili (la i-esima e la j-esima), condizionatamente alle altre $p-2$,** in particolare per lo studio della *indipendenza condizionata.*

Ci chiediamo qual è la funzione di regressione di $\mathbf{Y}_A$ su $\mathbf{Y}_B$, ossia la speranza matematica di $\mathbf{Y}_A$ condizionata ad un particolare valore $\mathbf{y}_B$ di $\mathbf{Y}_B$: in simboli $\mathrm{E}\left[\mathbf{Y}_A | \mathbf{Y}_B = \mathbf{y}_B\right]$.

Più in generale ci chiediamo direttamente qual è la distribuzione di $\mathbf{Y}_A$ condizionata ad un particolare valore $\mathbf{y}_B$ di $\mathbf{Y}_B$.

Passaggi per ricavare la densità delle distribuzione di condizionate nella normale multivariata

Per comodità lavoriamo con variabili $\mathbf{X}_A$, $\mathbf{X}_B$ con speranze matematiche nulle, ponendo:

$$\mathbf{X}_A = \mathbf{Y}_A - \boldsymbol{\mu}_A$$
$$\mathbf{X}_B = \mathbf{Y}_B - \boldsymbol{\mu}_B$$

Ovviamente la matrice di varianze e covarianze di $\mathbf{X}$ è uguale a quella di $\mathbf{Y}$ essendo $\mathbf{X}$ ottenuto da $\mathbf{Y}$ per traslazione:

$$\mathrm{V}\left[\mathbf{X}\right] = \mathrm{V}\left[\mathbf{Y}\right] = \boldsymbol{\Sigma}$$

E' opportuno richiamare le formule per la semplificazione [a] degli elementi dell'inversa della matrice partizionata delle varianze e covarianze di $\mathbf{Y}$ o di $\mathbf{X}$:

$$\boldsymbol{\Sigma}^{-1} = \left(\begin{array}{c|c} \boldsymbol{\Sigma}_{AA.B}^{-1} & -\boldsymbol{\Sigma}_{AA.B}^{-1}\boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1} \\ \hline -\boldsymbol{\Sigma}_{AA.B}^{-1}\boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1} & \boldsymbol{\Sigma}_{BB}^{-1}\left(\boldsymbol{\Sigma}_{AB}^{\mathsf{T}}\boldsymbol{\Sigma}_{AA.B}^{-1}\boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1} + \mathbf{I}\right) \end{array}\right)$$

avendo posto:

$$\boldsymbol{\Sigma}_{AA.B} = \boldsymbol{\Sigma}_{AA} - \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}\boldsymbol{\Sigma}_{AB}^{\mathsf{T}}.$$

Indichiamo con $\boldsymbol{\Sigma}^{IJ}$ il blocco corrispondente al posto di $\boldsymbol{\Sigma}_{IJ}$ nell'inversa $\boldsymbol{\Sigma}^{-1}$, avendo indicato con I e J due generici insiemi di indici ($I = A, B$; $J = A, B$), così che l'inversa sia data da:

$$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \boldsymbol{\Sigma}^{AA} & \boldsymbol{\Sigma}^{AB} \\ \boldsymbol{\Sigma}^{BA} & \boldsymbol{\Sigma}^{BB} \end{pmatrix} \qquad \text{con:}$$

$$
\begin{aligned}
\boldsymbol{\Sigma}^{AA} &= \boldsymbol{\Sigma}_{AA.B}^{-1};\\
\boldsymbol{\Sigma}^{AB} &= -\boldsymbol{\Sigma}_{AA.B}^{-1}\boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1};\\
\boldsymbol{\Sigma}^{BA} &= -\boldsymbol{\Sigma}_{BB}^{-1}\boldsymbol{\Sigma}_{AB}^{\mathsf{T}}\boldsymbol{\Sigma}_{AA.B}^{-1} = \left(\boldsymbol{\Sigma}^{AB}\right)^{\mathsf{T}};\\
\boldsymbol{\Sigma}^{BB} &= \boldsymbol{\Sigma}_{BB}^{-1}\left[\boldsymbol{\Sigma}_{AB}^{\mathsf{T}}\boldsymbol{\Sigma}_{AA.B}^{-1}\boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1} + \mathbf{I}\right].
\end{aligned}
$$

Non si confonda ad esempio $\boldsymbol{\Sigma}^{AA}$ (blocco dell'inversa $\boldsymbol{\Sigma}^{-1}$ corrispondente agli indici AA) con $\boldsymbol{\Sigma}_{AA}^{-1}$ (inversa del blocco di $\boldsymbol{\Sigma}$ corrispondente agli indici AA) (coincidono solo se $\boldsymbol{\Sigma}_{AB} = \mathbf{0}$)
Adesso siamo pronti per ricavare dai noti teoremi del calcolo delle probabilità la densità della distribuzione di $\mathbf{X}_A$ condizionatamente ad un valore fissato $\mathbf{x}_B$ del vettore aleatorio $\mathbf{X}_B$:

$$
f_{\mathbf{X}_A|\mathbf{X}_B=\mathbf{x}_B}(\mathbf{x}_A|\mathbf{X}_B = \mathbf{x}_B) = \frac{f_{\mathbf{X}_A\mathbf{X}_B}(\mathbf{x}_A, \mathbf{x}_B)}{f_{\mathbf{X}_B}(\mathbf{x}_B)}
$$

E' più comodo lavorare sui logaritmi ed in particolare su $-2\log f$ (in modo da trasformare solo le forme quadratiche a numeratore dell'esponente nella densità normale), indicando per brevità con K la costante di normalizzazione, che si può determinare dopo:

$$
\begin{aligned}
&-2\log[f(\mathbf{x}_A, \mathbf{x}_B)/f(\mathbf{x}_B)] =\\
&\quad= K + \mathbf{x}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{x} - \mathbf{x}_B^{\mathsf{T}}\boldsymbol{\Sigma}_{BB}^{-1}\mathbf{x}_B =\\
&\quad= K + \mathbf{x}_A^{\mathsf{T}}\boldsymbol{\Sigma}^{AA}\mathbf{x}_A + 2\mathbf{x}_A^{\mathsf{T}}\boldsymbol{\Sigma}^{AB}\mathbf{x}_B + \mathbf{x}_B^{\mathsf{T}}\boldsymbol{\Sigma}^{BB}\mathbf{x}_B - \mathbf{x}_B^{\mathsf{T}}\boldsymbol{\Sigma}_{BB}^{-1}\mathbf{x}_B =\\
&\qquad\text{(sostituendo gli opportuni blocchi di } \boldsymbol{\Sigma}^{-1}\text{)}\\
&\quad= K + \mathbf{x}_A^{\mathsf{T}}\boldsymbol{\Sigma}_{AA.B}^{-1}\mathbf{x}_A - 2\mathbf{x}_A^{\mathsf{T}}\boldsymbol{\Sigma}_{AA.B}^{-1}\boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}\mathbf{x}_B +\\
&\qquad+\mathbf{x}_B^{\mathsf{T}}\boldsymbol{\Sigma}_{BB}^{-1}\left[\boldsymbol{\Sigma}_{AB}^{\mathsf{T}}{}^{\mathsf{T}}\boldsymbol{\Sigma}_{AA.B}^{-1}\boldsymbol{\Sigma}_{AB}^{\mathsf{T}}\boldsymbol{\Sigma}_{BB}^{-1} + \mathbf{I}\right]\mathbf{x}_B - \mathbf{x}_B^{\mathsf{T}}\boldsymbol{\Sigma}_{BB}^{-1}\mathbf{x}_B =\\
&\quad= K + \mathbf{x}_A^{\mathsf{T}}\boldsymbol{\Sigma}_{AA.B}^{-1}\mathbf{x}_A - 2\mathbf{x}_A^{T}\boldsymbol{\Sigma}_{AA.B}^{-1}[\boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}\mathbf{x}_B] +\\
&\qquad+[\mathbf{x}_B^{\mathsf{T}}\boldsymbol{\Sigma}_{BB}^{-1}\boldsymbol{\Sigma}_{AB}^{\mathsf{T}}]\boldsymbol{\Sigma}_{AA.B}^{-1}[\boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}\mathbf{x}_B] =\\
&\quad= K + \left(\mathbf{x}_A - \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}\mathbf{x}_B\right)^{\mathsf{T}}\boldsymbol{\Sigma}_{AA.B}^{-1}\left(\mathbf{x}_A - \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}\mathbf{x}_B\right)
\end{aligned}
$$

Per cui è chiaro dall'ultima forma quadratica, che si tratta del numeratore dell'esponente di una distribuzione normale di parametri:

$$
\begin{aligned}
\boldsymbol{\mu}_{\mathbf{X}_A|\mathbf{X}_B=\mathbf{x}_B} &= \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}\mathbf{x}_B\\
\boldsymbol{\Sigma}_{\mathbf{X}_A|\mathbf{X}_B=\mathbf{x}_B} &= \boldsymbol{\Sigma}_{AA.B} = \boldsymbol{\Sigma}_{AA} - \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}\boldsymbol{\Sigma}_{AB}^{\mathsf{T}} = (\boldsymbol{\Sigma}^{AA})^{-1}
\end{aligned}
$$

La costante K è ricavabile dalla condizione di normalizzazione, ma si può comunque verificare effettuando il rapporto fra i termini costanti delle due densità, tenendo presente che per matrici partizionate si ha:

$$
|\boldsymbol{\Sigma}| = |\boldsymbol{\Sigma}_{AA} - \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}\boldsymbol{\Sigma}_{AB}^{T}||\boldsymbol{\Sigma}_{BB}| = |\boldsymbol{\Sigma}_{AA.B}||\boldsymbol{\Sigma}_{BB}|
$$

Per cui la distribuzione condizionata è:

$$
\mathbf{X}_{A|\mathbf{x}_B} \sim \mathcal{N}\left[\boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}\mathbf{x}_B; \boldsymbol{\Sigma}_{AA.B}\right]
$$

Ricordando che all'inizio della dimostrazione completa per comodità si era posto: $\mathbf{X}_A = \mathbf{Y}_A - \boldsymbol{\mu}_A$ e $\mathbf{X}_B = \mathbf{Y}_B - \boldsymbol{\mu}_B$, si risostituiscono $\mathbf{Y}_A, \mathbf{Y}_B$ in luogo degli scarti $\mathbf{X}_A, \mathbf{X}_B$ e si ha il risultato finale.

[a]Lo so, lo so, non sembra una semplificazione ma una complicazione, ma l'eleganza del risultato finale ci ripagherà...; ma ribadisco che si può saltare questa dimostrazione e saltare al risultato.

