

# Appunti di calcolo matriciale per statistici (solo un richiamo!!)

Marcello Chiodi

9 marzo 2020

*Avvertenza: in questa bozza figure e link esterni non attivi*

## Indice

<b>1</b>	<b>Richiami di algebra matriciale</b>	<b>3</b>
1.1	Introduzione	3
1.2	Rango di una matrice	4
1.2.1	Rango della matrice di varianze e covarianze di un vettore aleatorio:	5
1.3	Traccia di una matrice	6
1.3.1	Varianza generalizzata	6
1.4	Matrice inversa	7
1.4.1	Significato degli elementi dell'inversa di una matrice di varianze e covarianze	8
1.4.2	Inversa di una matrice simmetrica partizionata	8
1.4.3	Determinante e inversa di una matrice simmetrica orlata	9
1.4.4	Matrice inversa generalizzata	10
1.5	Matrici ortogonali	11
1.6	Matrici idempotenti	11
1.6.1	Esempio	12
1.6.2	Esempi di matrici idempotenti di rango 2	12
1.7	Forme quadratiche	13
1.7.1	Significato geometrico delle forme quadratiche	13
<b>2</b>	<b>Calcolo differenziale con vettori e matrici</b>	<b>13</b>
2.1	Gradiente di una funzione	13
2.2	Hessiano di una funzione	13
2.3	Derivate di forme lineari e quadratiche	14
2.4	Derivate di inverse e di determinanti	15
<b>3</b>	<b>Autovalori e autovettori</b>	<b>15</b>
3.1	Definizione di autovettore e autovalore	16
3.2	Proprietà generali degli autovalori	17
3.3	Autovalori e autovettori di matrici simmetriche	17
3.3.1	Diagonalizzazione di una matrice simmetrica	18
3.3.2	Decomposizione spettrale di una matrice simmetrica	19
3.3.3	Autovalori di inverse e di potenze	19
3.3.4	Autovalori di una forma quadratica definita positiva	20

<b>4</b>	<b>Vettori aleatori</b>	<b>21</b>
4.1	Momenti primo e secondo multivariati di vettori aleatori . . .	21
4.2	Momenti di una trasformata lineare di un vettore aleatorio . .	22
4.2.1	Costruzione di variabili correlate . . . . .	24
4.2.2	Funzione caratteristica di una trasformata lineare di un vettore aleatorio . . . . .	24
4.2.3	Momento primo di forme quadratiche . . . . .	25
4.2.4	Momenti di funzioni qualsiasi di vettori aleatori. . . . .	25
4.2.5	Esempio sulla distribuzione F . . . . .	26
4.2.6	Esempio sulla distribuzione Beta . . . . .	27
4.2.7	Esempio su variabili osservate . . . . .	28
<b>5</b>	<b>Analisi delle componenti principali (ACP), solo cenni</b>	<b>30</b>
5.1	Significato statistico e probabilistico delle componenti principali	33
<b>6</b>	<b>Variabili Statistiche Multiple</b>	<b>34</b>
6.1	Calcoli statistici in notazione vettoriale . . . . .	34
6.2	Espressione della varianza di una variabile statistica . . . . .	34
6.2.1	Espressione vettoriale della covarianza . . . . .	35
6.2.2	Espressione vettoriale della media aritmetica . . . . .	37
<b>7</b>	<b>Definizione della matrice dei dati</b>	<b>38</b>
7.1	Dati mancanti . . . . .	39
<b>8</b>	<b>I momenti primi e secondi multivariati</b>	<b>40</b>
8.1	La matrice di varianze e covarianze . . . . .	42
8.2	La matrice di correlazione . . . . .	42
8.3	esempio . . . . .	43
<b>9</b>	<b>La matrice degli scarti</b>	<b>45</b>
9.1	I momenti di combinazioni lineari di variabili statistiche multiple	47
9.2	Rango della matrice di varianza e covarianza ( $n \geq p$ ) . . . . .	49

# 1 Richiami di algebra elementare delle matrici per lo studio dei modelli statistici lineari

## 1.1 Introduzione

E' un richiamo delle nozioni e degli strumenti tecnici necessari per una trattazione agevole degli argomenti che coinvolgono, in varia misura, vettori di variabili casuali e vettori di variabili casuali normali. La strumentazione di calcolo vettoriale e matriciale è anche essenziale per lo studio delle variabili statistiche multiple e per lo studio dei modelli lineari.

Sebbene questi appunti siano stati concepiti come supporto ad alcuni dei miei corsi per studenti di area statistica, possono essere un breviario utile per corsi di analisi statistica multivariata e analisi dei modelli lineari di dipendenza.

- Ove possibile, viene enfatizzato il significato statistico e/o probabilistico delle proprietà delle matrici richiamate.
- In particolare verranno evidenziate alcune proprietà di matrici di varianze e covarianze, sia per vettori di variabili aleatorie che variabili statistiche multiple (rilevate attraverso una matrice di dati).
- Sebbene le proprietà del calcolo matriciale siano comunque importanti, ho evidenziato qui solo quelle che nel resto del corso vengono utilizzate.
- Presuppongo che lo studente che legge questa sezione abbia le necessarie nozioni di algebra lineare (*ed eventualmente questo è il momento buono per aggiornare o integrare le proprie nozioni*).
- Queste nozioni sono essenziali per lo studio dei modelli lineari: tale studio risulterà in questo modo molto scorrevole e di semplice comprensione (spero!).
- Lo studio delle proprietà delle matrici e dei vettori di variabili casuali è anche finalizzato allo studio delle variabili aleatorie con distribuzione normale multivariata, modello parametrico multivariato importante per uno studio approfondito dei modelli lineari.

In molti esempi e casi di studio esposti in queste pagine, si ha a che fare in vario modo con problemi che coinvolgono  $p$  variabili rilevate su  $n$  unità (in generale tratteremo nel corso di variabili quantitative e/o qualitative, ma in questo richiamo su matrici e vettori aleatori ci riferiamo solo a variabili quantitative; eventualmente qualche variabile può essere costituita solo da 0 e 1).

Accade spesso che di queste variabili una sia oggetto di interesse e che se ne voglia studiare la dipendenza dalle altre; in altre situazioni magari vogliamo studiare il comportamento simultaneo delle variabili.

Talora le  $n$  osservazioni sono da considerarsi come un campione casuale semplice da una qualche distribuzione multivariata, oppure come determinazioni di variabili che contengono delle componenti aleatorie (come per esempio nei modelli lineari)

In ogni caso non v'è dubbio che è utile definire (o ricordare) alcuni concetti relativi alle distribuzioni di vettori aleatori, per generalizzare la definizione di momento già nota nel caso univariato, almeno per il momento primo e

secondo; sebbene i risultati che vedremo abbiano validità generale, uno degli scopi sarà quello di impadronirci degli strumenti tecnici necessari per lo studio dei modelli lineari e della distribuzione normale multivariata e per apprezzarne l'importanza come modello di base per i modelli di dipendenza e di regressione lineare semplice e multipla.

Un altro motivo dell'importanza degli strumenti di questo capitolo è la familiarizzazione con il linguaggio dei vettori e delle matrici, che consente in molti problemi multivariati di adottare una notazione compatta, semplice e del tutto analoga a quella univariata, come si vede anche nel capitolo sulle matrici di dati e la notazione matriciale per i calcoli statistici. In effetti gli strumenti tecnici di questo capitolo sono necessari per lo studio dei seguenti argomenti:

- combinazioni (lineari) di variabili casuali
- distribuzione normale multivariata
- forme quadratiche in variabili casuali normali
- inferenza nei modelli statistici lineari
- regressione multipla
- GLM (generalized linear models)
- Modelli di dipendenza non parametrica
- analisi componenti principali per vettori aleatori
- analisi serie temporali
- analisi esplorativa dei dati
- tecniche multivariate

un esempio

La normale multivariata

## 1.2 Rango di una matrice

Il rango di una matrice  $\mathbf{A}$  qualsiasi,  $\rho(\mathbf{A})$ , è definito come:

**il massimo numero di righe (o colonne) linearmente indipendenti**  
oppure:

il massimo ordine per il quale si possono estrarre minori non tutti nulli da una matrice qualsiasi  $\mathbf{A}$ .

Alcune proprietà del rango di una matrice:

$$\begin{aligned}\rho(\mathbf{A}^T) &= \rho(\mathbf{A}) \\ \rho(\mathbf{A}^T \mathbf{A}) &= \rho(\mathbf{A} \mathbf{A}^T) = \rho(\mathbf{A}) \\ \rho(\mathbf{A} \mathbf{B}) &\leq \min \{ \rho(\mathbf{A}), \rho(\mathbf{B}) \} \\ \rho(\mathbf{A} + \mathbf{B}) &\leq \rho(\mathbf{A}) + \rho(\mathbf{B})\end{aligned}$$

### 1.2.1 Rango della matrice di varianze e covarianze di un vettore aleatorio:

- Se in un vettore aleatorio composto da  $p$  v.a. elementari, una componente è combinazione lineare delle altre, allora il rango della matrice di varianza e covarianza di  $\mathbf{X}$  risulta uguale a  $p - 1$ ;
- in generale il rango di  $V[\mathbf{X}]$  risulta uguale a  $p - k$  se  $k$  componenti sono ottenute attraverso combinazioni lineari (indipendenti) degli elementi di  $\mathbf{X}$ .
- Il rango di  $V[\mathbf{X}]$  risulta uguale esattamente a  $p$  (ossia a rango pieno) se e solo se le componenti di  $\mathbf{X}$  sono linearmente indipendenti.
- nel caso di matrici di dati occorrerà specificare che  $n \geq p$ . Se invece  $n < p$ , ossia le variabili sono più delle unità, il rango sarà senz'altro inferiore a  $p$  e quindi esisteranno senz'altro dei vincoli lineari fra le variabili

**Esempio 1.1** *Ad esempio sia  $\mathbf{X}$  una variabile aleatoria doppia, con componenti  $X_1$  e  $X_2$  con speranze matematiche nulle e matrice di varianze e covarianze (per ipotesi di rango 2):*

$$V[\mathbf{X}] = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$$

essendo  $\mathbf{X} = (X_1, X_2)^T$ . Se otteniamo ora un nuovo vettore aleatorio  $\mathbf{Y}$  a tre componenti, con:

$$\begin{aligned} \mathbf{y}_1 &= X_1 \\ \mathbf{y}_2 &= X_2 \\ \mathbf{y}_3 &= 2X_1 + 3X_2, \end{aligned}$$

abbiamo utilizzato in pratica una matrice di trasformazione:

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 3 \end{pmatrix}$$

così che  $\mathbf{Y} = \mathbf{A} \mathbf{X}$  corrisponde alla trasformazione prima definita.

Per ottenere la matrice di varianze e covarianze di  $\mathbf{Y}$  dovremo utilizzare la regola:

$$V[\mathbf{Y}] = \mathbf{A} V[\mathbf{X}] \mathbf{A}^T,$$

ottenendo:

$$V[\mathbf{Y}] = \begin{matrix} & \begin{matrix} c_1 & c_2 & c_3 \end{matrix} \\ \begin{matrix} r_1 : \\ r_2 : \\ r_3 : \end{matrix} & \begin{pmatrix} \sigma_1^2 & \sigma_{12} & 2\sigma_1^2 + 3\sigma_{12} \\ \sigma_{12} & \sigma_2^2 & 2\sigma_{12} + 3\sigma_2^2 \\ 2\sigma_1^2 + 3\sigma_{12} & 2\sigma_{12} + 3\sigma_2^2 & 4\sigma_1^2 + 12\sigma_{12} + 9\sigma_2^2 \end{pmatrix} \end{matrix}$$

E' immediato verificare che la terza riga (colonna) di  $V[\mathbf{Y}]$  si ottiene come combinazione lineare delle prime due righe:

$$r_3 = 2r_1 + 3r_2,$$

ossia lo stesso vincolo lineare esistente fra le componenti di  $\mathbf{y}$ .

Pertanto  $\rho(V[\mathbf{Y}]) = 3 - 1 = 2$ .

### Rango della matrice di varianze e covarianze e relazioni fra variabili

La sola conoscenza del rango di una matrice di varianze e covarianze ci dice poco sul tipo di interrelazioni (eventualmente lineari) esistenti fra le  $p$  componenti del vettore aleatorio: ci dice solo se esistono uno o più legami lineari esatti.

Esistono altre indicatori associati alle matrici di varianze e covarianze che ci consentono di sapere qualcosa di più su tali interrelazioni.