Distribuzioni condizionate nel caso generale di vettori aleatori normali

$$\mathbf{Y}_A | \{\mathbf{Y}_B = \mathbf{y}_B\} \sim \mathcal{N}_{n_A}\left[\boldsymbol{\mu}_A + \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}(\mathbf{y}_B - \boldsymbol{\mu}_B); \boldsymbol{\Sigma}_{AA.B}\right]$$

1. **La distribuzione condizionata è normale multivariata con parametri:**

$$\begin{aligned} \mathrm{E}\left[\mathbf{Y}_A | \{\mathbf{Y}_B = \mathbf{y}_B\}\right] &= \boldsymbol{\mu}_A + \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}(\mathbf{y}_B - \boldsymbol{\mu}_B); & (2.6) \\ \mathrm{V}\left[\mathbf{Y}_A | \{\mathbf{Y}_B = \mathbf{y}_B\}\right] &= \boldsymbol{\Sigma}_{AA.B}. & (2.7) \end{aligned}$$

2. **la funzione di regressione (speranza matematica condizionata) è lineare in $\mathbf{y}_B$;**

$$g(\mathbf{y}_B) = \boldsymbol{\mu}_A + \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}(\mathbf{y}_B - \boldsymbol{\mu}_B);$$

3. **la matrice di varianze e covarianze condizionate non dipende da $\mathbf{y}_B$ (omoscedasticità)** ed è espressa da:

$$\mathrm{V}\left[\mathbf{Y}_A | \{\mathbf{Y}_B = \mathbf{y}_B\}\right] = \boldsymbol{\Sigma}_{AA.B} = \boldsymbol{\Sigma}_{AA} - \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}\boldsymbol{\Sigma}_{AB}^{\mathsf{T}} =$$

$$= (\boldsymbol{\Sigma}^{AA})^{-1}$$

$\boldsymbol{\Sigma}^{AA}$ è il blocco di $\boldsymbol{\Sigma}^{-1}$ corrispondente agli indici AA

I vettori aleatori:

$$\mathbf{Y}_A - \boldsymbol{\mu}_A - \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}[\mathbf{Y}_B - \boldsymbol{\mu}_B] \qquad \text{e} \qquad \mathbf{Y}_B$$

(oppure $\mathbf{Y}_A - \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}\mathbf{Y}_B$ e $\mathbf{Y}_B$) risultano indipendenti (si verifica subito calcolando $E(\mathbf{Y}_A\mathbf{Y}_B^{\mathsf{T}})$)

Esempio numerico: Si consideri la matrice 3×3 di varianza e covarianza relativa ad una distribuzione normale multivariata a tre componenti:

$$\mathbf{\Sigma} = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix} \quad \text{e quindi correlazioni:}$$

$$\mathbf{R} = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{\sqrt{2}} \\ \frac{1}{2} & 1 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 1 \end{pmatrix} = \begin{pmatrix} 1.00 & 0.50 & 0.71 \\ 0.50 & 1.00 & 0.71 \\ 0.71 & 0.71 & 1.00 \end{pmatrix}$$

Consideriamo la distribuzione della variabile 1 condizionatamente alla 2 e alla 3. La matrice di varianze e covarianze va quindi partizionata nel seguente modo:

$$\mathbf{\Sigma} = \left(\begin{array}{c|cc} 2 & 1 & 1 \\ \hline 1 & 2 & 1 \\ 1 & 1 & 1 \end{array}\right).$$

Con riferimento alla notazione fin qui adottata l'insieme A è costituito dal solo indice 1 e l'insieme B dagli altri due indici (2 e 3); per cui si ha:

$$A = 1 \qquad B = 2,3 \qquad \text{e quindi:}$$

$$\mathbf{\Sigma}_{BB} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \qquad \text{e} \qquad \mathbf{\Sigma}^{-1} = \left(\begin{array}{r|rr} 1 & 0 & -1 \\ \hline 0 & 1 & -1 \\ -1 & -1 & 3 \end{array}\right)$$

$$\mathbf{\Sigma}^{AA} = 1$$

2.6.2 Significato degli elementi dell'inversa della matrice di varianze e covarianze.

E' possibile dare anche un significato agli elementi dell'inversa di $\mathbf{\Sigma}$, in termini di distribuzioni condizionate, essendo al solito $\mathbf{\Sigma}$ la matrice di varianze e covarianze di una distribuzione normale multivariata; si vedrà in altri capitoli se e in che misura tali concetti possano essere estesi al caso di variabili aleatorie non normali o, anche, nell'analisi di dati multivariati, al caso di variabili statistiche osservate.

`esempio da fare con file esempiocor Rmd`

`esempio da fare con file trivariatenormal Rmd`

Gli elementi diagonali dell'inversa: la correlazione multipla

Anche gli elementi sulla diagonale principale di $\mathbf{\Sigma}^{-1}$ sono interpretabili tenendo conto delle distribuzioni condizionate, ma in termini di variabilità di una variabile spiegata da tutte le altre, concetto che rivedremo poi nel caso di modelli lineari generali.

Infatti consideriamo ora l'insieme $\mathbf{Y}_A$ costituito da una sola variabile Y_i; nella notazione finora adottata A è uguale all'indice i e B all'insieme degli altri $p-1$

indici, ossia:

$$A = \{i\} \qquad B = \{1, 2, \ldots, i-1, i+1, \ldots, p\}$$

possiamo calcolare la varianza di $\mathbf{y}_i$ condizionata ai valori delle altre $p-1$ variabili, applicando la 2.7:

$$\mathrm{V}\left[\mathbf{Y}_A \middle| \{\mathbf{Y}_B = \mathbf{y}_B\}\right] = \mathrm{V}\left[\mathbf{Y}_i \middle| \{\mathbf{Y}_B = \mathbf{y}_B\}\right] = \left(\mathbf{\Sigma}^{i\,i}\right)^{-1}$$

Tenendo conto che in questo caso $\mathbf{\Sigma}^{AA} = \mathbf{\Sigma}^{i\,i} = c_{ii}$ si ha:

Varianza della distribuzione condizionata di una componente

$$\mathrm{V}\left[\mathbf{y}_i | \mathbf{y}_B\right] = (\mathbf{\Sigma}^{AA})^{-1} = \frac{1}{c_{ii}} = \frac{|\mathbf{\Sigma}|}{\sigma^{ii}}$$

Quindi l'inverso di un elemento diagonale dell'inversa della matrice di varianze e covarianze esprime la varianza della variabile di posto corrispondente condizionatamente alle altre $p-1$ variabili.

$$\max(\frac{1}{c_{ii}}) = \sigma_i^2 \qquad \min(c_{ii}) = \frac{1}{\sigma_i^2}$$

Il massimo di questa quantità è proprio la varianza della componente i -esima, ossia σ_i^2

Se $\mathbf{Z}$ è una matrice di correlazione, allora $1/c_{ii}$ indica la variabilità di Y_i non spiegata dalle altre $p-1$ variabili, per cui si può costruire il coefficiente di determinazione multipla:

$$R^2_{i.B} = 1 - \frac{|\mathbf{Z}|}{z^{ii}} = 1 - \frac{\mathrm{V}\left[\mathbf{y}_i\right] |\mathbf{Y}_B)}{\mathrm{V}\left[\mathbf{y}_i\right]}$$

Misura quanta parte della variabilità di $\mathbf{Y}_i$ è spiegata dalle altre p-1 variabili del vettore aleatorio $\mathbf{y_B}$

In generale l'indice di correlazione lineare multipla è dato da:

$$R_{i.B} = \sqrt{1 - \frac{|\mathbf{\Sigma}|}{\sigma_i^2 \sigma^{ii}}} = \sqrt{1 - \frac{1}{\sigma_i^2 c_{ii}}}$$

`esempio da fare con file inversa Rmd`

Gli elementi non diagonali dell'inversa: la correlazione parziale

Intanto, con riferimento ad una distribuzione normale multivariata con matrice di varianze e covarianze $\mathbf{\Sigma}$, si può dimostrare che, indicata per brevità con $\mathbf{C}$ la sua inversa, $\mathbf{C} = \mathbf{\Sigma}^{-1}$, allora:

$c_{ij} = 0$ è condizione necessaria e sufficiente perché le variabili $\mathbf{Y}_i$ e $\mathbf{Y}_j$ siano indipendenti condizionatamente alle altre $p-2$ variabili $\mathbf{Y}_B$.

Dettagli sull'indipendenza condizionata

Dalla densità normale multivariata si vede direttamente che:
se e solo se $c_{ij} = 0$ si ha la fattorizzazione:

$$f(\mathbf{y}) = f(\mathbf{y}_i, \mathbf{y}_B) f(\mathbf{y}_j, \mathbf{y}_B)$$

che è una condizione necessaria e sufficiente per l'indipendenza condizionata di due variabili aleatorie qualsiasi dotate di densità.
Infatti, ponendo $\mathbf{Y}_A = (\mathbf{y}_i, \mathbf{y}_j)^T$ e indicando con $\mathbf{Y}_B$ tutte le altre componenti, avendo indicato con $\mathbf{C}$ l'inversa della matrice di varianze e covarianze opportunamente partizionata:

$$\mathbf{C} = \left(\begin{array}{cc|c} c_{ii} & c_{ij} & \mathbf{c}_{iB}^{\mathsf{T}} \\ c_{ij} & c_{jj} & \mathbf{c}_{jB}^{\mathsf{T}} \\ \hline \mathbf{c}_{iB} & \mathbf{c}_{jB} & \mathbf{C}_{BB} \end{array}\right)$$

(per comodità di lettura ho portato gli indici (i, j) in prima e seconda posizione); si ha allora:

$$f(\mathbf{y}) \quad = \quad f(\mathbf{y}_i, \mathbf{y}_j, \mathbf{y}_B) = K \exp[-\frac{1}{2}\mathbf{y}^{\mathsf{T}}\mathbf{C}\mathbf{y}] =$$

$$= K \exp[-\frac{1}{2}(c_{ii}\mathbf{y}_i^2 + c_{jj}\mathbf{y}_j^2 + 2c_{ij}\mathbf{y}_i\mathbf{y}_j + 2y_i\mathbf{c}_{iB}^{\mathsf{T}}\mathbf{y}_B + 2y_j\mathbf{c}_{jB}^{\mathsf{T}}\mathbf{y}_B + \mathbf{y}_B^T\mathbf{C}_{BB}\mathbf{y}_B)]$$

Se ora $c_{ij} = 0$ allora si può facilmente operare su $f(\mathbf{y})$:

$$f(\mathbf{y}) = K \exp[-\frac{1}{2}(c_{ii}\mathbf{y}_i^2 + c_{jj}\mathbf{y}_j^2 + 2y_i\mathbf{c}_{iB}^{\mathsf{T}}\mathbf{y}_B + 2y_j vecc_{jB}^{\mathsf{T}}\mathbf{y}_B + \mathbf{y}_B^T\mathbf{C}_{BB}\mathbf{y}_B)] =$$

$$\underbrace{K \exp[-\frac{1}{2}(c_{ii}\mathbf{y}_i^2 + 2y_i\mathbf{c}_{iB}^{\mathsf{T}}\mathbf{y}_B + \mathbf{y}_B^{\mathsf{T}}\mathbf{C}_{BB}\mathbf{y}_B)] \times \exp[-(c_{jj}\mathbf{y}_j^2 + 2y_j vecc_{jB}^{\mathsf{T}}\mathbf{y}_B)/2]}_{g(\mathbf{y}_i, \mathbf{y}_B) \times g(\mathbf{y}_j, \mathbf{y}_B)}$$

in modo da ottenere la fattorizzazione desiderata in due funzioni, in cui non compaiono simultaneamente termini in $\mathbf{y}_i$ e $\mathbf{y}_j$

Per una interpretazione in generale del significato dei termini dell'inversa, e non solo per il caso estremo $c_{ij} = 0$, conviene riferirsi alle distribuzioni condizionate.

Dalla distribuzione di $\mathbf{Y}_A$ condizionata a $\mathbf{Y}_B = \mathbf{y}_B$, ponendo $\mathbf{Y}_A = (Y_i, Y_j)^{\mathsf{T}}$ (e quindi nella notazione della sezione precedente A è uguale alla coppia di indici i, j e B all'insieme degli altri $p-2$ indici) si ricava che essendo la distribuzione condizionata di $\mathbf{Y}_A$ ancora normale, l'indipendenza condizionata si ha se e solo se $\mathbf{y}_i$ e $\mathbf{y}_j$ risultano non correlati, condizionatamente a $\mathbf{Y}_B = \mathbf{y}_B$.

Si è visto che:

$$\mathrm{V}\left[(\mathbf{Y}_A|\mathbf{y}_B)\right] = (\mathbf{\Sigma}_{AA.B})^{-1} = \mathbf{\Sigma}_{AA} - \mathbf{\Sigma}_{AB}\mathbf{\Sigma}_{BB}^{-1}\mathbf{\Sigma}_{AB}^{\mathsf{T}} = (\mathbf{\Sigma}^{AA})^{-1}$$

cioè la varianza condizionata di $\mathbf{Y}_A$ è uguale all'inversa del blocco di elementi corrispondenti ad $\mathbf{Y}_A$ nell'inversa di $\mathbf{\Sigma}$.

Nel caso di due variabili i e j, occorre invertire la matrice 2×2 $\mathbf{\Sigma}^{AA}$ di elementi:

$$\mathbf{\Sigma}^{AA} = \begin{pmatrix} c_{ii} & c_{ij} \\ c_{ij} & c_{jj} \end{pmatrix}$$

e quindi:

$$(\mathbf{\Sigma}^{AA})^{-1} = \begin{pmatrix} c_{jj} & -c_{ij} \\ -c_{ij} & c_{ii} \end{pmatrix} / (c_{ii}c_{jj} - c_{ij}^2)$$

pertanto Y_i *e* Y_j *sono non correlati condizionatamente alle altre* $p-2$ *variabili (e quindi indipendenti data la normalità della distribuzione condizionata) se e solo se* $c_{ij} = 0$.

Ricordando la relazione che lega gli elementi c_{hk} dell'inversa di $\mathbf{\Sigma}$, ai cofattori σ^{hk} di σ_{hk} in $\mathbf{\Sigma}$, ossia:

$$c_{hk} = \frac{\sigma^{hk}}{|\mathbf{\Sigma}|}$$

dagli elementi di $(\mathbf{\Sigma}^{AA})^{-1}$ è possibile calcolare l'indice di correlazione lineare fra Y_i e Y_j condizionatamente a $\mathbf{Y}_B$:

Correlazione fra $\mathbf{Y}_i$ e $\mathbf{Y}_j$ (condizionatamente alle altre $p-2$ variabili)

$$corr(\mathbf{y}_i, \mathbf{y}_j|\mathbf{Y}_B = \mathbf{y}_B) = \rho_{ij.B} = \frac{-c_{ij}}{\sqrt{c_{ii}c_{jj}}} = \frac{\frac{-\sigma^{ij}}{|\mathbf{\Sigma}|}}{\sqrt{\frac{\sigma^{ii}}{|\mathbf{\Sigma}|}\frac{\sigma^{jj}}{|\mathbf{\Sigma}|}}} = \frac{-\sigma^{ij}}{\sqrt{\sigma^{ii}\sigma^{jj}}}$$

indice di correlazione lineare parziale ossia correlazione fra due variabili tenute costanti le altre $p-2$ variabili

L'indipendenza condizionata e l'indipendenza marginale sono due concetti diversi, e nessuno dei due implica l'altro.

figure ed esempi vari

```
controllare
inserire lucidi manuali ed esercizio
completare
```

Esempio di variabili condizionatamente non correlate

```
\figura{ese2003_multinormale1.jpg}{distribuzioni condizionate in una normale
multivariata}{ese2003_multinormale1}

\figura{ese2003_multinormale_numeri1.jpg}{distribuzioni condizionate in una
normale multivariatamatrice di varianze e covarianze e
inversa}{ese2003_multinormale_numeri1}

cor1=rbind(c(1,0.333,-.577),
            c(0.333,1,-.577),
            c(-.577,-.577,1))


inv1=rbind(c(1.5,0.0,0.866),
            c(0.0,1.5,0.866),
            c(0.866,0.866,2))

in ese2000\_correlaz1.nb



\figura{fig2000regr4.png}
{relazione fra due variabili in funzione
 del valore di una terza variabile}{regr4}


 grafico 3d con x=altezza, z=peso, y=var.categorica


esempio da fare con file trivariatenormal Rmd
```

da rifare del tutto questo esempio
Da questa matrice di correlazione si è calcolata l'inversa $\mathbf{C} = \mathbf{\Sigma}^{-1}$: e quindi si è calcolata la matrice $\mathbf{A}$ che ha come elemento generico:

$$r_{ij.B} = \frac{-c_{ij}}{\sqrt{c_{ii}c_{jj}}}$$

correlazione parziale fra due variabili, $\mathbf{X}_i$ e $\mathbf{X}_j$, tenute costanti le altre 5:

$$\mathbf{A} =$$

da ricalcolare
(ovviamente in questa matrice gli elementi diagonali non vanno considerati Si vede che la correlazione lineare (parziale, o meglio condizionata, ossia tenute costanti le altre variabili) fra le prime due variabili è di 0,245. Buona

parte quindi della correlazione marginale è indotta dall'influenza delle altre 5 variabili, ossia la covariazione delle prime due variabili insieme alle altre 5.
Se invece trasformiamo gli elementi diagonali di $\mathbf{C}$, mediante la relazione:

$$R^2_{i.B} = 1 - \frac{|\mathbf{Z}|}{z^{ii}} = 1 - \frac{1}{c_{ii}}$$

otteniamo i 7 indici di determinazione multipla, di ciascuna variabile condizionatamente alle altre 6:

$$0.827137, 0.896544, 0.848327, 0.297231, 0.722443, 0.756753, 0.82098$$

Si noti che la matrice di correlazione ha 7 autovalori dati da:

$$\boldsymbol{\lambda}^{\mathsf{T}} = 5.06451, 0.674288, 0.635871, 0.245914, 0.207684, 0.105888, 0.06584$$

La successione di tali valori indica chiaramente la presenza di correlazioni lineari fra combinazioni lineari di variabili molto forti.
`esempio da fare con file codice R Rmd`

Impiego delle informazioni dell'inversa C nell'analisi di dati multivariati.

Come si è visto, l'analisi degli elementi dell'inversa della matrice di correlazione può fornire degli elementi utili per indagare sulla dipendenza condizionata fra variabili.

cenno ai modelli grafici

In alcuni modelli di recente introduzione, i cosiddetti modelli grafici, le ralazioni fra le variabili vengono impostate in modo condizionato, parametrizzando opportunamente un modello di dipendenza normale attraverso gli elementi dell'inversa.

2.6.3 Esempi sulla differenza fra l'indipendenza condizionata e l'indipendenza marginale in tabelle $2 \times 2 \times 2$.

L'indipendenza condizionata e l'indipendenza marginale sono due concetti diversi, e nessuno dei due implica l'altro.

Per chiarire la differenza fra indipendenza marginale e indipendenza condizionata, ricorro qui ad un esempio relativo alla distribuzione di probabilità congiunta di tre variabili dicotomiche A,B e C: è vero che qui sto parlando di variabili normali, ma per spiegare nel modo più semplice possibile la differenza fra indipendenza marginale e indipendenza condizionata, conviene prendere tre sole variabili, e le più elementari, ossia con due sole modalità.

Si ha una tavola $2 \times 2 \times 2$ di tre mutabili A,B, e C ciascuna con due sole modalità. Le due tavole $A \times B$ condizionate ai valori di C sono:

$\mathbf{C} = c_1$	b_1	b_2	$tot.$		$\mathbf{C} = c_2$	b_1	b_2	$tot.$
a_1	$0,24$	$0,06$	$0,30$		a_1	$0,12$	$0,28$	$0,4$
a_2	$0,56$	$0,14$	$0,70$		a_2	$0,18$	$0,42$	$0,6$
$tot.$	$0,80$	$0,20$	$1,00$		$tot.$	$0,30$	$0,70$	$1,00$

In queste distribuzioni condizionate A e B sono indipendenti;
se P(C=c_1)= P(C=c_2)=$\frac{1}{2}$la tavola marginale $A \times B$ è:

$\mathbf{C}tot.$	b_1	b_2	$tot.$
a_1	$0,18$	$0,17$	$0,35$
a_2	$0,37$	$0,28$	$0,65$
$tot.$	$0,55$	$0,45$	$1,00$

Nella distribuzione marginale A e B non sono indipendenti.
Si può presentare il caso opposto, di caratteri indipendenti marginalmente e associati condizionatamente (paradosso di Simpson).
Si ha un'altra tavola $2 \times 2 \times 2$ di tre mutabili A,B, e C. Le due tavole $A \times B$ condizionate ai valori di C sono ora:

$\mathbf{C} = c_1$	b_1	b_2	$tot.$		$\mathbf{C} = c_2$	b_1	b_2	$tot.$
a_1	$0,5$	0	$0,5$		a_1	0	$0,5$	$0,5$
a_2	0	$0,5$	$0,5$		a_2	$0,5$	0	$0,5$
$tot.$	$0,5$	$0,5$	1		$tot.$	$0,5$	$0,5$	1

In queste distribuzioni condizionate A e B sono associati (addirittura sono massimamente associati)
Infatti se P($C = c_1$)= P(C=c_2)=$\frac{1}{2}$la tavola marginale $A \times B$ è:

$\mathbf{C}tot.$	b_1	b_2	$tot.$
a_1	$0,25$	$0,25$	$0,5$
a_2	$0,25$	$0,25$	$0,5$
$tot.$	$0,5$	$0,5$	1

Nella distribuzione marginale A e B sono indipendenti.

2.7 Utilità della distribuzione normale multivariata

In effetti quanto visto finora riguarda solo il modello teorico della normale multivariata, ossia le caratteristiche delle distribuzioni di vettori aleatori normali multivariati, che riassumo brevemente (e solo per le proprietà più rilevanti):

- è unimodale;

- dipende solo dai primi due momenti multivariati;
- la non correlazione è condizione necessaria e sufficiente per l'indipendenza di due componenti;
- l'indipendenza fra tutte le coppie di componenti (e quindi la non correlazione) è condizione necessaria e sufficiente per l'indipendenza fra tutte le componenti;
- ha contorni iper-ellissoidali;
- ha distribuzioni marginali normali multivariate;
- ha distribuzioni condizionate (o parziali) normali multivariate omoschedastiche e con funzioni di regressione lineari;
- combinazioni lineari di sue componenti sono ancora normali multivariate;
- si ottiene come distribuzione limite di processi multivariati come il teorema limite centrale multivariato.