## 1.3 Traccia di una matrice

La traccia di una matrice  $\mathbf{A}_{[p \times p]}$  quadrata,  $tr(\mathbf{A})$ , è definita come la somma degli elementi sulla diagonale principale:

$$tr(\mathbf{A}) = \sum_{i=1}^p a_{ii}$$

Alcune proprietà della traccia di una matrice:

•

$$tr(\mathbf{A}^T \mathbf{A}) = tr(\mathbf{A} \mathbf{A}^T) = \sum_{i=1}^p \sum_{j=1}^k a_{ij}^2$$

( $\mathbf{A}$  con  $p$  righe e  $k$  colonne)

- $tr(c\mathbf{A}) = c \, tr(\mathbf{A})$
- $tr(\mathbf{A} + \mathbf{B}) = tr(\mathbf{A}) + tr(\mathbf{B})$  ( $\mathbf{A}$  e  $\mathbf{B}$  quadrate dello stesso ordine)
- $tr(\mathbf{AB}) = tr(\mathbf{BA})$  ( $\mathbf{A}$  e  $\mathbf{B}$  quadrate dello stesso ordine)

### Traccia della Matrice di Varianze e covarianze

Se  $\mathbf{X}$  è un qualsiasi vettore aleatorio a  $p$  componenti con matrice di varianze e covarianze  $V[\mathbf{X}]$ , la traccia di  $V[\mathbf{X}]$  corrisponde alla somma delle varianze delle componenti di  $\mathbf{X}$ , ossia alla somma delle dispersioni lungo gli assi coordinati:

$$tr(V[\mathbf{X}]) = \sum_{i=1}^p V[X_i] = \sum_{i=1}^p \sigma_i^2$$

### 1.3.1 Varianza generalizzata

Un'altra misura di variabilità di una variabile aleatoria multipla  $\mathbf{X}$  è la varianza generalizzata (Wilks, 1932)

$$V_g[\mathbf{X}] = \text{Det}[V[\mathbf{X}]].$$

Il significato, anche in termini geometrici, di tale misura sarà più chiaro più avanti, in termini di autovalori e di ellissoidi di equiprobabilità per variabili normali multiple. Possiamo però vedere che  $V_g[\mathbf{X}]$  può essere nulla anche

se tutte le varianze sono maggiori di zero, e precisamente nel caso in cui  $V[\mathbf{X}]$  è di rango non pieno, ossia esiste almeno un vincolo lineare esatto fra le componenti di  $\mathbf{X}$ .

(La varianza generalizzata può essere ben interpretata per distribuzioni condizionate di variabili normali multivariate; ma anche come prodotto degli autovalori, ossia delle varianze delle componenti principali; o come volume dell'ellissoide di equiprobabilità in una normale multivariata)

## 1.4 Matrice inversa

Data una matrice quadrata  $\mathbf{A}$ , ( $\mathbf{A}_{[p \times p]}$ ), con  $|\mathbf{A}| \neq 0$ , si definisce inversa di  $\mathbf{A}$ , e si indica con  $\mathbf{A}^{-1}$ , una matrice tale che:

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_p \quad \text{matrice identità}$$

La condizione  $|\mathbf{A}| \neq 0$ , ossia che  $\mathbf{A}$  sia di rango pieno, è necessaria e sufficiente per l'esistenza e l'unicità di  $\mathbf{A}^{-1}$ .

E' noto infatti che l'elemento generico della matrice inversa  $(\mathbf{A})^{-1}$  è dato da:

$$\{\mathbf{A}^{-1}\}_{ij} = \frac{\mathbf{A}_{ji}}{|\mathbf{A}|}$$

essendo  $\mathbf{A}_{ij}$  il cofattore di  $a_{ij}$  ;

### Inversa di una matrice quadrata a rango pieno

Pertanto l'inversa è uguale alla trasposta della matrice aggiunta diviso il determinante della matrice.

$$\{\mathbf{A}^{-1}\}_{ij} = \frac{\mathbf{A}_{ji}}{|\mathbf{A}|}$$

essendo  $\mathbf{A}_{ij}$  il cofattore di  $a_{ij}$  ;

Ovviamente è una definizione utile solo per la dimostrare l'esistenza dell'inversa, ma non è conveniente numericamente per il calcolo: meglio ricorrere al metodo di Gauss-Siedel, o ad altri metodi di riduzione con la ricerca di elementi di pivot.

E' evidente che si farà ricorso, come sempre, a software matematico statistico, fornito sempre di buone routines per il calcolo dei determinanti e dell'inversa di una matrice: occorre comunque sempre accertarsi del grado di precisione fornito dal software usato, e cercare di usare la massima precisione numerica possibile; <sup>1</sup>

Il software R, con licenza di tipo *public domain*, ha degli algoritmi comunque ottimizzati per il calcolo matriciale (estremamente semplice da usare, dato che gli enti fondamentali in questo linguaggio sono le matrici e gli array, che si manipolano con funzioni che accettano matrici come argomenti).

Per impiegare al meglio R per l'algebra lineare, occorre installare R con le librerie *openblas*, in modo tale che R usi un set di istruzioni ottimizzate per buona parte di processori.

<sup>1</sup>ad esempio alcuni software nella risoluzione di sistemi di equazioni lineari, utilizzano una "extended precision calculation" che è sempre bene usare

ALCUNE PROPRIETÀ DELL'INVERSA DI UNA MATRICE		
$\mathbf{A}, \mathbf{B}$ quadrate di rango pieno		
$(\mathbf{A}^T)^{-1}$	=	$(\mathbf{A}^{-1})^T$
$(\mathbf{A}^{-1})^{-1}$	=	$\mathbf{A}$
$\mathbf{A}^{-1}$ è simmetrica	se e solo se	$\mathbf{A}$ è simmetrica
$\mathbf{A}^{-1}$ è diagonale	se e solo se	$\mathbf{A}$ è diagonale
$ \mathbf{A}^{-1} $	=	$ \mathbf{A} ^{-1}$
$(\mathbf{AB})^{-1}$	=	$\mathbf{B}^{-1}\mathbf{A}^{-1}$

#### 1.4.1    Significato degli elementi dell'inversa di una matrice di varianze e covarianze

Anche gli elementi dell' inversa di una matrice di varianze e covarianze hanno un preciso significato probabilistico statistico in termini di distribuzioni condizionate, [link con \(vedere anche → normale multivariata \)](#)

come si vedrà più avanti a proposito della normale multivariata.

Gli elementi non diagonali sono funzione della correlazione lineare condizionata, mentre gli elementi diagonali sono legati alla correlazione multipla.

livello  
avanzato

#### 1.4.2    Inversa di una matrice simmetrica partizionata

Supponiamo di avere una matrice simmetrica partizionata in quattro blocchi:

$$\mathbf{A} = \left( \begin{array}{c|c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \hline \mathbf{A}_{21} & \mathbf{A}_{22} \end{array} \right)$$

(primo e quarto blocco quadrati, e  $\mathbf{A}_{12} = \mathbf{A}_{21}$ ) Poniamo intanto: $\mathbf{A}_{11.2} = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{12}^T$  il motivo di questa notazione sarà chiarissimo (o almeno un po' meno oscuro di quanto non sia adesso) nel capitolo sulle distribuzioni condizionate di variabili normali multivariate. [link con distribuzioni condizionate di variab](#)

Si può dimostrare che, se esiste  $\mathbf{A}_{22}^{-1}$ , l'inversa della matrice partizionata può essere espressa come:

$$\mathbf{A}^{-1} = \left( \begin{array}{c|c} \mathbf{A}_{11.2}^{-1} & -\mathbf{A}_{11.2}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \hline -\mathbf{A}_{11.2}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} & \mathbf{A}_{22}^{-1}[\mathbf{A}_{12}^T\mathbf{A}_{11.2}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} + \mathbf{I}] \end{array} \right)$$

Si ha anche:

$$|\mathbf{A}| = |\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{12}^T||\mathbf{A}_{22}| = |\mathbf{A}_{11.2}||\mathbf{A}_{22}|$$

Il risultato sull'inversa di una matrice partizionata, arduo da ricordare, si dimostra effettuando il prodotto (sia destro che sinistro) per la matrice originaria partizionata  $\mathbf{A}$  e verificando che si ottiene la matrice identità.

Questo risultato è utile per ricavare le distribuzioni condizionate di variabili normali multivariate.

Nella regressione lineare multipla può servire il risultato particolare nel caso in cui  $\mathbf{A}_{11}$  è uno scalare  $a$  e quindi  $\mathbf{A}_{12}$  è un vettore riga che indico con  $\mathbf{y}^T$ . Il risultato è utile per esempio quando si aggiunge una riga, ossia si aggiunge una variabile, ad una matrice di varianze e covarianze di cui già si conosce l'inversa.

Abbiamo quindi:



$$\mathbf{A} = \left( \begin{array}{c|c} a & \mathbf{y}^T \\ \hline \mathbf{y} & \mathbf{A}_{22} \end{array} \right)$$

Si ha allora:

$$a_{11.2} = a - \mathbf{y}^T \mathbf{A}_{22}^{-1} \mathbf{y}$$

e quindi:

$$\mathbf{A}^{-1} = \left( \begin{array}{c|c} 1/a_{11.2} & -\mathbf{y}^T \mathbf{A}_{22}^{-1}/a_{11.2} \\ \hline -\mathbf{y}^T \mathbf{A}_{22}^{-1}/a_{11.2} & \mathbf{A}_{22}^{-1}[\mathbf{y} \mathbf{y}^T \mathbf{A}_{22}^{-1}/a_{11.2} + \mathbf{I}] \end{array} \right)$$

semplificabile in:

$$\mathbf{A}^{-1} = \frac{1}{a_{11.2}} \left( \begin{array}{c|c} 1 & -\mathbf{y}^T \mathbf{A}_{22}^{-1} \\ \hline -\mathbf{y}^T \mathbf{A}_{22}^{-1} & \mathbf{A}_{22}^{-1}(\mathbf{y} \mathbf{y}^T \mathbf{A}_{22}^{-1} + \mathbf{I} a_{11.2}) \end{array} \right)$$

Si ha in questo caso anche:

$$|\mathbf{A}| = |a - \mathbf{y}^T \mathbf{A}_{22}^{-1} \mathbf{y}| |\mathbf{A}_{22}| = |a_{11.2}| |\mathbf{A}_{22}|$$

### 1.4.3 Determinante e inversa di una matrice simmetrica orlata

Il risultato relativo a matrici  $\mathbf{A}$  simmetriche orlate, o partizionate in una riga e  $(p - 1)$  righe (e quindi 1 colonna e  $(p - 1)$  colonne) può essere ricavato in modo diretto senza far ricorso al risultato generale.

Il risultato è utile per ricavare le formule relative alla devianza spiegata nella regressione multipla e per ottenere gli indici di correlazione lineare parziale e sarà applicato a matrici di correlazione o di varianza e covarianza.

Sarà utile per attribuire un significato statistico agli elementi dell'inversa di una matrice di varianze e covarianze<sup>2</sup>

Supponiamo quindi di avere una matrice simmetrica  $\mathbf{A}$  di rango pieno  $p$  così partizionata:

$$\mathbf{A} = \left( \begin{array}{c|c} a & \mathbf{y}^T \\ \hline \mathbf{y} & \mathbf{A}_{22} \end{array} \right)$$

ove:

- $a$  è uno scalare
- $\mathbf{y}^T$  è un vettore riga
- $\mathbf{A}_{22}$  è una matrice di rango  $p - 1$  (ovviamente simmetrica) di cui si conoscono l'inversa  $(\mathbf{A}_{22})^{-1}$  e il determinante  $|\mathbf{A}|$ .

Troviamo prima il determinante di  $\mathbf{A}$  in funzione di quello di  $\mathbf{A}_{22}$ .

Consideriamo una matrice  $\mathbf{B}$  (partizionata in quattro parti delle stesse dimensioni delle parti di  $\mathbf{A}$ ) così definita:

$$\mathbf{B} = \left( \begin{array}{c|c} 1 & \mathbf{0}^T \\ \hline \mathbf{0} & (\mathbf{A}_{22})^{-1} \end{array} \right)$$

<sup>2</sup>Dal momento che il risultato verrà utilizzato più volte in questo testo, ho ritenuto utile inserirne anche una dimostrazione elementare, che non è comunque essenziale per l'impiego successivo che faremo del risultato di questa sezione nella regressione parziale e multipla

E' facile vedere, effettuando il prodotto  $\mathbf{AB}$ , che si ha:

$$\mathbf{AB} = \left( \begin{array}{c|c} a & \mathbf{y}^T \\ \hline \mathbf{y} & \mathbf{A}_{22} \end{array} \right) \left( \begin{array}{c|c} 1 & \mathbf{0}^T \\ \hline \mathbf{0} & (\mathbf{A}_{22})^{-1} \end{array} \right) = \left( \begin{array}{c|c} a & \mathbf{y}^T (\mathbf{A}_{22})^{-1} \\ \hline \mathbf{y} & \mathbf{I} \end{array} \right)$$

Per l'ultima matrice è facile vedere che:

$$\left| \left( \begin{array}{c|c} a & \mathbf{y}^T (\mathbf{A}_{22})^{-1} \\ \hline \mathbf{y} & \mathbf{I} \end{array} \right) \right| = a - \mathbf{y}^T (\mathbf{A}_{22})^{-1} \mathbf{y}$$

Mettendo insieme le relazioni precedenti e applicando le proprietà dei determinanti di prodotti di matrici si ha:

$$|\mathbf{A}||\mathbf{B}| = |\mathbf{AB}| = a - \mathbf{y}^T (\mathbf{A}_{22})^{-1} \mathbf{y}$$

Dal momento che si ha anche evidentemente:

$$|\mathbf{B}| = \left| \left( \begin{array}{c|c} 1 & \mathbf{0}^T \\ \hline \mathbf{0} & (\mathbf{A}_{22})^{-1} \end{array} \right) \right| = |(\mathbf{A}_{22})^{-1}| = 1/|\mathbf{A}_{22}|,$$

mettendo insieme le ultime due relazioni si ha infine:

$$|\mathbf{A}| = \frac{a - \mathbf{y}^T (\mathbf{A}_{22})^{-1} \mathbf{y}}{|\mathbf{B}|} = (a - \mathbf{y}^T (\mathbf{A}_{22})^{-1} \mathbf{y})|\mathbf{A}_{22}|. \quad (1)$$

Questo risultato consente semplicemente di ottenere esplicitamente il primo elemento dell'inversa di  $\mathbf{A}$ , ossia  $\{(\mathbf{A})^{-1}\}_{11}$ . Infatti:

$$\{(\mathbf{A})^{-1}\}_{11} = \frac{\text{cofattore}(\{\mathbf{A}\}_{11})}{|\mathbf{A}|}$$

Dato che:

$$\text{cofattore}(\{\mathbf{A}\}_{11}) = |\mathbf{A}_{22}| \quad \text{e} \quad |\mathbf{A}| = (a - \mathbf{y}^T (\mathbf{A}_{22})^{-1} \mathbf{y})|\mathbf{A}_{22}|$$

si ha:

$$\begin{aligned} \{(\mathbf{A})^{-1}\}_{11} &= \frac{\text{cofattore}(\{\mathbf{A}\}_{11})}{|\mathbf{A}|} = \frac{|\mathbf{A}_{22}|}{(a - \mathbf{y}^T (\mathbf{A}_{22})^{-1} \mathbf{y})|\mathbf{A}_{22}|} = \\ &= \frac{1}{a - \mathbf{y}^T (\mathbf{A}_{22})^{-1} \mathbf{y}} \end{aligned} \quad (2)$$

**livello  
avanza-  
to**

#### 1.4.4 Matrice inversa generalizzata

In certi casi, ad esempio per la risoluzione di sistemi di equazioni lineari a rango non pieno, conviene ricorrere alla cosiddetta inversa generalizzata.

inserire almeno un esempio numerico,  
se no il paragrafo non funziona

Data una matrice (anche rettangolare) di rango qualsiasi  $\mathbf{A}$ , si definisce inversa generalizzata di  $\mathbf{A}$ , e si indica con  $\mathbf{A}^-$ , una matrice tale che:

$$\mathbf{AA}^- \mathbf{A} = \mathbf{A}$$

L'inversa generalizzata di una qualsiasi matrice non è unica, tranne che per le matrici quadrate di rango pieno, per le quali si ha chiaramente:  $\mathbf{A}^- = \mathbf{A}^{-1}$

L'inversa generalizzata fornisce una delle soluzioni del sistema di equazioni lineari:

$$\mathbf{Ax} = \mathbf{b}$$

di rango anche non pieno, ovviamente nel caso in cui siano soddisfatte le condizioni per l'esistenza di soluzioni, ossia  $\rho(\mathbf{A}) = \rho(\mathbf{A}|\mathbf{b})$ .

Infatti con successive trasformazioni:

$$(\mathbf{A})\mathbf{x} = \mathbf{b}; \quad (\mathbf{AA}^-\mathbf{A})\mathbf{x} = \mathbf{b};$$

$$(\mathbf{AA}^-)(\mathbf{Ax}) = \mathbf{b} \quad (\mathbf{AA}^-)\mathbf{b} = \mathbf{b};$$

e infine:

$$\mathbf{A}(\mathbf{A}^-\mathbf{b}) = \mathbf{b},$$

per cui  $\mathbf{x} = \mathbf{A}^-\mathbf{b}$  è una soluzione del sistema originario.

Se la matrice  $\mathbf{A}$  è simmetrica valgono ulteriori proprietà. Si vedrà poi che mediante la decomposizione spettrale è possibile determinare una inversa generalizzata di una matrice simmetrica.

In effetti la definizione di inversa generalizzata è utile essenzialmente perchè consente di esprimere in modo compatto una generica soluzione di un sistema di equazioni lineari anche di rango non pieno.

## 1.5 Matrici ortogonali

Si definisce ortogonale una matrice quadrata  $\mathbf{A}$  di  $p$  righe e  $p$  colonne la cui trasposta coincide con l'inversa:

**Definizione di matrice ortogonale  $\mathbf{A}$**

$$\mathbf{A}^T = \mathbf{A}^{-1} \Rightarrow \mathbf{AA}^T = \mathbf{A}^T\mathbf{A} = \mathbf{I}$$

## 1.6 Matrici idempotenti

Si definisce idempotente una matrice quadrata  $\mathbf{A}$  di  $p$  righe e  $p$  colonne uguale al proprio quadrato:

**Definizione di matrice idempotente  $\mathbf{A}$**

$$\mathbf{A} = \mathbf{AA}$$

Se  $\mathbf{A}$  è idempotente allora valgono le seguenti proprietà:

$\mathbf{A} = \mathbf{A}\mathbf{A} = \dots = \mathbf{A}^n \quad \forall n, \quad n > 0$   $\mathbf{A}$  è uguale a tutte le sue potenze  
 $\mathbf{A}^n$  è idempotente  $\forall n, \quad n > 0$ . Tutte le potenze di  $\mathbf{A}$  sono idempotenti  
 $\mathbf{I} - \mathbf{A}$  è idempotente infatti:

$$\begin{aligned} [\mathbf{I} - \mathbf{A}][\mathbf{I} - \mathbf{A}] &= \mathbf{I}^2 - 2\mathbf{A} + \mathbf{A}^2 = \\ &= \mathbf{I} - 2\mathbf{A} + \mathbf{A} = \mathbf{I} - \mathbf{A} \end{aligned}$$

$$\text{tr}(\mathbf{A}) = \rho(\mathbf{A})$$

La traccia di  $\mathbf{A}$  è uguale al suo rango

Indicando con  $\lambda_i$  gli autovalori di  $\mathbf{A}$  si ha:

$$\begin{cases} \lambda_i = 1 & \text{se } i = 1, 2, \dots, \rho(\mathbf{A}) \\ \lambda_i = 0 & \text{se } i = \rho(\mathbf{A}) + 1, \dots, p \end{cases}$$

Infatti dal momento che gli autovalori delle potenze di una matrice sono uguali alle potenze degli autovalori, essendo  $\mathbf{A} = \mathbf{A}\mathbf{A}$ , si deve avere  $\lambda_i = \lambda_i^2$ , per cui  $\lambda_i$  può essere solo 0 o 1.

Risulta ovvio dalla definizione che l'unica matrice idempotente di rango pieno è la matrice identità; gli scalari idempotenti sono 0 e 1.

### 1.6.1 Esempio

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T, \quad \forall \mathbf{X}, \quad \text{purché esista: } (\mathbf{X}^T\mathbf{X})^{-1}$$

è una matrice idempotente (con  $\rho(\mathbf{H}) = \rho(\mathbf{X})$ ), come si verifica facilmente effettuando il prodotto:

$$\mathbf{H}\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T = \mathbf{H}.$$

Il concetto di matrice idempotente, in assoluto non particolarmente rilevante, è uno strumento tecnico che sarà utilissimo per lo studio delle proprietà delle forme quadratiche in variabili normali, e per lo studio di particolari quantità che scaturiscono dall'analisi dei modelli lineari; proprio nei modelli lineari la matrice  $\mathbf{H}$  viene chiamata *hat matrix*, per motivi chiariti in quel capitolo.

### 1.6.2 Esempi di matrici idempotenti di rango 2

Come è facile verificare mediante calcolo diretto, le seguenti matrici sono tutte idempotenti:

$$\mathbf{A}_1 = \begin{pmatrix} 9/17 & 8/17 & -2/17 & -2/17 \\ 8/17 & 9/17 & 2/17 & 2/17 \\ -2/17 & 2/17 & 8/17 & 8/17 \\ -2/17 & 2/17 & 8/17 & 8/17 \end{pmatrix}$$

$$\mathbf{A}_2 = \begin{pmatrix} 2/3 & -1/3 & 1/3 \\ -1/3 & 2/3 & 1/3 \\ 1/3 & 1/3 & 2/3 \end{pmatrix} \quad \mathbf{A}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/5 & 2/5 \\ 0 & 2/5 & 4/5 \end{pmatrix}$$

## 1.7 Forme quadratiche

Se  $\mathbf{A}$  è una matrice quadrata simmetrica  $p \times p$ , e  $t$  è un vettore di  $p$  componenti, si definisce forma quadratica la funzione omogenea di secondo grado:

$$\begin{aligned} \mathbf{Q} = \mathbf{t}^T \mathbf{A} \mathbf{t} &= a_{11}t_1^2 + a_{22}t_2^2 + \dots + a_{ii}t_i^2 + \dots + a_{pp}t_p^2 + \\ &+ 2a_{12}t_1t_2 + \dots + 2a_{ij}t_it_j + \dots + 2a_{p-1,p}t_{p-1}t_p \end{aligned}$$

### Forme Quadratiche positive

se  $\mathbf{t}^T \mathbf{A} \mathbf{t} > 0, \forall \mathbf{t}, \mathbf{t} \neq 0 : \Rightarrow \mathbf{A}$  è definita positiva  
se  $\mathbf{t}^T \mathbf{A} \mathbf{t} \geq 0, \forall \mathbf{t}, \mathbf{t} \neq 0 : \Rightarrow \mathbf{A}$  è semidefinita positiva

In effetti si dice definita (o semidefinita) sia la matrice che la forma quadratica corrispondente.

### 1.7.1 Significato geometrico delle forme quadratiche

Una forma quadratica definita positiva definisce un'ellissoide in uno spazio  $p$ -dimensionale mediante l'equazione  $\mathbf{t}^T \mathbf{A} \mathbf{t} = k$ . Il volume di tale ellissoide è funzione del determinante della matrice  $\mathbf{A}$ . Questo aspetto sarà importante quando si parlerà di distribuzione normale multivariata.

inserire  
esempi e  
figure

## 2 Calcolo differenziale con vettori e matrici

### 2.1 Gradiente di una funzione

Data una funzione di  $k$  variabili  $f(x_1, x_2, \dots, x_k)$ , si definisce gradiente della funzione il vettore (colonna!) formato dalle derivate parziali di  $f(\cdot)$  rispetto a ciascuna variabile:

$$\frac{\nabla f(x_1, x_2, \dots, x_k)}{(x_1, x_2, \dots, x_k)} = \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial f(x_1, x_2, \dots, x_k)}{\partial x_1} \\ \frac{\partial f(x_1, x_2, \dots, x_k)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x_1, x_2, \dots, x_k)}{\partial x_j} \\ \vdots \\ \frac{\partial f(x_1, x_2, \dots, x_k)}{\partial x_k} \end{pmatrix} \quad (3)$$

### 2.2 Hessiano di una funzione

Data una funzione di  $k$  variabili  $f(x_1, x_2, \dots, x_k)$ , si definisce Hessiano della funzione la matrice formata dalle derivate parziali seconde di  $f(\cdot)$  rispetto a ciascuna coppia di variabili:

$$\frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}^T} = \begin{pmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \dots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_j} & \dots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_k} \\ \dots & \dots & \dots & \dots & \dots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_1} & \dots & \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} & \dots & \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_k} \\ \dots & \dots & \dots & \ddots & \dots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_k \partial x_1} & \dots & \frac{\partial^2 f(\mathbf{x})}{\partial x_k \partial x_j} & \dots & \frac{\partial^2 f(\mathbf{x})}{\partial x_k^2} \end{pmatrix} \quad (4)$$

## 2.3 Derivate di forme lineari e quadratiche

*Gradiente di combinazioni lineari di variabili:*

$$\frac{\partial \mathbf{x}^T \mathbf{b}}{\partial \mathbf{x}} = \mathbf{b}$$

( $\mathbf{x}, \mathbf{b}$  vettori di  $p$  componenti)

Infatti:

$$\mathbf{x}^T \mathbf{b} = b_1 x_1 + b_2 x_2 + \dots + b_p x_p$$

per cui la singola derivata parziale è data da:

$$\frac{\partial \mathbf{x}^T \mathbf{b}}{\partial x_i} = b_i \quad i = 1, 2, \dots, p$$

e quindi il risultato in forma vettoriale:

$$\frac{\partial \mathbf{x}^T \mathbf{b}}{\partial \mathbf{x}} = \mathbf{b}$$

In generale, per il gradiente di un vettore di combinazioni lineari si ha:

$$\frac{\partial \mathbf{x}^T \mathbf{B}}{\partial \mathbf{x}} = \mathbf{B}$$

ove:  $\mathbf{x}$  è un vettore di  $p$  componenti e

$\mathbf{B}$  è una matrice di  $p \times k$  elementi e di elemento generico  $b_{ij}$

*Gradiente ed Hessiano di una forma quadratica:*

$$\frac{\partial \mathbf{x}^T \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{A} \mathbf{x}$$

$$\frac{\partial^2 \mathbf{x}^T \mathbf{A} \mathbf{x}}{\partial \mathbf{x} \partial \mathbf{x}^T} = 2\mathbf{A}$$

$\mathbf{x}$  vettore (colonna!) di  $p$  componenti  $\mathbf{A}$  è una matrice simmetrica di  $p \times p$  elementi e di elemento generico  $a_{ij}$

Infatti:

$$\begin{aligned} Q &= \mathbf{x}^T \mathbf{A} \mathbf{x} = \\ &= a_{11}x_1^2 + a_{22}x_2^2 + \dots + a_{ii}x_i^2 + \dots + a_{pp}x_p^2 + 2a_{12}x_1x_2 + \\ &\quad + \dots + 2a_{ij}x_ix_j + \dots + 2a_{p-1,p}x_{p-1}x_p \end{aligned}$$

$$\begin{aligned} \frac{\partial Q}{\partial x_i} &= \frac{\partial \mathbf{x}^T \mathbf{A} \mathbf{x}}{\partial x_i} = \\ &= 2a_{ii}x_i + 2a_{i1}x_1 + \dots + 2a_{ij}x_j + \dots + 2a_{ip}x_p \\ &= 2\mathbf{a}_i^T \mathbf{x} \end{aligned}$$

essendo  $\mathbf{a}_i^T$  l'  $i$ -esima riga di  $\mathbf{A}$ .