Non ci stiamo per ora ponendo il problema di adattare una tale distribuzione a dati osservati. In effetti in questo corso questo problema verrà affrontato solo marginalmente: l'importanza del modello normale multivariato in questo corso sta nel fatto che è un modello utile per la definizione di relazioni di dipendenza in media *esattamente* lineari ed omoschedastiche, con distribuzione normale attorno al valor medio.

Un altro aspetto rilevante della distribuzione normale multivariata, è che rappresenta, nei casi regolari, la distribuzione limite degli stimatori di massima verosimiglianza: sfrutteremo questa proprietà in capitoli successivi (per esempio nella regressione non lineare o nei modelli lineari generalizzati) quando in assenza di risultati esatti sulla distribuzione degli stimatori, ricorreremo a risultati asintotici.

2.8 Alcuni risultati sulla distribuzione di forme quadratiche in variabili normali.

In questa sezione riassumo alcuni dei risultati più interessanti riguardanti la distribuzione di forme quadratiche in variabili normali: questi risultati sono poco essenziali per quanto riguarda l'essenza della distribuzione normale multivariata, ma rappresentano uno strumento tecnico indispensabile per il successivo studio dei modelli lineari, poichè costituiscono la base teorica per l'inferenza esatta nei modelli lineari classici.

2.8.1 Indipendenza di forme quadratiche e combinazioni lineari di variabili normali.

Sia $\mathbf{X}$ un vettore di variabili casuali a p componenti indipendenti, ciascuna distribuita secondo una normale standardizzata, ossia

$$\mathbf{X} \sim N_p(0_p, \mathbf{I}_p).$$

Valgono alcuni teoremi sull'indipendenza fra forme quadratiche in $\mathbf{X}$ e combinazioni lineari in $\mathbf{X}$, che si basano sulle proprietà dei vettori dei coefficienti che determinano le forme quadratiche e le combinazioni lineari.

- Si abbiano due forme quadratiche in variabili normali indipendenti $\mathbf{X}$:

  $$\mathbf{Q}_1 = \mathbf{X}^\mathsf{T}\mathbf{A}_1\mathbf{X} \qquad \text{e} \qquad \mathbf{Q}_2 = \mathbf{X}^\mathsf{T}\mathbf{A}_2\mathbf{X}$$

  Le due forme quadratiche $\mathbf{Q}_1$ e $\mathbf{Q}_2$ sono indipendenti se e solo se $\mathbf{A}_1\mathbf{A}_2 = \mathbf{0}_{(pxp)}$ (essendo ovviamente $\mathbf{A}_1$ e $\mathbf{A}_2$ matrici quadrate simmetriche, ed essendo $\mathbf{0}_{(p\times p)}$ una matrice quadrata composta di zeri);

- Si abbia la forma quadratica

  $$\mathbf{Q} = \mathbf{X}^\mathsf{T}\mathbf{A}\mathbf{X}$$

  e la combinazione lineare $Z = \mathbf{b}^\mathsf{T}\mathbf{X}$.

  La forma quadratica $\mathbf{Q}$ e la combinazione lineare Z sono indipendenti

  se e solo se $\mathbf{A}\mathbf{b} = \mathbf{0}_p$

  (essendo $\mathbf{b}$ un vettore di p elementi e 0_p il vettore nullo di p componenti)

2.8.2 Distribuzione di forme quadratiche in variabili normali standardizzate e indipendenti.

In questa sezione affrontiamo il problema della distribuzione di particolari forme quadratiche in variabili normali, indipendenti e non: la finalità sarà chiara quando si studieranno le proprietà degli stimatori e dei test nei modelli lineari (modelli di regressione semplice e multipla, di analisi della varianza etc.); si tratta molto semplicemente di generalizzare alcuni risultati noti sulla v.c. χ^2: è ragionevole aspettarsi che forme quadratiche in variabili normali multivariate siano talora riconducibili a variabili χ^2.

Sia $\mathbf{X}$ un vettore di variabili casuali a p componenti indipendenti, ciascuna distribuita secondo una normale standardizzata, ossia

$$\mathbf{X} \sim \mathcal{N}_p\left[\mathbf{0}_p, \mathbf{I}_p\right].$$

E' noto che:

$$\sum_{i=1}^{p} \mathbf{X}_i^2 \sim \chi^2_{p,} \qquad (\text{oppure}\mathbf{X}^\mathsf{T}\mathbf{X} \sim \chi^2_p).$$

In effetti questa è proprio la definizione di una variabile casuale di tipo chi-quadrato con p gradi di libertà, che risulta avere una distribuzione gamma di parametro di forma $\alpha = p/2$ e parametro di scala $\lambda = \frac{1}{2}$.

Più in generale ci chiediamo se si può ricavare la distribuzione di una forma quadratica qualsiasi in variabili normali standardizzate, ossia

$$\mathbf{Q} = \mathbf{X}^\mathsf{T}\mathbf{A}\mathbf{X},$$

e per quali matrici $\mathbf{A}$ questa forma quadratica risulta ancora distribuita come una chi-quadrato.

E' facile vedere che la forma quadratica $\mathbf{Q} = \mathbf{X}^\mathsf{T}\mathbf{A}\mathbf{X}$ si distribuisce come

$$\sum_{i=1}^{p} \lambda_i \chi_1^2$$

, ove i λ_i sono gli autovalori di $\mathbf{A}$.

$$\mathbf{Q} = \mathbf{X}^\mathsf{T}\mathbf{A}\mathbf{X} \sim \sum_{i=1}^{p} \lambda_i \chi_1^2$$

Questo risultato si ricava facilmente dalla decomposizione spettrale della matrice $\mathbf{A}$, in quanto si può scrivere:

$$\mathbf{A} = \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}^\mathsf{T}, \quad \text{per cui:} \quad \mathbf{Q} = \mathbf{X}^\mathsf{T}\mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}^\mathsf{T}\mathbf{X},$$

($\mathbf{\Gamma}$ è una matrice le cui colonne sono gli autovettori di $\mathbf{A}$, (ortogonali: $\mathbf{\Gamma}^\mathsf{T}\mathbf{\Gamma} = \mathbf{I}$, $\mathbf{\Lambda}$ è una matrice diagonale i cui elementi sono gli autovalori di $\mathbf{A}$) e il vettore aleatorio $\mathbf{W} = \mathbf{\Gamma}^\mathsf{T}\mathbf{X}$ è ovviamente normale a componenti standardizzate e indipendenti, data l'ortogonalità di $\mathbf{\Gamma}$ (una rotazione ortogonale di una iper-sfera conduce sempre ad una iper-sfera!). Quindi segue facilmente in modo naturale il risultato scritto prima. Esprimendo in modo più formale:

poniamo $\mathbf{W} = \mathbf{\Gamma}^\mathsf{T}\mathbf{X}$,si ha per i momenti di $\mathbf{W}$:

$$\begin{aligned} \mathrm{E}\left[\mathbf{W}\right] &= \mathbf{\Gamma}^\mathsf{T}\mathrm{E}\left[\mathbf{X}\right] = \mathbf{0} \\ \mathrm{V}\left[\mathbf{W}\right] &= \mathbf{\Gamma}^\mathsf{T}\mathrm{V}\left[\mathbf{X}\right]\mathbf{\Gamma} = \mathbf{\Gamma}^\mathsf{T}\mathbf{I}_p\mathbf{\Gamma} = \mathbf{I}_p \end{aligned}$$

Il vettore aleatorio $\mathbf{W}$ è dunque composto da p variabili normali, standardizzate e indipendenti. Tornando ora alla forma quadratica $\mathbf{Q}$ si ha:

$$\mathbf{Q} = \mathbf{X}^\mathsf{T}\mathbf{A}\mathbf{X} = \mathbf{X}^\mathsf{T}\mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}^\mathsf{T}\mathbf{X} = \mathbf{W}^\mathsf{T}\mathbf{\Lambda}\mathbf{W} = \sum_{i=1}^{p} \lambda_i W_i^2$$

Le componenti W_i^2 (*seqin*) sono chiaramente distribuite come delle chi-quadrato (indipendenti) con un grado di libertà.

Pertanto $\mathbf{Q}$ è distribuita come una combinazione lineare di p variabili casuali chi-quadrato indipendenti con un grado di libertà, con coefficienti dati dagli autovalori di $\mathbf{A}$.

E' quindi possibile calcolare i momenti di $\mathbf{Q}$ in quanto combinazione lineare di v.c. χ_1^2 indipendenti:

$$\begin{aligned} \mathrm{E}\left[\mathbf{Q}\right] &= \sum_{i=1}^{p} \lambda_i \mathrm{E}\left[\chi_1^2\right] = \sum_{i=1}^{p} \lambda_i \\ \mathrm{V}\left[\mathbf{Q}\right] &= \sum_{i=1}^{p} \lambda_i^2 \mathrm{V}\left[(\chi_1^2\right] = 2\sum_{i=1}^{p} \lambda_i^2 \end{aligned}$$

Distribuzione di una forma quadratica in variabili normali

Supponiamo di avere la forma quadratica: $\mathbf{Q} = \mathbf{X}^{\mathsf{T}}\mathbf{A}\mathbf{X}$
Se (*e solo se*) gli autovalori di $\mathbf{A}$ sono tutti uguali a 0 o a 1, ossia *se (e solo se)* $\mathbf{A}$ *è idempotente*, $\mathbf{Q} = \mathbf{X}^{\mathsf{T}}\mathbf{A}\mathbf{X}$ si distribuisce come una variabile casuale χ^2_r, essendo r il rango di $\mathbf{A}$, ossia il numero degli autovalori λ_i uguali ad uno.

Infatti si vede immediatamente che, se $\mathbf{A}$ è idempotente di rango r, si ha:

$$\lambda_1 = \lambda_2 = \ldots = \lambda_r = 1; \qquad \text{e} \qquad \lambda_{r+1} = \lambda_{r+2} = \ldots = \lambda_p = 0;$$

per cui:

$$\sum_{i=1}^{p} \lambda_i \chi_1^2 \sim \sum_{i=1}^{r} 1 \times \chi_1^2 + \sum_{i=r+1}^{p} 0 \times \chi_1^2 = \sum_{i=1}^{r} \chi_1^2 \sim \chi_r^2$$

per la proprietà additiva delle v.c. χ^2.

Per dimostrare che l'idempotenza di $\mathbf{A}$ è condizione necessaria e sufficiente perchè $\mathbf{Q}$ sia distribuita come una chi-quadrato (prima abbiano visto che l'idempotenza di $\mathbf{A}$ è condizione sufficiente), conviene ricorrere alla funzione caratteristica di $\mathbf{Q}$, che è data da:

$$\begin{aligned}\phi_{\mathbf{Q}}(t) &= \mathrm{E}\left[\exp(it\mathbf{X}^{\mathsf{T}}\mathbf{A}\mathbf{X})\right] = \mathrm{E}\left[\exp(it\sum_{i=1}^{p}\lambda_i\mathbf{X}_i^2)\right] = \prod_{i=1}^{p}\mathrm{E}\left[\exp(it\lambda_i\mathbf{X}_i^2)\right] = \\ &= \prod_{i=1}^{p}(1-2it\lambda_i)^{-\frac{1}{2}}\end{aligned}$$

(dato che $\mathbf{X}_i^2$ è distribuito come una chi-quadrato con un grado di libertà, l'ultimo passaggio deriva dalla funzione caratteristica della v.c. chi-quadrato).