Quindi segue il risultato in forma vettoriale, tenendo presente che derivando rispetto a tutti gli elementi di  $\mathbf{x}$ , le righe  $\mathbf{a}_i$  ricostituiscono la matrice  $\mathbf{A}$ :

$$\frac{\partial \mathbf{x}^T \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{A} \mathbf{x}$$

Derivando ancora, si ottiene facilmente il risultato per le derivate seconde di una forma quadratica:

$$\frac{\partial^2 \mathbf{x}^T \mathbf{A} \mathbf{x}}{\partial \mathbf{x} \partial \mathbf{x}^T} = 2\mathbf{A}$$

Jacobiano di una trasformazione lineare:

Sia  $\mathbf{A}$  una matrice quadrata; data la trasformazione lineare:

$$\mathbf{x} = \mathbf{A}\mathbf{y} + \mathbf{b},$$

lo Jacobiano di tale trasformazione è dato da:

$$J(\mathbf{y}) = \left| \frac{\partial \mathbf{x}}{\partial \mathbf{y}} \right| = |\mathbf{A}|$$

ossia il valore assoluto del determinante di  $\mathbf{A}$ .

## 2.4 Derivate di inverse e di determinanti

Sia  $\mathbf{B} = \mathbf{A}^{-1}$ , di elemento generico:  $b_{ij} = \mathbf{A}_{ji}/|\mathbf{A}|$ , indicato con  $\mathbf{A}_{ij}$  il cofattore di  $a_{ij}$  in  $\mathbf{A}$  si può dimostrare che:

$$\frac{\partial b_{ij}}{\partial a_{hk}} = -b_{ih}b_{kj} = \mathbf{A}_{hi}\mathbf{A}_{jk}/|\mathbf{A}|^2$$

Se  $\mathbf{A} = \mathbf{A}^T$  (ossia  $\mathbf{A}$  è simmetrica)

$$\frac{\partial |\mathbf{A}|}{\partial a_{ii}} = \mathbf{A}_{ii}$$

$$\frac{\partial |\mathbf{A}|}{\partial a_{ij}} = 2\mathbf{A}_{ij} \quad i \neq j$$

(ricordando che  $|\mathbf{A}| = \sum_{i=1}^p a_{ij}\mathbf{A}_{ij}$ ).

## 3 Autovalori e autovettori

Gli autovalori e gli autovettori <sup>3</sup> sono delle quantità associate ad una matrice quadrata, che ne riassumono alcune caratteristiche essenziali.

In particolare per una matrice simmetrica si possono dimostrare proprietà molto forti.

Se poi la matrice simmetrica è una matrice di varianze e covarianze, si possono attribuire particolari significati a tali quantità, sia nel caso di matrici di varianze e covarianze di vettori di variabili aleatorie che nel caso di matrici di varianze e covarianze empiriche di vettori di variabili statistiche osservate, sebbene la loro interpretabilità, dal punto di vista dello statistico, non sia sempre agevole, se non in particolari contesti.

Nell'analisi esplorativa dei dati sono importanti per misurare la correlazione generale fra tutte le variabili, per determinare il grado di collinearità presente in un insieme di dati multivariati o in un vettore di variabili aleatorie e per trovare un sistema di riferimento ortogonale (per rotazione).

In questo corso saranno impiegati per scopi esplorativi e per lo studio della multicollinearità nella regressione multipla; per quanto riguarda i vettori

---

<sup>3</sup>Termini italiani: autovalore, radice caratteristica; Termini inglesi: characteristic roots, eigenvalue Termini italiani: autovettore, vettore caratteristico Termini inglesi: eigenvector

aleatori, si vedrà presto l'interpretazione migliore degli autovettori e degli autovalori per vettori aleatori distribuiti secondo una normale multivariata.

Nelle pagine che seguono vengono brevemente richiamate le proprietà algebriche e geometriche degli autovalori e degli autovettori, con riferimento in particolare alle caratteristiche che verranno successivamente sfruttate nel corso. Resta sottinteso che si tratta semplicemente di un richiamo di nozioni che in modo più completo e sistematico vanno approfondite, se non lo si è già fatto, in un corso di algebra lineare.

### 3.1 Definizione di autovettore e autovalore

Data la matrice quadrata  $\mathbf{A}$ , si vuole trovare la soluzione non banale<sup>4</sup>  $\boldsymbol{\gamma}$  del sistema di equazioni:

**Autovettore di una matrice quadrata  $\mathbf{A}$**

$$\mathbf{A}\boldsymbol{\gamma} = \lambda\boldsymbol{\gamma}$$

Si vuole quindi trovare un vettore  $\boldsymbol{\gamma}$  la cui proiezione secondo lo spazio definito da  $\mathbf{A}$  sia parallela al vettore stesso.

Si tratta di un sistema omogeneo nell'incognita  $\boldsymbol{\gamma}$ , infatti:

$$\mathbf{A}\boldsymbol{\gamma} - \lambda\boldsymbol{\gamma} = \mathbf{0}_p$$

e quindi:

$$[\mathbf{A} - \lambda\mathbf{I}_p]\boldsymbol{\gamma} = \mathbf{0}_p$$

Condizione necessaria per avere una soluzione  $\boldsymbol{\gamma}$  diversa dal vettore nullo è che:

$$|\mathbf{A} - \lambda\mathbf{I}_p| = 0.$$

La precedente è un'equazione di grado  $p$  in  $\lambda$ , per cui vi saranno  $p$  autovalori complessi (distinti e non):

autovalori

$$\lambda_1, \lambda_2, \dots, \lambda_i, \dots, \lambda_p.$$

L'equazione è di grado  $p$  in quanto sviluppando il determinante di  $\mathbf{A} - \lambda\mathbf{I}_p$ , il primo termine è:  $\prod_{i=1}^p (a_{ii} - \lambda)$  che risulta essere di grado  $p$  in  $\lambda$

Di solito si conviene di normalizzare gli autovettori in modo tale che:

$$\boldsymbol{\gamma}^T \boldsymbol{\gamma} = 1.$$

Infatti in corrispondenza di ciascun autovalore  $\lambda_i$  vi sarà certamente un'infinità di autovettori proporzionali  $\boldsymbol{\gamma}_i$  (Si vede subito dalla definizione di autovettore: se  $\boldsymbol{\gamma}_i$  è un autovettore lo è anche  $k\boldsymbol{\gamma}_i$ ).

In ogni caso resta l'ambiguità sul segno di  $\boldsymbol{\gamma}$ .

---

<sup>4</sup>Una soluzione non banale è una soluzione con elementi non tutti nulli



### 3.2 Proprietà generali degli autovalori

Dall'equazione fondamentale:

$$|\mathbf{A} - \lambda \mathbf{I}_p| = 0,$$

si vede che il polinomio di grado  $p$  in  $\lambda$ :

$$q(\lambda) = |\mathbf{A} - \lambda \mathbf{I}_p|$$

si può esprimere in funzione delle  $p$  radici complesse  $\lambda_i$ :

$$q(\lambda) = |\mathbf{A} - \lambda \mathbf{I}_p| = \prod_{i=1}^p (\lambda_i - \lambda)$$

(si può dimostrare dalle proprietà relative alla fattorizzazione dei polinomi).

Per cui si ha subito (sfruttando le proprietà dei polinomi):

#### Traccia e determinante in funzione degli autovalori

$$|\mathbf{A}| = \prod_{i=1}^p \lambda_i$$

Il determinante di una matrice è uguale al prodotto dei suoi autovalori.

$$tr(\mathbf{A}) = \sum_{i=1}^p \lambda_i$$

La traccia di una matrice è uguale alla somma dei suoi autovalori.

### 3.3 Autovalori e autovettori di matrici simmetriche

Per gli autovalori e gli autovettori di una matrice simmetrica  $\mathbf{A}$  si possono dimostrare proprietà molto forti, corrispondenti a molte caratteristiche essenziali della matrice (in generale molte proprietà valgono anche per matrici hermitiane, ossia con elementi  $a_{ij}$  e  $a_{ji}$  complessi coniugati, tuttavia per gli argomenti da noi trattati è sufficiente riferirci a matrici simmetriche reali)

Se  $\mathbf{A}$  è simmetrica tutti gli autovalori e gli autovettori sono reali, per cui convenzionalmente gli autovalori  $\lambda_i$  vengono indicizzati in ordine decrescente:

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_i \geq \dots \geq \lambda_p.$$

Se  $\mathbf{A}$  è simmetrica, il numero degli autovalori non nulli è uguale a  $\rho(\mathbf{A})$  (rango di  $\mathbf{A}$ ). Se per  $i \neq j$  i corrispondenti autovalori  $\lambda_i$  e  $\lambda_j$  sono distinti si ha:

$$\begin{cases} \boldsymbol{\gamma}_i^T \boldsymbol{\gamma}_j &= 0 \\ \boldsymbol{\gamma}_i^T \mathbf{A} \boldsymbol{\gamma}_j &= 0 \end{cases} \quad (\text{ortogonalità})$$

Infatti  $\lambda_i$  e  $\lambda_j$ , insieme ai corrispondenti autovettori, forniscono due soluzioni distinte del sistema di equazioni:  $\mathbf{A}\boldsymbol{\gamma} = \lambda\boldsymbol{\gamma}$ , e quindi valgono contemporaneamente i due gruppi di eguaglianze:

$$\begin{cases} \mathbf{A}\boldsymbol{\gamma}_i &= \lambda_i \boldsymbol{\gamma}_i \\ \mathbf{A}\boldsymbol{\gamma}_j &= \lambda_j \boldsymbol{\gamma}_j \end{cases}$$

Premoltiplicando ambo i membri del primo sistema per  $\boldsymbol{\gamma}_j^T$  e i due membri del secondo per  $\boldsymbol{\gamma}_i^T$  otteniamo due eguaglianze fra scalari:

$$\begin{cases} \gamma_i^T \mathbf{A} \gamma_i = \gamma_i^T \lambda_i \gamma_i \\ \gamma_i^T \mathbf{A} \gamma_j = \gamma_i^T \lambda_j \gamma_j \end{cases}$$

in cui i primi membri sono uguali, perchè  $\gamma_j^T \mathbf{A} \gamma_i$  è la trasposta di  $\gamma_i^T \mathbf{A} \gamma_j$ , ed essendo scalari sono uguali, per cui uguagliando i secondi membri si ha:

$$\gamma_j^T \lambda_i \gamma_i = \gamma_i^T \lambda_j \gamma_j$$

e quindi:

$$\gamma_i^T \gamma_j (\lambda_i - \lambda_j) = 0$$

e infine, avendo supposto distinti i due autovalori,  $(\lambda_i - \lambda_j) \neq 0$ , per cui deve essere:

$$\gamma_i^T \gamma_j = 0.$$

Saranno quindi nulli anche i primi membri, per cui:

$$\gamma_i^T \mathbf{A} \gamma_j = 0.$$

In ogni caso si può dimostrare per ogni autovalore di molteplicità  $m$ ,  $m$  autovettori corrispondenti possono essere rimpiazzati da  $m$  loro combinazioni lineari indipendenti. Gli autovettori possono essere scelti in modo da soddisfare i vincoli di ortogonalità per ogni coppia  $i \neq j$

$$\gamma_i^T \gamma_j = 0 \quad \text{ed anche} \quad \gamma_i^T \mathbf{A} \gamma_j = 0$$

Pertanto se  $\mathbf{\Gamma}$  è la matrice che ha come colonne gli autovettori  $\mathbf{y}_i$ , allora per l'ortogonalità fra gli autovettori si ha:

$$\mathbf{\Gamma}^T \mathbf{\Gamma} = \mathbf{I};$$

ed anche:

$$\mathbf{\Gamma}^{-1} = \mathbf{\Gamma}^T,$$

e quindi:

$$\mathbf{\Gamma} \mathbf{\Gamma}^T = \mathbf{I}.$$

(queste ultime proprietà valgono comunque per matrici ortogonali)

### 3.3.1 Diagonalizzazione di una matrice simmetrica

Dalla definizione di autovettore si ha anche l'importante proprietà:  
(avendo posto  $\mathbf{\Lambda} = \text{Diag}(\boldsymbol{\lambda})$ ).

$$\mathbf{\Gamma}^T \mathbf{A} \mathbf{\Gamma} = \text{Diag}(\boldsymbol{\lambda}) = \mathbf{\Lambda} \tag{5}$$

Dalla definizione si ha infatti:

$$\mathbf{A} \gamma_i = \lambda_i \gamma_i$$

Premoltiplicando ambo i membri per  $\gamma_j^T$  si ha:

$$\begin{cases} \gamma_i^T \mathbf{A} \gamma_i = \lambda_i & \text{se } i = j \\ \gamma_j^T \mathbf{A} \gamma_i = 0 & \text{se } i \neq j \end{cases}$$

La diagonalizzazione di una matrice simmetrica sarà importante quando  $\mathbf{A}$  è una matrice di varianze e covarianze.

diagonalizzazione

Dal risultato fondamentale sulla diagonalizzazione di una matrice simmetrica si può ricavare un altro risultato molto utile:

Data una matrice simmetrica definita positiva  $\mathbf{A}$  di rango pieno è possibile sempre trovare una matrice  $\mathbf{B}$  tale che:

$$\mathbf{B}^T \mathbf{A} \mathbf{B} = \mathbf{I}$$

E' facile vedere che le colonne della matrice  $\mathbf{B}$  si ottengono riscalandogli autovettori di  $\mathbf{A}$ , ossia con:  $\gamma_i / \sqrt{\lambda_i}$  (dal momento che la matrice è di rango pieno i suoi autovalori sono tutti positivi)

### 3.3.2 Decomposizione spettrale di una matrice simmetrica

Dalla relazione  $\mathbf{\Gamma}^T \mathbf{A} \mathbf{\Gamma} = \mathbf{\Lambda}$ , si ha anche, premoltiplicando ambo i membri per  $\mathbf{\Gamma}$  e postmoltiplicando per  $\mathbf{\Gamma}^T$ :

#### Decomposizione canonica (o spettrale) di $\mathbf{A}$

$$\mathbf{A} = \mathbf{\Gamma} \mathbf{\Lambda} \mathbf{\Gamma}^T = \lambda_1 \gamma_1 \gamma_1^T + \lambda_2 \gamma_2 \gamma_2^T + \dots + \lambda_p \gamma_p \gamma_p^T$$

relazione fondamentale per la ricostruzione di una matrice simmetrica a partire dagli autovettori. I primi  $k$  termini ( $k < p$ ) forniscono un'approssimazione della matrice  $\mathbf{A}$  di rango  $k$ .