Ancora si può osservare che $(1-2it\lambda_i)$ è un autovalore della matrice: $(\mathbf{I}-2it\mathbf{A})$ e quindi la produttoria di tali autovalori $(1-2it\lambda_i)$ è uguale al determinante della suddetta matrice:

$$\phi(t) = \prod_{i=1}^{p}(1-2it\lambda_i)^{-\frac{1}{2}} = |\mathbf{I}-2it\mathbf{A}|^{-\frac{1}{2}}$$

Perché $\mathbf{Q}$ sia distribuita come una chi-quadrato, occorre che la sua funzione caratteristica $\phi_{\mathbf{Q}}(t)$ sia identicamente uguale a quella di una v.c. χ^2 per qualsiasi valore dell'argomento t. La funzione caratteristica di una v.c. χ^2 con ν gradi di libertà è data da:

$$\phi_{\chi^2}(t) = (1-2it)^{-\nu/2},$$

mentre per la funzione caratteristica di $\mathbf{Q}$ si è visto che:

$$\phi_{\mathbf{Q}}(t) = \prod_{i=1}^{p}(1-2it\lambda_i)^{-\frac{1}{2}}.$$

Per avere $\phi_{\chi^2}(t) = \phi_{\mathbf{Q}}(t)$ per qualsiasi t, occorre che i coefficienti λ_i siano o zero o uno, di modo che i corrispondenti termini della produttoria in $\phi_{\mathbf{Q}}(t)$ siano uguali ad uno, se $\lambda_i = 0$, oppure a $(1-2it)^{-\frac{1}{2}}$, se $\lambda_i = 1$; se sono r (rango di $\mathbf{A}$) quelli uguali ad uno, si avrà in definitiva:

$$\phi_{\mathbf{Q}}(t) = (1-2it)^{-r/2},$$

che è la funzione caratteristica di una chi-quadrato con r gradi di libertà.

Ad esempio si consideri la matrice seguente:

$$\mathbf{A} = \begin{pmatrix} 16/25 & 12/25 \\ 12/25 & 9/25 \end{pmatrix}$$

Tale matrice simmetrica risulta idempotente di rango 1, come è facile verificare effettuando il prodotto $\mathbf{AA}$, oppure verificando che $\lambda_1 = 1$ e $\lambda_2 = 0$.
Supponendo di avere un vettore aleatorio $\mathbf{X}$ costituito da due variabili casuali normali standardizzate e indipendenti, X_1 e X_2 la forma quadratica $\mathbf{Q} = \mathbf{X}^{\mathsf{T}}\mathbf{AX}$ risulta data da:

$$\mathbf{Q} = a_{11}X_1^2 + a_{22}X_2^2 + 2a_{12}X_1X_2 = (16X_1^2 + 9X_2^2 + 24X_1X_2)/25,$$

e infine:

$$\mathbf{Q} = [(4/5)X_1 + (3/5)X_2]^2$$

E' immediato verificare che $\mathbf{Q}$ si distribuisce secondo una chi-quadrato con un grado di libertà, senza bisogno di applicare il teorema generale sulla distribuzione delle forme quadratiche. Infatti la variabile:

$$Z = (4/5)X_1 + (3/5)X_2$$

è distribuita normalmente (in quanto combinazione lineare di variabili normali) con media zero e varianza unitaria.
Infatti:

$$\begin{aligned} \mathrm{E}[Z] &= (4/5)\mathrm{E}[X_1] + (3/5)\mathrm{E}[X_2] = 0 \\ \mathrm{V}[Z] &= (4/5)^2\mathrm{V}[X_1] + (3/5)^2\mathrm{V}[X_2] = 16/25 + 9/25 = 1 \end{aligned}$$

($Cov[X_1, X_2] = 0$ per l'indipendenza).
Quindi $\mathbf{Q}$ è uguale al quadrato di una normale standardizzata, e quindi segue una distribuzione chi-quadrato con un grado di libertà.

Forme quadratiche idempotenti:
somma dei quadrati degli scarti dalla media.

Prendiamo ora in esame una forma quadratica già nota [2], ossia la somma dei quadrati degli scarti dalla propria media aritmetica di n variabili casuali normali indipendenti X_i. Tipicamente le variabili saranno quelle corrispondenti ad un campione a n componenti i.i.d. e quindi il vettore aleatorio è $X = \{X_1, X_2, \ldots, X_i, \ldots, X_n\}^{\mathsf{T}}$.

[2] se non fosse già nota al lettore è il momento buono per leggerne una dimostrazione, anche se un po' troppo generale

Interessa dunque la distribuzione della quantità:

$$\mathbf{Q} = \sum_{i=1}^{n} (\mathbf{X}_i - M)^2$$

avendo indicato con M la variabile casuale media aritmetica delle n componenti X_i:

$$M = \sum_{i=1}^{n} X_i / n$$

che si può anche scrivere:

$$M = \frac{\mathbf{1_n}^\mathsf{T} \mathbf{X}}{n},$$

essendo $\mathbf{1_n}$ un vettore di n elementi uguali ad uno:

$$\mathbf{1_n} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

Allora la somma dei quadrati degli scarti si può scrivere in notazione vettoriale con semplici passaggi:

$$\begin{aligned} \mathbf{Q} &= \sum_{i=1}^{n} (\mathbf{X}_i - M)^2 = [\mathbf{X} - \mathbf{1_n} M]^\mathsf{T} [\mathbf{X} - \mathbf{1_n} M] = \\ &= [\mathbf{X} - \frac{\mathbf{1_n 1_n}^\mathsf{T} \mathbf{X}}{n}]^\mathsf{T} [\mathbf{X} - \frac{\mathbf{1_n 1_n}^\mathsf{T} \mathbf{X}}{n}] = \\ &= \mathbf{X}^\mathsf{T} [\mathbf{I} - \frac{\mathbf{1_n 1_n}^\mathsf{T}}{n}]^\mathsf{T} [\mathbf{I} - \frac{\mathbf{1_n 1_n}^\mathsf{T}}{n}] \mathbf{X} \end{aligned}$$

Posto ora $\mathbf{U} = \frac{\mathbf{1_n 1_n}^\mathsf{T}}{n}$, è facile vedere che $\mathbf{U}$ è idempotente e simmetrica di rango 1: è composta da $n \times n$ elementi tutti uguali a $\frac{1}{n}$; quindi sono idempotenti (ma di rango $n-1$) anche $\mathbf{I} - \mathbf{U}$, e $[\mathbf{I} - \mathbf{U}]^\mathsf{T} [\mathbf{I} - \mathbf{U}]$, per cui possiamo scrivere:

$$\mathbf{Q} = \sum_{i=1}^{n} (\mathbf{X}_i - M)^2 = \mathbf{X}^\mathsf{T} [\mathbf{I} - \mathbf{U}] \mathbf{X}$$

e $\mathbf{Q}$ è distribuita secondo una χ^2_{n-1}.

Esempio numerico;

Con $n = 5$ si supponga di avere le 5 osservazioni $x_i : 3, 5, 8, 9, 10$, con media aritmetica $M = 7$.

La somma dei quadrati degli scarti (osservati!) è data da:

$$\mathbf{Q} = \sum_{i=1}^{n} (x_i - M)^2 = 16 + 4 + 1 + 4 + 9 = 34.$$

E' facile vedere che la matrice $\mathbf{U}$ è data da:

$$\mathbf{U} = \begin{pmatrix} 0,2 & 0,2 & 0,2 & 0,2 & 0,2 \\ 0,2 & 0,2 & 0,2 & 0,2 & 0,2 \\ 0,2 & 0,2 & 0,2 & 0,2 & 0,2 \\ 0,2 & 0,2 & 0,2 & 0,2 & 0,2 \\ 0,2 & 0,2 & 0,2 & 0,2 & 0,2 \end{pmatrix}$$

Indicato quindi con x il vettore delle 5 osservazioni, si verifichi il risultato fornito dal prodotto $\mathbf{x}^\mathsf{T}[\mathbf{I}-\mathbf{U}]\mathbf{x}$:

$$\begin{aligned}\mathbf{x}^\mathsf{T}[\mathbf{I}-\mathbf{U}]\mathbf{x} &= (3\,5\,8\,9\,10)\begin{pmatrix}0,8 & -0,2 & -0,2 & -0,2 & -0,2\\ -0,2 & 0,8 & -0,2 & -0,2 & -0,2\\ -0,2 & -0,2 & 0,8 & -0,2 & -0,2\\ -0,2 & -0,2 & -0,2 & 0,8 & -0,2\\ -0,2 & -0,2 & -0,2 & -0,2 & 0,8\end{pmatrix}\begin{pmatrix}3\\5\\8\\9\\10\end{pmatrix} =\\ &= 3^2\times 0,8+5^2\times 0,8+\ldots+10^2\times 0,8-2\times 0,2\times 3\times 5-\ldots = 34\end{aligned}$$

2.8.3 La distribuzione dell'esponente della distribuzione normale multivariata.

Sappiamo già che il doppio dell'esponente della distribuzione normale univariata, $\frac{(\mathbf{X}-\mathrm{E}[\mathbf{X}])^2}{\mathrm{V}[\mathbf{X}]}$, si distribuisce secondo una variabile casuale χ^2. Vediamo come si generalizza questo risultato nel caso normale multivariato.

Sia $\mathbf{Y}$ un vettore di variabili casuali a p componenti, distribuito secondo una normale multivariata qualsiasi, ossia

$$\mathbf{Y} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

Si può dimostrare che la variabile casuale corrispondente alla forma quadratica che figura al numeratore dell'esponente della funzione di densità, ossia:

$$\mathbf{Q} = (\mathbf{Y}-\boldsymbol{\mu})^\mathsf{T}\,(\boldsymbol{\Sigma})^{-1}\,(\mathbf{Y}-\boldsymbol{\mu}),$$

segue una distribuzione chi-quadrato con p gradi di libertà.

Infatti il risultato si mostra facilmente ricorrendo ad una opportuna trasformazione lineare (già impiegata in questo capitolo)

$$\mathbf{X} = \mathbf{B}^\mathsf{T}[\mathbf{Y}-\boldsymbol{\mu}],$$

in cui $\mathbf{B}$ è tale che:

$$\mathbf{B}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{B} = \mathbf{I}, \qquad \text{e} \qquad (\boldsymbol{\Sigma})^{-1} = \mathbf{B}\mathbf{B}^\mathsf{T}.$$

---

verificare

--- e quindi:

$$\mathrm{V}\,[\mathbf{X}] = \mathbf{B}^\mathsf{T}\mathrm{V}\,[\mathbf{Y}-\boldsymbol{\mu}]\,\mathbf{B} = \mathbf{B}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{B} = \mathbf{I}$$

Pertanto:

$$\begin{aligned}\mathbf{Q} &= (\mathbf{Y}-\boldsymbol{\mu})^\mathsf{T}\,(\boldsymbol{\Sigma})^{-1}\,(\mathbf{Y}-\boldsymbol{\mu}) = (\mathbf{Y}-\boldsymbol{\mu})^\mathsf{T}\mathbf{B}\mathbf{B}^\mathsf{T}(\mathbf{Y}-\boldsymbol{\mu}) =\\ &= [(\mathbf{Y}-\boldsymbol{\mu})^\mathsf{T}\mathbf{B}][\mathbf{B}^\mathsf{T}(\mathbf{Y}-\boldsymbol{\mu})]\\ &= \mathbf{X}^\mathsf{T}\mathbf{X} \sim \chi^2_p\end{aligned}$$

Per cui $\mathbf{Q}$ si distribuisce come la somma dei quadrati di p variabili normali standardizzate e indipendenti, ossia come una chi-quadrato con p gradi di libertà.