### 3.3.3 Autovalori di inverse e di potenze

Vediamo che relazioni esistono fra gli autovalori e gli autovettori di una matrice e quelli della sua inversa e delle sue potenze.

Operiamo ancora sull'equazione che definisce gli autovalori e gli autovettori:

$$\mathbf{A} \gamma_i = \lambda_i \gamma_i$$

Se il rango di  $\mathbf{A}$  è pieno, premoltiplicando ambo i membri per  $\lambda_i^{-1} \mathbf{A}^{-1}$ , si ottiene

$$\lambda_i^{-1} \mathbf{A}^{-1} \mathbf{A} \gamma_i = \lambda_i^{-1} \mathbf{A}^{-1} \lambda_i \gamma_i \Rightarrow \lambda_i^{-1} \gamma_i = \mathbf{A}^{-1} \gamma_i$$

e si vede facilmente che:

$$\lambda_i(\mathbf{A}^{-1}) = [\lambda_i(\mathbf{A})]^{-1}$$

(a meno di un riordinamento degli indici)

Qualunque sia il rango di  $\mathbf{A}$ , premoltiplicando ripetutamente ambo i membri per  $\mathbf{A}$ , si dimostra per induzione che:

$$\lambda_i(\mathbf{A}^k) = [\lambda_i(\mathbf{A})]^k$$

In entrambi i casi gli autovettori sono sempre quelli di  $\mathbf{A}$ .

Matrice	$\lambda$	$\gamma$	Decomposizione canonica
$\mathbf{A}$	$\lambda_i$	$\gamma_i$	$\mathbf{A} = \mathbf{\Gamma} \mathbf{\Lambda} \mathbf{\Gamma}^T = \sum_{i=1}^p \lambda_i \gamma_i \gamma_i^T$
$\mathbf{A}^{-1} ( \mathbf{A}  \neq 0)$	$\lambda_i^{-1}$	$\gamma_i$	$\mathbf{A}^{-1} = \mathbf{\Gamma} \mathbf{\Lambda}^{-1} \mathbf{\Gamma}^T = \sum_{i=1}^p \gamma_i \gamma_i^T / \lambda_i$
$\mathbf{A}^k$ intero	$\lambda_i^k$	$\gamma_i$	$\mathbf{A}^k = \mathbf{\Gamma} \mathbf{\Lambda}^k \mathbf{\Gamma}^T = \sum_{i=1}^p \lambda_i^k \gamma_i \gamma_i^T$

### 3.3.4 Autovalori di una forma quadratica definita positiva

#### Autovalori di una forma quadratica definita positiva

Una matrice simmetrica  $\mathbf{A}$  è definita positiva, *se e solo se* tutti i suoi autovalori sono positivi.

$\mathbf{A}$  è semidefinita positiva, *se e solo se* tutti i suoi autovalori sono non negativi.

Infatti ricorrendo agli autovalori ed agli autovettori di  $\mathbf{A}$  si può scrivere  $\mathbf{A}$  secondo la decomposizione canonica  $\mathbf{A} = \mathbf{\Gamma} \mathbf{\Lambda} \mathbf{\Gamma}^T$ :

$$\mathbf{Q}(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} = \mathbf{x}^T \mathbf{\Gamma} \mathbf{\Lambda} \mathbf{\Gamma}^T \mathbf{x}$$

Ponendo ora  $\mathbf{y} = \mathbf{\Gamma}^T \mathbf{x}$ , si ha:

$$\mathbf{Q}(\mathbf{x}) = \mathbf{y}^T \mathbf{\Lambda} \mathbf{y} = \sum_{i=1}^p \lambda_i y_i^2$$

da cui deriva il risultato sulla positività di  $\mathbf{Q}(\mathbf{x})$ .

Si vede anche che una forma quadratica si può sempre esprimere come somma ponderata di quadrati di variabili ruotate secondo gli autovettori di  $\mathbf{A}$ .

Infatti si è può sempre trasformare un ellissoide qualsiasi, mediante opportune trasformazioni lineari ortogonali, in un ellissoide ad assi paralleli a quelli coordinati, e quindi, mediante cambiamenti di scala, in un' ipersfera.

**un esempio**

Questi concetti saranno impegnati nella sezione sull'analisi delle componenti principali (5 )

## 4 Richiami di alcune proprietà dei vettori aleatori.

### 4.1 Momenti primo e secondo multivariati di vettori aleatori

Sia  $\mathbf{X}$  un qualsiasi vettore di variabili casuali, sia discrete che continue, con  $p$  componenti:

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

Definiamo i primi due momenti di un vettore aleatorio, con una notazione analoga a quella del caso univariato:

momenti  
di un  
vettore  
aleatorio

#### Momento primo e secondo multivariati

*vettore di speranze matematiche:*

$$\mathbf{E}[\mathbf{X}] = \boldsymbol{\mu} \quad \text{momento primo (multivariato) dall'origine}$$

*matrice di varianze e covarianze:*

$$\mathbf{V}[\mathbf{X}] = \mathbf{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T] \quad \text{momento secondo (multivariato) centrale}$$

*Ovviamente nella definizione si presuppone l'esistenza dei momenti primi e secondi delle varie componenti e coppie di componenti.*

- $\boldsymbol{\mu}$  è un vettore di  $p$  elementi, con elemento generico:

$$\mathbf{E}[X_i] = \mu_i$$

- $\mathbf{V}[\mathbf{X}]$  è una matrice simmetrica  $p \times p$  di elemento generico:

$$\sigma_{ij} = \{\mathbf{V}[\mathbf{X}]\}_{ij} = \mathbf{E}[(X_i - \mu_i)(X_j - \mu_j)] = \mathbf{E}[X_i X_j] - \mu_i \mu_j$$

e quindi in definitiva si ha:

$$\mathbf{E}[\mathbf{X}] = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_i \\ \vdots \\ \mu_p \end{pmatrix} \quad \mathbf{V}[\mathbf{X}] = \begin{pmatrix} \sigma_1^2 & \dots & \sigma_{1i} & \dots & \sigma_{1p} \\ \dots & \dots & \dots & \dots & \dots \\ \sigma_{1i} & \dots & \sigma_i^2 & \dots & \sigma_{ip} \\ \dots & \dots & \dots & \ddots & \dots \\ \sigma_{1p} & \dots & \sigma_{ip} & \dots & \sigma_p^2 \end{pmatrix}$$

Per gli elementi sulla diagonale principale di  $\mathbf{V}[\mathbf{X}]$ , ossia per le varianze delle singole componenti, invece della notazione  $\sigma_{ii}$  si impiega la notazione  $\sigma_i^2$  per uniformità col simbolismo nel caso univariato.

#### Momenti centrati e momenti secondi dall'origine

Vale la nota relazione in termini di momenti multivariati dall'origine:

$$\mathbf{V}[\mathbf{X}] = \mathbf{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T] = \mathbf{E}[\mathbf{X}\mathbf{X}^T] - \boldsymbol{\mu}\boldsymbol{\mu}^T$$

Si può definire la *matrice di correlazione*,  $R(\mathbf{X})$ , di elemento generico:

$$\rho_{ij} = \{R(\mathbf{X})\}_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$$

che, ovviamente, è simmetrica ed ha elementi diagonali tutti uguali ad uno:

$$R(\mathbf{X}) = \begin{pmatrix} 1 & \dots & \rho_{1i} & \dots & \rho_{1p} \\ \dots & \dots & \dots & \dots & \dots \\ \rho_{1i} & \dots & 1 & \dots & \rho_{ip} \\ \dots & \dots & \dots & \dots & \dots \\ \rho_{1p} & \dots & \rho_{ip} & \dots & 1 \end{pmatrix}$$

Dati gli argomenti che qui trattiamo, evidentemente abbiamo supposto di avere un numero  $p$  fissato di variabili, e non una sequenza di variabili aleatorie anche infinita, come avviene per esempio nella definizione di processi aleatori.

E'ovviamente possibile definire momenti multivariati di  $\mathbf{X}$  centrali e non centrali di ordine superiore rispetto al secondo, ma per gli argomenti ora trattati non è necessario.

Come per le variabili aleatorie semplici i momenti di ordine 3 e 4 forniscono degli indici di forma, i momenti multivariati di ordine superiore al secondo forniscono degli indici di forma multivariati, degli indicatori di allontanamento dalla multinormalità, indici di non linearità delle regressioni e di eteroscedasticità.

[link con sezione uso dei momenti bivariati nell'analisi dei residui](#)

In effetti la matrice di varianze e covarianze fornisce informazioni solo sulla variabilità delle singole componenti e sulle loro correlazioni lineari, sia per le distribuzioni congiunte che per quelle condizionate (elementi della matrice inversa). Per le combinazioni lineari di variabili si useranno gli autovalori e gli autovettori della matrice di varianze e covarianze.

momenti  
multiva-  
riati

[chiarire](#)

link o riferimento

(vedere anche → schema delle relazioni lineari)

Come chiarito nella parte sulla normale multivariata, in analogia al caso univariato, la normale multivariata dipende solo dai primi due momenti multivariati, per cui la conoscenza della matrice di varianza e covarianza è in quel caso sufficiente per valutare qualsiasi relazione di tipo lineare fra componenti

[chiarire](#)

[link con normale multivariata](#)

## 4.2 Momenti di una trasformata lineare di un vettore aleatorio

Sia ora  $\mathbf{Y}$  una v.c. a  $k$  componenti, ottenuta mediante una qualsiasi trasformazione lineare di  $\mathbf{X}$ :

$$\mathbf{Y} = \mathbf{A}_{[k \times p]} \mathbf{X} + c_{[k \times 1]}$$

La matrice  $\mathbf{A}$  ha  $k$  righe e  $p$  colonne e per il resto è qualsiasi, nel senso che il suo rango può anche essere inferiore a  $\min(k, p)$ . Il vettore  $c$  ha  $k$  elementi. Con semplici passaggi si vede come data la matrice  $\mathbf{A}$  e il vettore  $c$  è possibile ottenere tutti i momenti di  $\mathbf{Y}$  in funzione di quelli di  $\mathbf{X}$ :

$$\begin{aligned} E[\mathbf{Y}] &= E[\mathbf{A}\mathbf{X} + c] = \mathbf{A}E[\mathbf{X}] + c = \mathbf{A}\boldsymbol{\mu} + c \\ V[\mathbf{Y}] &= V[\mathbf{A}\mathbf{X} + c] = E[(\mathbf{A}\mathbf{X} + c - \mathbf{A}\boldsymbol{\mu} - c)(\mathbf{A}\mathbf{X} + c - \mathbf{A}\boldsymbol{\mu} - c)^T] = \\ &= E[\mathbf{A}(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T \mathbf{A}^T] = \mathbf{A}V[\mathbf{X}] \mathbf{A}^T \end{aligned}$$

**Momenti di una trasformazione lineare di un vettore  
aleatorio  $\mathbf{X}$ :  $\mathbf{Y} = \mathbf{A}\mathbf{X} + c$**

$$\begin{aligned} E[\mathbf{A}\mathbf{X} + c] &= \mathbf{A}E[\mathbf{X}] + c && \text{Speranza matematica} \\ V[\mathbf{A}\mathbf{X} + c] &= \mathbf{A}V[\mathbf{X}]\mathbf{A}^T && \text{Matrice di varianze e covarianze} \end{aligned}$$

In particolare se  $k = 1$  allora  $\mathbf{A}$  è un vettore riga  $\mathbf{b}^T$ ,  $c$  è uno scalare e  $\mathbf{Y}$  è una v.c. semplice  $y$  (ossia scalare) e si ha:

$$y = \mathbf{b}^T \mathbf{X} + c$$

e quindi:

$$\begin{aligned} E[y] &= \mathbf{b}^T E[\mathbf{X}] + c = \\ &= b_1\mu_1 + b_2\mu_2 + \dots + b_p\mu_p + c \\ V[y] &= \mathbf{b}^T V[\mathbf{X}] \mathbf{b} = \\ &= b_1^2\sigma_1^2 + b_2^2\sigma_2^2 + \dots + b_i^2\sigma_i^2 + \dots \\ &+ \dots + b_p^2\sigma_p^2 + 2b_1b_2\sigma_{12} + \dots + 2b_ib_j\sigma_{ij} + \dots + 2b_{p-1}b_p\sigma_{p-1,p} \end{aligned}$$

Formule più complesse valgono per i momenti multivariati di ordine superiore al secondo, ma è sempre possibile ricavare tutti i momenti (multivariati) di grado  $m$  di  $\mathbf{Y}$ , sia centrali che non centrali, a partire dalla conoscenza della matrice di trasformazione  $\mathbf{A}$  e dei momenti multivariati di grado  $1, 2, \dots, m$  di  $\mathbf{X}$ .

**Esempio 4.1** Si hanno  $n$  variabili casuali  $X_i$  normali indipendenti con  $E[X_i] = \mu_i$  e  $V[X_i] = \sigma_i^2$ ,

Quali sono i primi due momenti della nuova variabile aleatoria  $Z$  definita dalla relazione:

$$Z = \sum_{i=1}^n X_i^2$$

E' facile vedere che per ogni  $X_i$  si ha:

$$\begin{aligned} E[X_i^2] &= \mu_i^2 + \sigma_i^2, \\ V[X_i^2] &= E[X_i^4] - (E[X_i^2])^2 = \\ &= \mu_i^4 + 6\sigma_i^2\mu_i^2 + 3\sigma_i^4 - (\mu_i^2 + \sigma_i^2)^2 \\ &= 2(\sigma_i^4 + 2\sigma_i^2\mu_i^2) \end{aligned}$$

ricordando le proprietà dei momenti della normale.

Infine si ha il risultato richiesto:

$$\begin{aligned} E[Z] &= \sum_{i=1}^n (\mu_i^2 + \sigma_i^2); \\ V[Z] &= 2 \sum_{i=1}^n (\sigma_i^4 + 2\sigma_i^2\mu_i^2). \end{aligned}$$

#### 4.2.1 Costruzione di variabili correlate

Un metodo semplice per la costruzione di  $p$  variabili aleatorie  $Y_j$ ,  $j = 1, 2, \dots, p$  correlate a partire da  $p+1$  variabili aleatorie indipendenti  $X_j$  ( $j = 0, 1, \dots, p$ ), è quello di sommare la componente  $X_0$  a tutte le altre  $p$  componenti  $X_j$ ,  $j = 1, 2, \dots, p$ . In dettaglio otteniamo ora un nuovo vettore aleatorio  $\mathbf{Y}$  a  $p$  componenti correlate, ponendo:

$$\begin{pmatrix} Y_1 = X_0 + X_1 \\ \vdots \\ Y_j = X_0 + X_j \\ \vdots \\ Y_p = X_0 + X_p \end{pmatrix} \quad X_i \perp\!\!\!\perp X_j \quad (i, j = 0, 1, \dots, p; i \neq j)$$

In pratica la componente  $X_0$  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  $\mathbf{Y}$  (anche considerando  $\mathbf{X}$  dotato di densità, spesso è complicato integrare rispetto a  $\mathbf{X}_0$  nella densità congiunta di  $X_0, X_1, \dots, X_p$ ).

**Esempio 4.2** 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  $\mathbf{Y}$ .

$$\begin{aligned} V[Y_j] &= V[X_0] + V[X_j] \\ \text{Cov}(Y_j, Y_k) &= \text{Cov}(X_0 + X_j, X_0 + X_k) \\ &= V[X_0] \quad \text{data l'indipendenza delle } X \\ \text{corr}(Y_j, Y_k) &= \frac{V[X_0]}{\sqrt{(V[X_0] + V[X_j])(V[X_0] + V[X_k])}} \end{aligned}$$

e quindi per la correlazione fra due generiche componenti si ha (dividendo numeratore e denominatore per  $V[X_0]$ ):

$$\text{corr}(Y_j, Y_k) = \frac{1}{\sqrt{(1 + \frac{V[X_j]}{V[X_0]})(1 + \frac{V[X_k]}{V[X_0]})}}$$

La correlazione fra le componenti di  $\mathbf{Y}$  è quindi una funzione crescente di  $V[X_0]$ .

#### 4.2.2 Funzione caratteristica di una trasformata lineare di un vettore aleatorio

Se  $\mathbf{Y} = \mathbf{A}\mathbf{X} + c$

Allora la funzione caratteristica di  $\mathbf{Y}$  si ricava da quella di  $\mathbf{X}$  mediante la relazione:

$$\phi_{\mathbf{Y}}(\mathbf{t}) = \exp[i\mathbf{t}^T c] \phi_{\mathbf{X}}(\mathbf{A}\mathbf{t})$$

#### Matrici di varianze e covarianze

Una matrice di varianze e covarianze è sempre semidefinita positiva, sia per variabili aleatorie multiple che per variabili statistiche multiple.

livello  
avanza-  
to



Infatti, dato un vettore aleatorio  $\mathbf{X}$ , la varianza di una sua qualsiasi combinazione lineare  $\mathbf{Y} = \mathbf{t}^T \mathbf{X}$  (con  $\mathbf{t} \neq 0$ ) come è noto è data da:

$$V[\mathbf{Y}] = V[\mathbf{t}^T \mathbf{X}] = \mathbf{t}^T V[\mathbf{X}] \mathbf{t};$$

essendo  $V[\mathbf{Y}] \geq 0$ , in quanto una varianza è sempre non negativa, allora:

$$\mathbf{t}^T V[\mathbf{X}] \mathbf{t} \geq 0, \forall \mathbf{t}, \mathbf{t} \neq 0$$

e quindi, secondo la definizione data prima,  $V[\mathbf{X}]$  è una matrice semidefinita positiva; è definita positiva se si esclude il caso di collinearità esatta fra le  $p$  variabili, e quindi se  $V[\mathbf{X}]$  è di rango pieno  $p$  (e quindi  $V[\mathbf{Y}] > 0$ ).

Si può anche dimostrare, ma ometto la dimostrazione, che  $\mathbf{A}$  è semidefinita positiva, se e solo se si può scrivere come:

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

(con  $\mathbf{X}$  qualsiasi, anche rettangolare).

Per esempio se  $\mathbf{X}$  è una matrice di dati ( $n$  osservazioni e  $p$  variabili), e  $\mathbf{Z}$  la matrice degli scarti dalle rispettive medie aritmetiche, allora  $\mathbf{Z}^T \mathbf{Z}$  è la matrice delle devianze e codevianze delle  $p$  variabili che, come si sa, è semidefinita positiva.

Analogamente vengono definite le forme quadratiche definite negative e semidefinite negative.

### 4.2.3 Momento primo di forme quadratiche

Se  $\mathbf{X}$  è un vettore aleatorio dotato dei primi due momenti, per una forma quadratica in  $\mathbf{X}$ , con  $E[\mathbf{X}] = \boldsymbol{\mu}$  si ha:

$$E[\mathbf{X}^T \mathbf{A} \mathbf{X}] = \text{tr}(\mathbf{A} V[\mathbf{X}]) + \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu}$$

livello  
avanza-  
to

### 4.2.4 Momenti di funzioni qualsiasi di vettori aleatori.

In queste righe parlo principalmente di combinazioni lineari di variabili aleatorie, e qualche volta di forme quadratiche; qualche volta però occorre trattare variabili che sono funzioni non lineari di altre variabili e di volerle calcolare, se non la distribuzione esatta, almeno i momenti in forma approssimata.

E' intuitivo che la strada maestra per la soluzione di un simile problema è quello dell'approssimazione attraverso sviluppi in serie, che darà risultati migliori, quanto più la funzioni da trattare siano linearizzabili: è possibile trovare delle approssimazioni per i momenti di funzioni qualsiasi di vettori aleatori con momenti noti, attraverso opportuni sviluppi in serie, possibilmente troncati ai primi termini per evitare formule troppo complesse.

Sia  $\mathbf{g}(\cdot)$  un vettore di  $k$  funzioni reali di  $p$  variabili reali, e si abbia quindi la generica trasformazione di vettori aleatori:

$$\mathbf{Y} = \mathbf{g}(\mathbf{X}),$$

in cui  $\mathbf{X}$  è un vettore aleatorio a  $p$  componenti e  $\mathbf{Y}$  è un vettore aleatorio a  $k$  componenti.

Indichiamo con  $\frac{\partial \mathbf{g}(\mathbf{x})}{\partial \mathbf{x}}$  la matrice  $k \times p$  delle derivate parziali delle componenti di  $\mathbf{g}(\cdot)$  rispetto agli elementi di  $\mathbf{x}$ .<sup>5</sup>

<sup>5</sup>ho ovviamente dato per assunta l'esistenza delle derivate opportune

Sviluppando  $\mathbf{g}(\cdot)$  in serie di Taylor troncata al primo termine attorno al vettore speranza matematica di  $\mathbf{X}$ , ossia  $E[\mathbf{X}]$ , si hanno le espressioni più semplici:

$$\mathbf{Y} \approx \mathbf{g}(E[\mathbf{X}]) + \left( \frac{\partial \mathbf{g}(x)}{\partial x} \right)_{x=E[\mathbf{X}]} (\mathbf{X} - E[\mathbf{X}]) \quad (6)$$

( $\frac{\partial \mathbf{g}(\mathbf{x})}{\partial \mathbf{x}}$  è calcolato nel punto  $\mathbf{x} = E[\mathbf{X}]$ ).

Prendendo la speranza matematica di ambo i membri si ha:

$$E[\mathbf{Y}] \approx E[\mathbf{g}(E[\mathbf{X}])] + E \left[ \left( \frac{\partial \mathbf{g}(\mathbf{x})}{\partial \mathbf{x}} \right)_{\mathbf{x}=E[\mathbf{X}]} [\mathbf{X} - E[\mathbf{X}]] \right] = \mathbf{g}(E[\mathbf{X}])$$

(il secondo addendo del secondo membro si annulla perchè  $E[\mathbf{X} - E[\mathbf{X}]] = \mathbf{0}$  e  $\left( \frac{\partial \mathbf{g}(\mathbf{x})}{\partial \mathbf{x}} \right)_{\mathbf{x}=E[\mathbf{X}]}$  è una costante e non una variabile aleatoria.)

Sostituendo nell'espressione precedente (6):

$$\mathbf{Y} \approx E[\mathbf{Y}] + \left( \frac{\partial \mathbf{g}(\mathbf{x})}{\partial \mathbf{x}} \right)_{\mathbf{x}=E[\mathbf{X}]} [\mathbf{X} - E[\mathbf{X}]]$$

per cui:

$$\mathbf{Y} - E[\mathbf{Y}] \approx \left( \frac{\partial \mathbf{g}(\mathbf{x})}{\partial \mathbf{x}} \right)_{\mathbf{x}=E[\mathbf{X}]} [\mathbf{X} - E[\mathbf{X}]]$$

che è una relazione lineare approssimata fra gli scarti dei vettori aleatori. Applicando i teoremi sulle trasformazioni lineari di vettori aleatori (visti in 4.2) si ha:

$$V[\mathbf{Y}] \approx \left( \frac{\partial \mathbf{g}(\mathbf{x})}{\partial \mathbf{x}} \right)_{\mathbf{x}=E[\mathbf{X}]} V[\mathbf{X}] \left( \frac{\partial \mathbf{g}(\mathbf{x})}{\partial \mathbf{x}} \right)_{\mathbf{x}=E[\mathbf{X}]}^T$$

(In tutte le formule precedenti  $\frac{\partial \mathbf{g}}{\partial x}$  è calcolata nel punto  $\mathbf{x} = E[\mathbf{X}]$ )

Nel caso univariato ( $p = k = 1$ )  $\mathbf{g}$  è una funzione di una sola variabile  $g(\cdot)$ :

$$V[y] \approx [g(x)]^2 V[X]$$

$$E[\mathbf{g}(\mathbf{X})] \approx \mathbf{g}(E[\mathbf{X}]) \quad (7)$$

$$V[\mathbf{Y}] \approx \left( \frac{\partial \mathbf{g}(\mathbf{x})}{\partial \mathbf{x}} \right)_{\mathbf{x}=E[\mathbf{X}]} V[\mathbf{X}] \left( \frac{\partial \mathbf{g}(\mathbf{x})}{\partial \mathbf{x}} \right)_{\mathbf{x}=E[\mathbf{X}]}^T \quad (8)$$

#### 4.2.5 Esempio sulla distribuzione F

A chiarimento di queste formule, e solo per dare un'idea del grado di approssimazione in alcuni casi di cui si conosce la soluzione esatta, riporto alcuni esempi, comunque non essenziali per gli argomenti immediatamente successivi. Come esempio si consideri la variabile casuale F di Snedecor, data dal rapporto di due variabili casuali  $\chi^2$  indipendenti, divise per i rispettivi gradi di libertà. Per valutare l'approssimazione fornita dalle formule del paragrafo precedente, appliciamole per ottenere delle espressioni approssimate dei primi due momenti di F, le cui espressioni esatte sono comunque note.

La funzione di trasformazione è:

$$F = \frac{X_1/\nu_1}{X_2/\nu_2},$$

essendo  $X_1$  e  $X_2$  due variabili casuali  $\chi^2$  indipendenti, rispettivamente con  $\nu_1$  e  $\nu_2$  gradi di libertà; quindi si ha in questo esempio  $k = 1$  e  $p = 2$ .