In definitiva:

Distribuzione dell'esponente di una normale multivariata

se $\mathbf{Y} \sim (\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, allora

$$(\mathbf{Y} - \boldsymbol{\mu})^{\mathsf{T}} (\boldsymbol{\Sigma})^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \sim \chi^2_p$$

Esempio numerico

$$\mathbf{Y} \sim (N_2(\mathbf{0}, \boldsymbol{\Sigma}),$$

$$\text{con}\boldsymbol{\Sigma} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \qquad \text{e matrice di correlazione:} \qquad R = \begin{pmatrix} 1 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 1 \end{pmatrix}$$

e quindi

$$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix};$$

pertanto la forma quadratica:

$$\mathbf{Q} = \mathbf{Y}^{\mathsf{T}} \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \mathbf{Y}\mathbf{Y} = \mathbf{y}_1^2 + 2\mathbf{y}_2^2 - 2\mathbf{y}_1^2\mathbf{y}_2^2 \sim \chi^2_2$$

segue una distribuzione chi-quadrato con due gradi di libertà.

`eventualmente dimostrarlo per via`
`diretta nell'esempio`

2.8.4 Teorema di Cochran:

Supponiamo di avere una somma di quadrati di p variabili normali standardizzate e indipendenti, ossia:

$$\mathbf{Q} = \mathbf{X}^{\mathsf{T}}\mathbf{X}$$

o, più in generale, una forma quadratica:

$$\mathbf{Q} = \mathbf{X}^{\mathsf{T}}\mathbf{A}\mathbf{X},$$

con $\mathbf{A}$ idempotente di rango p . In questo caso il numero delle componenti di $\mathbf{X}$ potrà essere in generale maggiore di p; il punto essenziale è che $\mathbf{Q}$ abbia una distribuzione chi-quadrato con p gradi di libertà.

Supponiamo di saper scomporre algebricamente $\mathbf{Q}$ nella somma di k forme quadratiche:

$$\mathbf{Q} = \mathbf{X}^{\mathsf{T}}\mathbf{X} = \sum_{i=1}^{k} \mathbf{X}^{\mathsf{T}}\mathbf{A}_i\mathbf{X} = \sum_{i=1}^{k} \mathbf{Q}_i,$$

avendo posto: $\mathbf{Q}_i = \mathbf{X}^\mathsf{T}\mathbf{A}_i\mathbf{X}$, ed essendo per ipotesi:

$$\mathbf{Q} \sim \chi^2_p$$

Il teorema di Cochran stabilisce delle relazioni di importanza fondamentale in merito alle caratteristiche delle distribuzioni delle singole componenti $\mathbf{Q}_i$.

TEOREMA DI COCHRAN

Una qualsiasi delle seguenti tre condizioni implica le altre due:

1. la somma dei gradi di libertà delle forme quadratiche deve eguagliare p:

$$\sum_{i=1}^{k} \rho(\mathbf{A}_i) = p = \rho(\mathbf{A})$$

(in generale la somma dei ranghi delle singole componenti deve eguagliare il rango di $\mathbf{A}$)

2. tutte le k forme quadratiche $\mathbf{X}^\mathsf{T}\mathbf{A}_i\mathbf{X}$ hanno una distribuzione χ^2 che corrisponde a : tutte le $\mathbf{A}_i$ devono essere idempotenti;

3. tutte le k forme quadratiche $\mathbf{Q}_i = \mathbf{X}^\mathsf{T}\mathbf{A}_i\mathbf{X}$ sono a due a due indipendenti, che corrisponde a: $\mathbf{A}_i\mathbf{A}_j = 0$ per qualsiasi coppia $i \neq j$.

Ai fini pratici se per esempio se si vuole applicare ad una particolare scomposizione la proprietà 2, per poi dedurne la 1 e la 3, non è necessario esplicitare le singole matrici $\mathbf{A}_i$, ma è sufficiente sapere che si è scomposta $\mathbf{Q}$ in forme quadratiche nelle variabili aleatorie $\mathbf{X}_i$.

Importanza del teorema di Cochran

L'importanza di tale teorema nell'ambito della teoria normale sui modelli lineari è cruciale; in generale a ciascuna delle k componenti si farà corrispondere una particolare fonte di variabilità o un gruppo di parametri.

In effetti esiste una formulazione ancora più generale del teorema, che prende in considerazione distribuzioni χ^2 non centrali, ossia forme quadratiche in variabili normali con speranza matematica diversa da zero, utile per la generalizzazione alla distribuzione di determinate quantità test non solo sotto H_0 ma anche sotto H_1. Per non appesantire questi appunti non riporto questa generalizzazione.

Esempio. Come esempio si rifletta sulla nota scomposizione per la somma dei quadrati di n variabili normali standardizzate indipendenti:

$$\sum_{i=1}^{n} \mathbf{X}_i^2 = \sum_{i=1}^{n} (\mathbf{X}_i - M)^2 + nM^2$$

Per applicare il teorema di Cochran è sufficiente far vedere che i due addendi sulla destra sono forme quadratiche in variabili normali di rango $n-1$ e 1: è immediato verificarlo senza bisogno di esplicitare le matrici, perché $\sum_{i=1}^{n} (\mathbf{X}_i - M)^2$ è palesemente una forma quadratica con un vincolo lineare ($\sum_{i=1}^{n} (\mathbf{X}_i - M) = 0$), mentre M^2 ha ovviamente un solo grado di libertà, quindi i due termini sono indipendenti e distribuiti come delle v.c. χ^2 con i rispettivi gradi di libertà.

2.9 Stimatori di massima verosimiglianza dei parametri di una normale multivariata

In questo capitolo sulla normale multivariata non ci siamo mai preoccupati della stima dei parametri: ho solo esposto le proprietà formali di una distribuzione normale multivariata [3] di cui conosciamo tutti i parametri, ossia gli elementi di $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$. Non affronterò in modo completo il problema dell'inferenza sui parametri sulla base di campioni provenienti da una normale multivariata, ma mi limiterò in queste pagine al problema della stima puntuale dei parametri, che è ovviamente basilare.

Supponiamo di avere un campione (multivariato) casuale di ampiezza n estratto da una normale multivariata a p componenti, ossia una matrice $\mathbf{X}$ $(n \times p)$ di dati, le cui righe sono le n determinazioni di una variabile normale multipla:

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{i1} & x_{i2} & \dots & x_{ij} & \dots & x_{ip} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nj} & \dots & x_{np} \end{pmatrix}$$

In analogia al caso univariato, i momenti primi e secondi calcolati sul campione multivariato sono le stime di massima verosimiglianza dei corrispondenti parametri della distribuzione di provenienza; in sintesi:

[3] che ritengo di importanza fondamentale come premessa allo studio dei modelli lineari e dei modelli di dipendenza in genere. Comunque ribadisco che uno può studiare i modelli lineari a livello elementare introduttivo senza conoscere la normale multivariata, ma uno statistico non può...

Stimatore di massima verosimiglianza di μ

Lo stimatore di massima verosimiglianza del vettore delle speranze matematiche $\boldsymbol{\mu}$ di una variabile normale multipla è dato dal vettore $M(\mathbf{X})$ delle medie aritmetiche di un campione multivariato $\mathbf{x}$ di n osservazioni i.i.d. estratto dalla corrispondente distribuzione.
Tale stimatore, come nel caso univariato, è non distorto.

Stimatore di massima verosimiglianza di Σ

Lo stimatore di massima verosimiglianza della matrice di varianze e covarianze $\boldsymbol{\Sigma}$ di tale variabile è dato dalla matrice delle varianze e covarianze empiriche calcolata su un campione multivariato di n osservazioni i.i.d. estratto dalla corrispondente distribuzione.

$$\hat{\boldsymbol{\Sigma}} = \mathrm{V}[\mathbf{X}] = \frac{\mathbf{Z}^{\mathsf{T}}\mathbf{Z}}{n}$$

(essendo $\mathbf{Z}$ la matrice degli scarti)
Tale stimatore, come nel caso univariato, è invece distorto.
E' possibile costruire uno stimatore corretto (o non distorto) moltiplicando sia le varianze che le covarianze empiriche per il fattore correttivo $\frac{n}{n-1}$, ottenendo quindi lo stimatore:

$$\mathbf{S} = \mathrm{V}[\mathbf{X}]\frac{n}{n-1} = \frac{\mathbf{Z}^{\mathsf{T}}\mathbf{Z}}{n-1}$$

2.9.1 Dimostrazione

Avvertenza

È opportuno a questo punto richiamare e rivedere le proprietà viste precedentemente sulla derivazione di forme quadratiche di determinanti e di matrici inverse. link con proprietà matrici

Stimatori di Massima Verosimiglianza nella normale Multivariata: Dimostrazione

Per ricavare gli stimatori di massima verosimiglianza dei parametri di una normale multivariata costruiamo come sempre la verosimiglianza, o meglio il suo logaritmo, supponendo di avere n osservazioni indipendenti ciascuna con p componenti.
Per comodità e perché questo facilita i passaggi successivi, consideriamo co-

me parametri gli elementi c_{ij} di $\mathbf{C}$, l'inversa della matrice $\mathbf{\Sigma}$ di varianze e covarianze, oltre ovviamente al vettore delle speranze matematiche $\boldsymbol{\mu}$.
Sappiamo dalle proprietà degli stimatori di massima verosimiglianza che la parametrizzazione è irrilevante ai fini della determinazione degli stimatori puntuali.
Costruiamo la quantità:

$$-2\log L(\boldsymbol{\mu};\mathbf{C})$$

(essendo $L(\boldsymbol{\mu},\mathbf{C})$ la verosimiglianza campionaria (rispetto a $\boldsymbol{\mu}$ e $\mathbf{C}$), sulla base di un campione di n osservazioni indipendenti (si riveda la parte iniziale sulla normale multivariata, per questa parametrizzazione, in particolare l'equazione 2.5):

$$l(\boldsymbol{\mu},\mathbf{C};\mathbf{X}) = -2logL(\boldsymbol{\mu},\mathbf{C};\mathbf{X}) = k - nlog|\mathbf{C}| + \sum_{i=1}^{n}(\mathbf{x}_i-\boldsymbol{\mu})^{\mathsf{T}}\mathbf{C}(\mathbf{x}_i-\boldsymbol{\mu})$$

essendo $\mathbf{x}_i$ il vettore colonna osservato (a p componenti) relativo all'i-esima osservazione. Procedendo a derivare prima rispetto al vettore $\boldsymbol{\mu}$ si ha:

$$\frac{\partial l(\boldsymbol{\mu},\mathbf{C};\mathbf{X})}{\partial\boldsymbol{\mu}} = -2\sum_{i=1}^{n}\mathbf{C}(\mathbf{x}_i-\boldsymbol{\mu}) = -2\mathbf{C}\sum_{i=1}^{n}(\mathbf{x}_i-\boldsymbol{\mu})$$

E' immediato vedere che $\frac{\partial l(\boldsymbol{\mu},\mathbf{C};\mathbf{X})}{\partial\boldsymbol{\mu}}$ si annulla se:

$$2\mathbf{C}\sum_{i=1}^{n}(\mathbf{x}_i-\hat{\mu}) = 0$$

ossia (dato che $\mathbf{C}$ è di rango pieno!) solo quando:

$$\sum_{i=1}^{n}(\mathbf{x}_i-\hat{\mu}) = 0,$$

ed infine:

$$\sum_{i=1}^{n}\mathbf{x}_i = n\hat{\mu} \quad \text{e} \quad \hat{\mu} = \sum_{i=1}^{n}\mathbf{x}_i/n = M(\mathbf{X}) = \begin{pmatrix} M_1 \\ \dots \\ M_j \\ \dots \\ M_p \end{pmatrix}$$

Per quanto riguarda invece le derivate rispetto agli elementi di $\mathbf{C}$ conviene distinguere gli elementi diagonali c_{jj} da quelli fuori dalla diagonale $c_{jk}(k \neq j)$:

$$\begin{aligned} \frac{\partial l(\boldsymbol{\mu},\mathbf{C};\mathbf{X})}{\partial c_{jj}} &= \\ &= -\frac{n}{|\mathbf{C}|}\frac{\partial|\mathbf{C}|}{\partial c_{jj}} + \frac{\partial\sum_{i=1}^{n}(\mathbf{x}_i-\boldsymbol{\mu})^{\mathsf{T}}\mathbf{C}(\mathbf{x}_i-\boldsymbol{\mu})}{\partial c_{jj}} \qquad j=1,2,\dots,p \end{aligned}$$

Per il primo addendo a secondo membro ricordiamo il risultato generale per i determinanti di matrici simmetriche:

$$\frac{\partial|\mathbf{C}|}{\partial c_{jj}} = \mathbf{C}_{jj}$$

essendo $\mathbf{C}_{rs}$il cofattore di c_{rs} in $\mathbf{C}$, ____-
verificare

____-

mentre per il secondo addendo ovviamente si tratta di termini lineari in $\mathbf{C}$, per cui basterà nella sommatoria selezionare solo le componenti opportune dei vettori $(\mathbf{x}_i - \boldsymbol{\mu})$, ossia solo quelle che moltiplicano c_{jj} :

$$\frac{\partial l(\boldsymbol{\mu}, \mathbf{C}; \mathbf{X})}{\partial c_{jj}} = -n\frac{\mathbf{C}_{jj}}{|\mathbf{C}|} + \sum_{i=1}^{n}(x_{ij} - \mu_j)^2.$$

Si vede subito che:

$$\frac{\mathbf{C}_{jj}}{|\mathbf{C}|} = \sigma_j^2$$

dal momento che $\mathbf{C} = \boldsymbol{\Sigma}^{-1}$ e quindi $\boldsymbol{\Sigma} = \mathbf{C}^{-1}$ e gli elementi di un'inversa sono proprio dati dai rapporti fra cofattori e determinante.
Per trovare le espressioni degli stimatori $\hat{\sigma}_j^2$ occorre annullare le precedenti derivate, avendo sostituito alle speranza matematiche μ_j gli stimatori di massima verosimiglianza M_j. Pertanto:

$$\frac{\partial l(\boldsymbol{\mu}, \mathbf{C}; \mathbf{X})}{\partial c_{jj}} = 0 \to -n\frac{\mathbf{C}_{jj}}{|\mathbf{C}|} + \sum_{i=1}^{n}(x_{ij} - M_j)^2 = 0;$$

e quindi:

$$\hat{\sigma}_J^2 = \frac{\sum_{i=1}^{n}(x_{ij} - M_j)^2}{n}$$

Deriviamo adesso rispetto agli elementi non diagonali $c_{jk}(k \neq j\ )$:

$$\begin{aligned}\frac{\partial l(\boldsymbol{\mu}\mathbf{C}; \mathbf{X})}{\partial c_{jk}} &= \\ &= \ -\frac{n}{|\mathbf{C}|}\frac{\partial|\mathbf{C}|}{\partial c_{jk}} + \frac{\partial \sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})^{\mathsf{T}}\mathbf{C}(\mathbf{x}_i - \boldsymbol{\mu})}{\partial c_{jk}} \qquad j, k = 1, 2, \ldots, p; \quad k \neq j\end{aligned}$$

Procediamo come prima, e per il primo addendo a secondo membro ricordiamo il risultato generale per i determinanti di matrici simmetriche:

$$\frac{\partial|\mathbf{C}|}{\partial c_{jk}} = 2C_{jk}$$

cofattore di c_{jk} in $\mathbf{C}$, $k \neq j$.
Mentre per il secondo addendo selezioniamo le componenti dei vettori $(\mathbf{x}_i - \boldsymbol{\mu})$ che moltiplicano c_{jk}:

$$\frac{\partial l(\boldsymbol{\mu}, \mathbf{C}; \mathbf{X})}{\partial c_{jk}} = -2n\frac{\mathbf{C}_{jk}}{|\mathbf{C}|} + 2\sum_{i=1}^{n}(x_{ij} - \mu_j)(x_{ik} - \mu_k)$$

Ancora si ha:

$$\frac{\mathbf{C}_{jk}}{|\mathbf{C}|} = \sigma_{jk}$$

e per trovare le espressioni degli stimatori $\hat{\sigma}_{jk}$ occorre annullare le precedenti derivate, avendo sostituito alle speranze matematiche μ_j gli stimatori di massima verosimiglianza M_j. Pertanto:

$$\frac{\partial l(\boldsymbol{\mu}, \mathbf{C}; \mathbf{X})}{\partial c_{jk}} = 0 \Rightarrow -2n\frac{\mathbf{C}_{jk}}{|\mathbf{C}|} + 2\sum_{i=1}^{n}(x_{ij} - M_j)(x_{ik} - M_k) = 0;$$

e quindi:

$$\hat{\sigma}_{jk} = \frac{\sum_{i=1}^{n}(x_{ij} - M_j)(x_{ik} - M_k)}{n}$$

e quindi in definitiva il risultato prima anticipato:

$$\hat{\boldsymbol{\Sigma}} = \mathrm{V}[\mathbf{X}] = \mathrm{V}[\mathbf{Z}] = \mathbf{Z}^{\mathsf{T}}\mathbf{Z}/n$$

essendo $\mathbf{X}$ il campione multivariato originario e $\mathbf{Z}$ la matrice degli scarti

In effetti, dal momento che gli unici parametri della distribuzione normale multivariata sono il vettore delle medie e la matrice di varianza e covarianza, per ottenere gli stimatori di massima verosimiglianza (e quindi stimatori puntuali!) di tutte le quantità necessarie per calcolare le distribuzioni congiunte, marginali, condizionate e per le componenti principali da un campione proveniente da una normale multivariata, si impiegheranno le stesse formule già viste per la distribuzione teorica, sostituendo ai momenti primi e secondi teorici quelli empirici stimati dal campione, dal momento che lo stimatore di massima verosimiglianza di una funzione dei parametri $\mathbf{g}(\boldsymbol{\theta})$ è dato dalla stessa funzione dello stimatore di Massima verosimiglianza, $\mathbf{g}(\hat{\boldsymbol{\theta}})$.

Per esempio, lo stimatore di massima verosimiglianza degli autovalori della matrice di varianze e covarianze $\boldsymbol{\Sigma}$, ossia del vettore delle varianze delle componenti principali, si ottiene ovviamente dagli autovalori dello stimatore di massima verosimiglianza $\hat{\boldsymbol{\Sigma}}$:

$$\hat{\lambda}(\boldsymbol{\Sigma}) = \lambda\left(\hat{\boldsymbol{\Sigma}}\right)$$

Inferenza nel caso normale sugli autovalori

Sebbene solitamente si impieghino le tecniche di analisi delle componenti principali a scopo esplorativo, è interessante accennare al caso in cui si abbia a disposizione un campione multivariato estratto da una distribuzione normale; abbiamo infatti già visto come per una distribuzione normale multivariata gli autovalori e gli autovettori assumano dei significati ben precisi.

Evidentemente gli stimatori di massima verosimiglianza degli autovalori e degli autovettori sono forniti dagli autovalori e dagli autovettori della matrice di varianze e covarianze campionaria (che è lo stimatore di massima verosimiglianza della matrice di varianze e covarianze teorica); dal momento che per gli stimatori delle varianze e delle covarianze per campioni provenienti da una normale valgono la proprietà di regolarità ei teoremi che forniscono le distribuzioni campionarie e che garantiscono la consistenza degli stimatori insieme con loro correttezza asintotica, dobbiamo aspettarci che anche per gli autovalori e gli autovettori ricavati da tali matrici campionarie valgano delle proprietà di consistenza e di correttezza asintotica.

In effetti qui mi limito a riportare un risultato asintotico che riguarda la distribuzione degli autovalori per campioni provenienti da una distribuzione normale multivariata, utile tuttavia per avere un'idea dell'ordine di grandezza della variabilità dovuta al campionamento.

Asintoticamente gli l_j, stime campionarie dei veri autovalori λ_j, ottenute da un campione di n osservazioni estratto da una normale multivariata, si distribuiscono secondo una normale multivariata a componenti indipendenti, con valore atteso $\mathrm{E}\left[l_j\right] = \lambda_j$ e varianza campionaria : $\mathrm{V}\left[l_j\right] = \frac{2\lambda_j^2}{n-1}$

(si ricordi il caso particolare di matrici di varianze e covarianze diagonali: questi risultati coincidono con quelli classici della distribuzione di una varianza campionaria!)

In sintesi sia ha:

$$n \to \infty : \qquad \mathbf{l} \sim \mathcal{N}_p\left(\boldsymbol{\lambda}, \frac{2\boldsymbol{\lambda}^2}{n-1}\right)$$

Casi interessanti:

$$H_0 : \lambda_j = 1, j = 1, 2, \ldots, p$$

che corrisponde al caso di indipendenza fra le variabili (standardizzate).

È da intendersi che questi risultati sono semplicemente delle approssimazioni ma danno delle indicazioni sull'ordine di grandezza dell'errore campionario.

2.10 Un test di Multinormalità: cenni

Quando si ha a disposizione un campione di dati multivariato, molto spesso è necessario verificare se è plausibile l'ipotesi di provenienza da un universo normale multivariato.

Un modo semplice per verificare la normalità di un campione di osservazioni multivariate, consiste ovviamente nell'effettuare dei test di normalità su ciascuna delle distribuzioni univariate.

Ricordo che la normalità delle distribuzioni marginali è una condizione necessaria ma non sufficiente per la normalità multivariata: pertanto i test sulla normalità delle distribuzioni marginali costituiscono uno sbarramento preliminare, nel senso che se danno esito negativo possiamo senz'altro scartare l'ipotesi di multinormalità, altrimenti occorrerà procedere col saggiare l'ipotesi di normalità multivariata con test basati sulla distribuzione congiunta.