Definendo quindi il vettore aleatorio  $\mathbf{X} = [X_1, X_2^T]$ , si ha per i primi due momenti, come è noto dalle proprietà della variabile  $\chi^2$ :

$$E[\mathbf{X}] = \begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix} \quad V[\mathbf{X}] = \begin{pmatrix} 2\nu_1 & 0 \\ 0 & 2\nu_2 \end{pmatrix}$$

L'approssimazione (del primo ordine) al momento primo di  $F$  è data da:

$$E[F] \approx F_{x=E[\mathbf{X}]} = \left( \frac{X_1/\nu_1}{X_2/\nu_2} \right)_{x=E[\mathbf{X}]} = \frac{\nu_1/\nu_1}{\nu_2/\nu_2} = 1$$

Ricordando le proprietà della variabile  $F$  di Snedecor, sappiamo che il momento primo esatto è dato da:

$$E[F] = \frac{\nu_2}{\nu_2 - 2}$$

l'approssimazione coincide col valore esatto solo al divergere di  $\nu_2$ ; infatti:

$$\lim_{\nu_2 \rightarrow \infty} E[F] = \lim_{\nu_2 \rightarrow \infty} \frac{\nu_2}{\nu_2 - 2} = 1.$$

Passando ora al calcolo dell'approssimazione alla varianza di  $F$  occorre valutare il gradiente di  $F$  (rispetto a  $\mathbf{X}$ ) nel punto  $E[\mathbf{X}]$ :

$$\left( \frac{\partial F}{\partial x} \right)_{x=E[\mathbf{X}]} = \left( \frac{\nu_2}{\nu_1 X_2}, \frac{-\nu_2 X_1}{\nu_1 X_2^2} \right)_{\mathbf{X}=\{\nu_1, \nu_2\}^T}^T = \left( \frac{1}{\nu_1}, \frac{-1}{\nu_2} \right)^T$$

ed infine sostituire nella formula:

$$\begin{aligned} V[F] &\approx \left( \frac{\partial F}{\partial x} \right)_{x=E[\mathbf{X}]}^T V[\mathbf{X}] \left( \frac{\partial F}{\partial x} \right)_{x=E[\mathbf{X}]} = \\ &= \left( \frac{1}{\nu_1}, \frac{-1}{\nu_2} \right)^T \begin{pmatrix} 2\nu_1 & 0 \\ 0 & 2\nu_2 \end{pmatrix} \left( \frac{1}{\nu_1}, \frac{-1}{\nu_2} \right) = \\ &= \frac{2}{\nu_1} + \frac{2}{\nu_2}. \end{aligned}$$

Sappiamo che la varianza esatta della  $F$  di Snedecor è data da:

$$V[F] = \frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 4)(\nu_2 - 2)^2}$$

E' facile vedere che il rapporto fra l'approssimazione ed il valore esatto della varianza di  $F$  tende a 1 al divergere di  $\nu_1$  e  $\nu_2$ ; infatti è facile vedere che:

$$\lim_{\nu_1, \nu_2 \rightarrow \infty} \frac{\frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 4)(\nu_2 - 2)^2}}{\frac{2}{\nu_1} + \frac{2}{\nu_2}} = 1.$$

#### 4.2.6 Esempio sulla distribuzione Beta

Come altro esempio consideriamo la variabile casuale Beta, funzione di due variabili gamma indipendenti secondo la funzione di trasformazione:

$$B(X_1, X_2) = \frac{X_1}{X_1 + X_2},$$

essendo  $X_1$  e  $X_2$  due variabili casuali gamma indipendenti, con parametri di scala unitari e parametri di forma rispettivamente  $\alpha$  e  $\beta$ ; quindi si ha

anche in questo esempio  $k = 1$  e  $p = 2$ . Definendo quindi il vettore aleatorio  $\mathbf{X} = \{X_1, X_2\}^T$ , si ha per i primi due momenti, come è noto dalle proprietà della variabile gamma:

$$\mathbb{E}[\mathbf{X}] = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \quad \mathbb{V}[\mathbf{X}] = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}$$

L'approssimazione (del primo ordine) al momento primo di  $B(X_1, X_2)$  è data da:

$$\mathbb{E}[B(X_1, X_2)] \approx B(X_1, X_2)_{\mathbf{X}=\mathbb{E}[\mathbf{X}]} = \left( \frac{X_1}{X_1 + X_2} \right)_{\mathbf{X}=(\alpha, \beta)^T} = \frac{\alpha}{\alpha + \beta}$$

Ricordando le proprietà della variabile B, vediamo che questa approssimazione coincide con il valore esatto:

$$\mathbb{E}[B(X_1, X_2)] = \frac{\alpha}{\alpha + \beta}.$$

Per quanto riguarda la varianza, il gradiente di B calcolato in corrispondenza del valore atteso è dato da:

$$\left( \frac{\partial B}{\partial \mathbf{X}} \right)_{\mathbf{X}=\mathbb{E}[\mathbf{X}]} = \left( \frac{X_2}{(X_1 + X_2)^2}, \frac{-X_1}{(X_1 + X_2)^2} \right)^T_{\mathbf{X}=\{\alpha, \beta\}^T} = \left( \frac{\beta}{(\alpha + \beta)^2}, \frac{-\alpha}{(\alpha + \beta)^2} \right)^T$$

e infine sostituendo nella formula:

$$\begin{aligned} \mathbb{V}[B] &= \left( \frac{\partial B}{\partial x} \right)_{x=\mathbb{E}[\mathbf{X}]}^T \mathbb{V}[\mathbf{X}] \left( \frac{\partial B}{\partial x} \right)_{x=\mathbb{E}[\mathbf{X}]} = \\ &= \left( \frac{\beta}{(\alpha + \beta)^2}, \frac{-\alpha}{(\alpha + \beta)^2} \right)^T \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} \left( \frac{\beta}{(\alpha + \beta)^2}, \frac{-\alpha}{(\alpha + \beta)^2} \right) = \\ &\quad \text{(dopo alcune semplificazioni elementari)} \\ &= \frac{\alpha\beta}{(\alpha + \beta)^3}. \end{aligned}$$

Ricordando adesso che la varianza esatta della variabile Beta è data da:

$$\mathbb{V}[B] = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

Stavolta il rapporto fra l'approssimazione ed il valore esatto della varianza di B è dato da:

$$\frac{V_{\text{appr}}(B)}{\mathbb{V}[B]} = \frac{(\alpha + \beta + 1)}{(\alpha + \beta)}$$

si vede facilmente che questa quantità tende a 1 al divergere di  $\alpha$  oppure di  $\beta$  (mentre per il rapporto F occorre la divergenza di entrambi i parametri).

#### 4.2.7 Esempio su variabili osservate

[esempio da fare con file esempioBMP1 Rmd](#)

Un altro esempio è tratto da variabili statistiche osservate: su un insieme di 1432 bambini sono state rilevate le variabili altezza e peso. Il vettore delle medie e la matrice di varianze e covarianze empiriche sono riportate di seguito:

Variabile	Media	Varianza
ALTEZZA (Metri)	1,5192	0,0103
PESO (Kilogrammi)	44,9909	115,6358

La matrice di varianza e covarianza delle variabili peso e altezza è data da:

$$V[\mathbf{X}] = \begin{pmatrix} 115,6358 & 0,7851 \\ 0,7851 & 0,0103 \end{pmatrix}$$

Su questi 1432 soggetti viene calcolata la nuova variabile BMI (Body Mass Index), data da:

$$BMI = \frac{Peso^2}{Altezza}$$

Questa variabile è impiegata in campo biomedico come indicatore per valutare il grado di adiposità di un soggetto.<sup>6</sup>

Se vogliamo una valutazione approssimata della media di BMI, senza calcolare materialmente i valori sui 1432 soggetti (magari perchè non disponiamo delle singole osservazioni), ma basandoci sui momenti delle variabili altezza e peso otteniamo:

$$M(BMI) \approx \frac{M(Peso)}{(M(Altezza))^2} = \frac{44,9909}{1,5192^2} = 19,4937$$

Per quanto riguarda la varianza si ha (indicando con  $X_1$  la variabile Peso e con  $X_2$  la variabile Altezza) per il gradiente di BMI:

$$BMI = \frac{X_1}{X_2^2}$$

$$\begin{aligned} \left( \frac{\partial BMI}{\partial x} \right)_{x=E[\mathbf{X}]} &= \left( \frac{1}{X_2^2}, \frac{-2X_1}{X_2^3} \right)_{x=E[\mathbf{X}]}^T = \left( \frac{1}{1,5192^2}, \frac{-2 \times 44,9909}{1,5192^3} \right)^T \quad (9) \\ &= (0,4333; -25,6631)^T. \quad (10) \end{aligned}$$

Quindi, sostituendo nella relazione:

$$\begin{aligned} V[BMI] &\approx \left( \frac{\partial BMI}{\partial x} \right)_{x=E[\mathbf{X}]}^T V[\mathbf{X}] \left( \frac{\partial BMI}{\partial x} \right)_{x=E[\mathbf{X}]} = \\ &\quad \text{si ottiene il valore approssimato della varianza di BMI:} \\ V[BMI] &\approx \{0,4333; -25,6631\} \begin{pmatrix} 115,6358 & 0,7851 \\ 0,7851 & 0,0103 \end{pmatrix} \{0,4333; -25,6631\}^T \\ &= 11,0337 \end{aligned}$$

Il grado di validità di queste approssimazioni può essere verificato confrontando con i valori esatti dei primi due momenti di BMI calcolati sui 1432 valori trasformati a partire dai valori singoli originari:

<sup>6</sup>un adulto non dovrebbe superare un valore di 25

BMI= (Kg/ $mt^2$ )		
Media	Varianza	
19,3103	10,4356	Valori esatti
19,4937	11,0337	Valori approssimati

Ovviamente l'utilità di tali formule approssimate si ha quando non è possibile calcolare i momenti esatti (nel caso di variabili aleatorie) o se non sono disponibili i dati relativi alle singole osservazioni, per il calcolo dei valori trasformati, ma solo i primi due momenti delle variabili originarie.

## 5    Analisi delle componenti principali (ACP), solo cenni

Questa parte ha molto in comune da un punto di vista tecnico con l'analisi delle componenti principali per variabili statistiche multiple osservate.

In effetti, l'analisi in componenti principali per vettori aleatori assume un significato molto forte se il vettore aleatorio segue una distribuzione normale multivariata. Maggiori dettagli nella parte sulla normale multivariata e nelle parti sulle variabili statistiche osservate.

Sia  $\mathbf{X}$  un vettore aleatorio di  $p$  componenti dotato di momenti multivariati di primo e secondo ordine:

$$\left\{ \begin{array}{l} E[\mathbf{X}] = \mathbf{0} \quad (\text{valore atteso nullo}) \\ V[\mathbf{X}] = \mathbf{\Sigma} \end{array} \right. ;$$

Se le variabili sono standardizzate, cosa consigliabile per variabili statistiche osservate, allora  $\mathbf{\Sigma}$  è la matrice di correlazione.

Si vuole trovare una nuova variabile casuale  $Z$  (unidimensionale), combinazione lineare di  $\mathbf{X}$  , che abbia la massima varianza possibile <sup>7</sup>, ossia si cerca un vettore di coefficienti  $\mathbf{y}$  tali che:

$$Z = \mathbf{y}^T \mathbf{X} \quad \text{ha varianza massima}$$

$$\text{col vincolo} \quad \mathbf{y}^T \mathbf{y} = 1$$

( $\mathbf{y}$  è un vettore normalizzato; il vincolo sui coefficienti è necessario, altrimenti sarebbe possibile trovare combinazioni di varianza grande a piacere).

---

<sup>7</sup>ma qualche volta cercheremo la minima varianza!

### Ricerca della prima componente principale

Occorre massimizzare rispetto a  $\mathbf{y}$  la varianza di  $Z = \mathbf{y}^T \mathbf{X}$ :

$$V[Z] = \mathbf{y}^T \Sigma \mathbf{y}$$

col vincolo che  $\mathbf{y}$  sia un vettore di lunghezza 1:

$$\mathbf{y}^T \mathbf{y} = 1$$

Il lagrangiano per questo problema è dato da:

$$\mathcal{L}(\mathbf{y}, \lambda) = \mathbf{y}^T \Sigma \mathbf{y} - \lambda \mathbf{y}^T \mathbf{y} + \lambda$$

Derivando rispetto a  $\mathbf{y}$  si ha:

$$2\Sigma \mathbf{y} - 2\lambda \mathbf{y} = \mathbf{0},$$

e quindi:

$$\Sigma \mathbf{y} = \lambda \mathbf{y}.$$

La soluzione  $\mathbf{y}$  è dunque fornita dagli autovettori di  $\Sigma$

Per stabilire quale autovalore fornisce il massimo della funzione obiettivo, premoltiplichiamo nell'ultima equazione ambo i membri per  $\mathbf{y}^T$ :

$$\mathbf{y}^T \Sigma \mathbf{y} = \mathbf{y}^T \lambda \mathbf{y}$$

Da questa uguaglianza vediamo che:

- il primo membro è uguale a  $V[Z]$ ;
- Il secondo membro è uguale a  $\lambda$ , per soddisfare il vincolo  $\mathbf{y}^T \mathbf{y} = 1$ ;

In definitiva si ha:  $V[Z] = \lambda$ , per cui l'ottimo si ha in corrispondenza del massimo autovalore di  $\Sigma$ . Pertanto la soluzione ottima  $\mathbf{y}$  è data dall'autovettore  $\boldsymbol{\gamma}_1$  corrispondente al 1° autovalore  $\lambda_1$  di  $V[\mathbf{X}]$

La nuova variabile  $Z$  di varianza massima è dunque data da:

$$Z_1 = \boldsymbol{\gamma}_1^T \mathbf{X}.$$

Per comodità indico questa nuova variabile con  $Z_1$  anziché con  $Z$ .