Se l'insieme in esame è costituito da molte variabili non sarà possibile utilizzare i normali test di bontà dell'adattamento; tuttavia è possibile ottenere delle informazioni eventualmente anche grafiche trasformando opportunamente l'insieme di dati multivariato.

Come si è visto prima, la forma quadratica ad esponente della densità normale ha una distribuzione proporzionale a quella di una χ^2 con p gradi di libertà.

Infatti se:

$$\mathbf{Y} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

si è già visto prima che la variabile casuale

$$\mathbf{Q} = (\mathbf{Y} - \boldsymbol{\mu})^\mathsf{T} \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \sim \chi_p^2$$

Pertanto se trasformiamo ognuno degli n vettori osservati $\mathbf{x}_i$ a p componenti secondo la stessa relazione, dovremo aspettarci che questi n valori trasformati q_i seguano ciascuno una distribuzione χ^2 con p gradi di libertà:

$$Q_i = (\mathbf{x}_i - \boldsymbol{\mu})^\mathsf{T} ({}^{-1}\mathbf{x}_i - \boldsymbol{\mu}) \boldsymbol{\Sigma} \sim \chi_p^2$$

(le n trasformate Q_i risultano indipendenti per l'indipendenza ipotizzata degli n vettori osservati $\mathbf{x}_i$)

Quindi, se è valida l' ipotesi di multinormalità, il vettore delle n trasformate q_i costituisce un campione casuale semplice estratto da una distribuzione χ^2 con p gradi di libertà. In effetti le quantità che si usano effettivamente per il calcolo delle q_i sono gli stimatori di $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$, M e $\mathbf{S}$, e non i parametri veri (usualmente incogniti); questo fa sì che le quantità:

$$\hat{q}_i = (\mathbf{x}_i - M)^\mathsf{T} \mathbf{S}^{-1} (\mathbf{x}_i - M)$$

seguano una distribuzione χ_p^2 solo approssimativamente; l'approssimazione è soddisfacente per campioni grandi.

In effetti, un'informazione utile si ricava dalla rappresentazione grafica di tali valori trasformati in corrispondenza dei percentili teorici di una variabile χ^2; un altro elemento di cui si potrebbe tenere conto nella costruzione di un test di normalità è dato dagli angoli che i vettori osservati formano con il centroide M del campione; tuttavia adesso per semplicità non vedremo quest'ulteriore possibilità.

Questa trasformazione ci darà fra l'altro la possibilità di creare (sebbene in modo non univoco) un ordinamento in un insieme di dati multivariati, cosa molto utile quando nell'analisi esplorativa occorre trovare *outliers* ossia dati lontani rispetto alla massa dei dati, facendo uso della cosiddetta distanza di MahalaNobis

`esempio da fare con file MahalaNobis Rmd`

Capitolo 3

Alcune distribuzioni multivariate non normali

L'estensione al caso multivariato di distribuzioni non-normali a componenti non indipendenti è sempre ardua, perché le possibilità di estensione di sistemi di curve univariate non normali al caso multivariato possono essere di diversa natura, mentre dalla distribuzione normale univariata si può arrivare alla sua estensione multivariata con diverse impostazioni giungendo sempre alla stessa forma multivariata; ad esempio:

- dalla densità o dalla funzione caratteristica, sostituendo ad un quadrato una forma quadratica;
- se $\mathbf{x}^{\mathsf{T}}\mathbf{a}$ è normale per qualsiasi $\mathbf{a}$, allora x è normale multivariato.
- come distribuzione di $\mathbf{x} = \boldsymbol{\mu} + \mathbf{A}\mathbf{y}$(con $\mathbf{y}$ a componenti indipendenti)
- da distribuzioni condizionate normali e omoscedastiche con funzioni di regressione lineari.

3.1 Una particolare distribuzione beta multivariata (distribuzione di Dirichlet)

La distribuzione di Dirichlet a k componenti, che costituisce una particolare generalizzazione multivariata della distribuzione Beta, è definita come segue:

- si considerino $k+1$ v.a. indipendenti $\mathbf{X}_i (i = 0, 1, 2, \ldots, k)$, ciascuna con distribuzione Gamma con lo stesso parametro di scala λ e di parametri di forma c_i ;
- indicata con S la loro somma, $S = \sum_{i=0}^{k} \mathbf{X}_i$, la distribuzione di Dirichlet è la distribuzione congiunta delle k nuove variabili definite dalle relazioni:

$$\mathbf{y}_i = \mathbf{X}_i / S i = 1, 2, \ldots, k.$$

La densità di tale distribuzione è data da:

$$f(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_k) = \prod_{i=1}^{k} \mathbf{y}_i^{ci-1}[1 - \sum_{i=1}^{k} \mathbf{y}_i]^{c0-1} \mathbf{\Gamma}(\sum_{i=0}^{k} c_i) / \prod_{i=0}^{k} \mathbf{\Gamma}(c_i),$$

ed è definita sul simplesso:

$$\mathbf{y}_i (0, i = 1, 2, \ldots, k; \sum_{i=1}^{k} \mathbf{y}_i \leq 1.$$

Questa distribuzione è importante ad esempio per la descrizione della distribuzione simultanea di rapporti di composizione; si vedano nelle figure che seguono, per il caso bivariato, alcuni esempi di densità per diverse combinazioni dei parametri c_0, c_1 e c_2 (indicati nel seguito con a, b, c nel caso bivariato).

Se $c_i (1 (i = 0, 1, \ldots, k)$, la densità ha sempre un massimo unico in corrispondenza di:

$$\mathbf{y}_i^* = (c_i - 1) / \sum_{i=0}^{k} (c_i - 1), (i = 1, 2, \ldots, k).$$

- Tutte le distribuzioni marginali univariate sono delle distribuzioni Beta.
- Le distribuzioni condizionate sono ancora delle Beta
- Nella distribuzione bivariata (indicando le due componenti con **X** ,**y** , e i parametri con a,b,c)) la distribuzione di **y** condizionata a **X** =x è proporzionale ad una variabile con distribuzione Beta univariata. In particolare si dimostra che:
- $\mathbf{y}/(1 - x) \| \mathbf{X} = x$ si distribuisce come una $Beta[b, c]$
- per cui $\mathrm{E}[\mathbf{y}]$ varia linearmente con x, ma anche $\mathrm{V}[\mathbf{y}]$ varia con x

Esempi di densità di distribuzioni di Dirichlet:

$$\begin{array}{ccc} c_0 = 1,2 & c_1 = 0,9 & c_2 = 0,9 \\ c_0 = 1,2 & c_1 = 1,3 & c_2 = 1,8 \\ c_0 = 3 & c_1 = 4 & c_2 = 5. \end{array}$$

3.2 Altri esempi di distribuzioni multivariate non normali

Distribuzione Logistica Doppia di densità:

$$F(x, \mathbf{y}) = 1/(1 + Exp[-x] + Exp[-\mathbf{y}])$$

3.3 Costruzione di variabili correlate

Qualche volta può essere utile costruire in modo semplice delle variabili correlate a partire da variabili indipendenti. Il metodo più semplice consiste nell'aggiungere a tutte le variabili aleatorie un'ulteriore variabile aleatoria, secondo quanto specificato di seguito.

costruzione di variabili correlate

Uno schema generale di costruzione di variabili aleatorie correlate da p+1 variabili aleatorie indipendenti $\mathbf{X}_j$ (j=0,1, ...,p), è quello di considerare p variabili aleatorie sommando a tutte la componente $\mathbf{X}_0$. In dettaglio otteniamo ora un nuovo vettore aleatorio $\mathbf{Y}$ a p componenti, ponendo:

$$\begin{pmatrix} \mathbf{y}_1 = \mathbf{X}_0 + \mathbf{X}_1 \\ \dots \\ \mathbf{y}_j = \mathbf{X}_0 + \mathbf{X}_j \\ \dots \\ \mathbf{y}_p = \mathbf{X}_0 + \mathbf{X}_p \end{pmatrix}$$

In pratica la componente $\mathbf{X}_0$ è quella che determina la covarianza fra le componenti di $\mathbf{Y}$.
E' facile calcolare i momenti di $\mathbf{Y}$ da quelli di $\mathbf{X}$, mentre può essere in generale arduo calcolare la distribuzione di $\mathbf{Y}$ (è spesso è complicato integrare rispetto a $\mathbf{X}_0$ nella densità congiunta di $\mathbf{X}_0$, $\mathbf{X}_1$, ..., $\mathbf{X}_p$,).
Come esercizio si calcoli la correlazione e la covarianza fra due generiche componenti di $\mathbf{Y}$ o, direttamente, la matrice di varianze e covarianze e la matrice di correlazione di $\mathbf{Y}$.

$$V(\mathbf{y}_j) = V(\mathbf{X}_0) + V(\mathbf{X}_j); Cov(\mathbf{y}_j, \mathbf{y}_k) = V(\mathbf{X}_0)$$

Capitolo 4

link esterni e argomenti mancanti

```
\label{sec:corrparz}
\label{sec:acp}
\label{sec:modlin}
\label{sec:normalemulti}
\label{eq:distcondnormbiv1}
\esempiormd{esempiocor}
\esempiormd{esempioBMP1}
\esempiormd{MahalaNobis}
\esempiormd{trivariatenormal}
\esempiormd(inversa)
\esempiormd{codice R}
\esempiormd{PCA}
statistica3_2020provetrinorm1.R

\href{./examples/stat3_firstslides2.html}{un esempio}
\href{./examples/MLA2019normale.pdf}{La normale multivariata}
```

forse bisogna uniformare la notazione sui cofattori

cenno a risultati sull'inferenza multivariata? Wilks, etc.

dire qualcosa su unicitÃ , suff. congiunta, fattorizzazione verosimiglianza, efficienza, etc.???

verificare se la matrice Ã¨ unica nella diagonalizzazione

```
nella dirichlet

\begin{fig}
\begin{fig}

in bivar1.nb

\end{fig}


FIG2000REGR_ETERO1.STG

\end{fig}

Esempi di densitÃ  di distribuzioni di Dirichlet:


\section{La distribuzione multinomiale}\label{sec:multinomiale}

\begin{fig}
in bivar1.nb
\end{fig}

Distribuzione Esponenziale Bivariata  $(a=0,7)$
$$
F(\vecx,\vecy)=
(1-Exp[-\vecx])(1-Exp[-\vecy])( (1+a Exp[-\vecx-\vecy])
$$

\begin{fig}
in bivar1.nb
\end{fig}

Distribuzione Bivariata Dirichlet
 (\mBeta-bivariata)  $a=1,5;
\vecb=1,6; c=2,1$
\begin{fig}
in bivar1.nb
\end{fig}

Distribuzione Bivariata Dirichlet
```

```
 (\mBeta-bivariata)  $a=4;
\vecb=4; c=3$

\begin{fig}

in bivar1.nb

\end{fig}
Distribuzione Bivariata Dirichlet
(\mBeta-bivariata)  $a=1,1;
\vecb=1,1; c=0,9$
\begin{fig}
in bivar1.nb
\end{fig}
```