Se ora vogliamo trovare una nuova variabile semplice  $Z_2$ , ancora combinazione lineare di  $\mathbf{X}$ , che abbia ancora la maggior varianza, ma con l'ulteriore vincolo di non essere correlata con la prima componente trovata  $Z_1$ , dobbiamo impostare un nuovo problema di massimo:

### Ricerca della seconda componente principale

2° Problema di ottimo vincolato:

Occorre massimizzare la varianza di  $Z_2 = \mathbf{y}^T \mathbf{X}$

$$V[Z_2] = \mathbf{y}^T \Sigma \mathbf{y}$$

col 1° vincolo che  $\mathbf{y}$  sia sempre un vettore di lunghezza 1:

$$\mathbf{y}^T \mathbf{y} = 1$$

col 2° vincolo che  $Z_2$  non sia correlata con la prima componente  $Z_1$ :

$$\mathbf{y}^T \Sigma \boldsymbol{\gamma}_1 = 0$$

Il lagrangiano per questo secondo problema è dato da:

$$\mathcal{L}(\mathbf{y}, \lambda) = \mathbf{y}^T \Sigma \mathbf{y} - \lambda \mathbf{y}^T \mathbf{y} + \lambda - \delta \mathbf{y}^T \Sigma \boldsymbol{\gamma}_1$$

Derivando rispetto a  $\mathbf{y}$  si ha:

$$2\Sigma \mathbf{y} - 2\lambda \mathbf{y} - \delta \Sigma \boldsymbol{\gamma}_1 = \mathbf{0}$$

Si può vedere come la soluzione a questo secondo problema è fornita dall'autovettore di  $\Sigma$  corrispondente al secondo autovalore  $\lambda_2$ ; inoltre

$$V[Z_2] = \lambda_2.$$

In definitiva, ripetendo il procedimento fino a giungere a  $Z_p$ , è possibile trovare  $p$  nuove variabili aleatorie, combinazioni lineari di  $\mathbf{X}$ , a due a due non correlate, e tali che ciascuna  $Z_i$  abbia varianza massima subordinatamente al vincolo di non correlazione con le precedenti variabili  $Z_1, Z_2, \dots, Z_{i-1}$  ed al vincolo di unitarietà dei vettori dei coefficienti.

Ciascuna variabile è data da:

$$Z_i = \boldsymbol{\gamma}_i^T \mathbf{X}, \quad \text{con varianza: } V[Z_i] = \lambda_i$$

In definitiva attraverso la matrice  $\mathbf{\Gamma}$  costituita dagli autovettori di  $\Sigma$ , è possibile trovare un nuovo vettore aleatorio  $\mathbf{Z}$ , dato da:

$$\mathbf{Z} = \mathbf{\Gamma}^T \mathbf{X},$$

tale che:  $V[\mathbf{Z}] = \mathbf{\Lambda}$ .

Dal momento che  $\mathbf{\Gamma}^T \mathbf{\Gamma} = \mathbf{I}$ , la trasformazione corrisponde ad una rotazione ortogonale degli assi. I nuovi assi definiti dallo spazio degli autovettori sono detti *assi principali* e le nuove variabili  $\mathbf{Z}$  sono le *componenti principali* di  $\mathbf{X}$ .

componenti  
principali



### Combinazione lineare di *minima* varianza

In effetti, potremmo impostare il problema al contario, al fine di trovare un modo per misurare l'esistenza di vincoli lineari *quasi esatti* fra le variabili.

Dal momento che un vincolo lineare esatto fra variabili aleatorie corrisponde ad una variabile con varianza nulla, potremmo cercare, fra le combinazioni generate da rotazioni ortogonali degli assi, quella di minima varianza, ossia quella che più si avvicina, con i vincoli imposti, ad una combinazione lineare esatta:

la soluzione stavolta è data dall'autovettore  $\gamma_p$  corrispondente a  $\lambda_p$ , autovalore più piccolo.

## 5.1 Significato statistico e probabilistico delle componenti principali

Possiamo ora migliorare l'informazione fornita dal rango di una matrice di varianza e covarianza, sia essa teorica o empirica. Infatti anche se  $V[\mathbf{X}]$  è a rango pieno, se dall'esame della sequenza degli autovalori risulta che il più piccolo degli autovalori è molto vicino a zero (relativamente all'ordine di grandezza degli autovalori stessi), ciò implica che esiste una combinazione lineare delle componenti del vettore aleatorio  $\mathbf{X}$  a varianza molto bassa.

E' interessante notare che in questo caso la varianza generalizzata  $|V[\mathbf{X}]|$  risulterà piccola rispetto a  $tr(V[\mathbf{X}])$ , a conferma del fatto che la varianza generalizzata fornisce delle informazioni non tanto sulla variabilità delle singole componenti, quanto sulla variazione congiunta.

Altre interpretazioni geometriche delle componenti principali si hanno per vettori aleatori normali, in termini di assi degli ellissoidi di equiprobabilità, come descritto nelle pagine sulla normale multivariata.

Un' interpretazione più vicina alla logica della regressione lineare si vedrà quando si introdurrà l'analisi delle componenti principali per variabili statistiche multiple osservate.

In effetti l'analisi in componenti principali viene spesso usata nell'analisi esplorativa di dati, specialmente in presenza di un gran numero di variabili rilevate per cercare di lavorare su poche variabili che spieghino molta variabilità dell'insieme dei dati (ossia quelle definite dai primi autovettori)

L'utilità pratica di questa tecnica sta nella possibilità di attribuire un significato ai vari fattori. Questo aspetto esplorativo sarà per ora tralasciato.

Non affronto per niente in questo corso, l'argomento dell'eventuale ricerca di combinazioni non lineari di variabili aleatorie che ne spieghino buona parte della varianza.

Nel caso di componenti quadratiche il problema è analiticamente affrontabile sebbene computazionalmente più pesante.

## 6 Variabili Statistiche Multiple

In questa breve sezione richiamo la notazione per insiemi di dati multivariati, attraverso le *matrici di dati*.

In questo contesto vengono rivisti i concetti di momento primo e secondo di variabili statistiche multiple, per i quali è utile adottare un simbolismo compatto, e se ne vedrà il significato. La notazione sarà utile quando si introdurranno anche con degli esempi alcune misure di interdipendenza lineare compresa la **correlazione lineare parziale** (nel cap. ??) e l' **analisi in componenti principali** (nel cap. ??) e poi nello studio dei modelli lineari ??

### 6.1 Calcoli statistici in notazione vettoriale

In questa sezione mi limito a sottolineare come alcuni calcoli statistici per una e due variabili possano essere riespressi in notazione vettoriale: tale notazione per calcoli statistici non è particolarmente utile se effettivamente si lavora solo con una o due variabili, ma è indispensabile quando trattiamo variabili statistiche multiple.

### 6.2 Espressione della varianza di una variabile statistica

Se abbiamo un vettore di  $n$  osservazioni  $\mathbf{x}$  (di media  $M(\mathbf{x})$ ) e il corrispondente vettore degli scarti  $\mathbf{z}$  : <sup>8</sup>

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{pmatrix} \quad \mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_i \\ \vdots \\ z_n \end{pmatrix} = \begin{pmatrix} x_1 - M(\mathbf{x}) \\ x_2 - M(\mathbf{x}) \\ \vdots \\ x_i - M(\mathbf{x}) \\ \vdots \\ x_n - M(\mathbf{x}) \end{pmatrix}$$

essendo l' $i$ -esimo scarto:

$$z_i = x_i - M(\mathbf{x}) \quad i = 1, 2, \dots, n$$

E' facile vedere che:

$$nV[\mathbf{x}] = \sum_{i=1}^n (x_i - M(\mathbf{x}))^2 =$$

$$= \sum_{i=1}^n z_i^2 = (z_1, z_2, \dots, z_i, \dots, z_n) \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_i \\ \vdots \\ z_n \end{pmatrix} = \mathbf{z}^T \mathbf{z}$$

e quindi:

---

<sup>8</sup>Anche se superfluo, ricordo che in tutto questo testo i vettori sono sempre colonna, per cui se devo indicare un vettore riga userò un vettore trasposto.

### Espressione vettoriale della varianza

$$V[\mathbf{x}] = \frac{\mathbf{z}^T \mathbf{z}}{n}$$

Per indicare gli scarti di una variabile dalla propria media userò qualche volta la notazione  $\bar{x}$  e qualche volta ricorrerò ad un simbolo specifico, come  $\mathbf{z}$ , opportunamente definito nel testo.

#### 6.2.1 Espressione vettoriale della covarianza

In modo simile, possiamo utilizzare una notazione vettoriale per indicare la **covarianza** fra due variabili statistiche  $\mathbf{x}$  e  $\mathbf{y}$ , per le quali abbiamo  $n$  coppie di osservazioni  $(x_i, y_i)$ ; indichiamo con  $\bar{x}_i$  e  $\bar{y}_i$  gli scarti dalle rispettive medie della  $\mathbf{x}$  e della  $\mathbf{y}$ :

$$\bar{x}_i = x_i - M(\mathbf{x}) \quad \text{e} \quad \bar{y}_i = y_i - M(\mathbf{y}) \quad i = 1, 2, \dots, n$$

e corrispondentemente i vettori degli scarti:

$$\bar{\mathbf{x}}^T = \{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_i, \dots, \bar{x}_n\} \quad \bar{\mathbf{y}}^T = \{\bar{y}_1, \bar{y}_2, \dots, \bar{y}_i, \dots, \bar{y}_n\}$$

Pertanto è immediato vedere come esprimere in notazione vettoriale la covarianza fra le due variabili statistiche  $\mathbf{x}$  e  $\mathbf{y}$ :

$$\begin{aligned} n\text{Cov}[\mathbf{x}, \mathbf{y}] &= \sum_{i=1}^n (x_i - M(\mathbf{x}))(y_i - M(\mathbf{y})) = \sum_{i=1}^n \bar{x}_i \bar{y}_i = \\ &= (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_i, \dots, \bar{x}_n) \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_i \\ \vdots \\ \bar{y}_n \end{pmatrix} = \bar{\mathbf{x}}^T \bar{\mathbf{y}} \end{aligned}$$

ed infine:

$$\text{Cov}[\mathbf{x}, \mathbf{y}] = \frac{\bar{\mathbf{x}}^T \bar{\mathbf{y}}}{n}$$

Ricordo la proprietà per la quale si può esprimere la covarianza senza ricorrere alle somme di scarti che diventa, ancora in notazione vettoriale:

$$\begin{aligned} \text{Cov}[\mathbf{x}, \mathbf{y}] &= \frac{\sum_{i=1}^n (x_i - M(\mathbf{x}))(y_i - M(\mathbf{y}))}{n} = \frac{\sum_{i=1}^n x_i y_i}{n} - M(\mathbf{x})M(\mathbf{y}) = \\ &= \frac{\{x_1, x_2, \dots, x_i, \dots, x_n\} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix}}{n} - M(\mathbf{x})M(\mathbf{y}) \end{aligned}$$

**Espressione vettoriale della covarianza**

$$\text{Cov}[\mathbf{x}, \mathbf{y}] = \frac{\mathbf{x}^T \mathbf{y}}{n} - M(\mathbf{x})M(\mathbf{y})$$

Bozze MARCELLO CHIODI 2020

### 6.2.2 Espressione vettoriale della media aritmetica

E' facile vedere che in notazione matriciale possiamo esprimere anche una media aritmetica, anche se per ora l'utilità della notazione non è grande.

Da ora in poi indicheremo con  $\mathbf{1}_k$  un vettore colonna di  $k$  elementi tutti uguali ad 1:

$$\mathbf{1}_k = \begin{pmatrix} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \quad (k \text{ volte})$$

Con l'introduzione di questo nuovo elemento possiamo scrivere:

#### Espressione vettoriale della media aritmetica

$$M(\mathbf{x}) = \frac{\sum_{i=1}^n x_i}{n} = \frac{\mathbf{x}^T \mathbf{1}_n}{n}$$

Praticamente la moltiplicazione di un vettore riga per un vettore  $\mathbf{1}_n$  ci permette di scrivere una sommatoria semplice in termini di prodotto vettoriale

## 7 Definizione della matrice dei dati

Supponiamo di avere l'informazione relativa a  $n$  unità su cui sono state rilevate  $p$  variabili statistiche *quantitative*.

in questa fase di definizione del simbolismo che adotteremo per un insieme di dati multivariato, non ci preoccuperemo del fatto che queste unità costituiscano una popolazione completa o invece un campione (casuale semplice o di qualsiasi altro tipo): supponiamo che si tratti comunque dell'intera informazione disponibile dall'osservazione, comunque essa sia stata ottenuta.

Quando introdurremo particolari modelli di dipendenza e di interdipendenza, faremo delle opportune assunzioni sul modello generatore dei dati.

**L'informazione completa è per noi costituita da una matrice di dati  $\mathbf{X}_{[n \times p]}$ .**

La matrice  $\mathbf{X}$  ( $n$  righe e  $p$  colonne), di elemento generico  $x_{ij}$  è data dai valori osservati di  $p$  variabili (che per ora supporremo quantitative), per ciascuna delle  $n$  unità statistiche:

$$\mathbf{X}_{[n \times p]} = \begin{pmatrix} X_1 & X_2 & \dots & X_j & \dots & X_p \\ x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1p} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \hline x_{i1} & x_{i2} & \dots & x_{ij} & \dots & x_{ip} \\ \hline \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \hline x_{n1} & x_{n2} & \dots & x_{nj} & \dots & x_{np} \end{pmatrix} \begin{matrix} U_1 \\ \dots \\ \dots \\ \hline U_i \\ \hline \dots \\ \hline U_n \end{matrix}$$

$$\text{Medie} = \{M_1, M_2, \dots, M_j, \dots, M_p\}$$

L'informazione relativa ad una unità  $U_i$  è dunque costituita dalla riga  $i$ -esima delle  $p$  osservazioni relative alle  $p$  variabili:

$$\begin{matrix} i\text{-esima unità (riga)} \\ U_i = \{x_{i1}; x_{i2}; \dots; x_{ij}; \dots; x_{ip}\}^T; \quad i = 1, 2, \dots, n \end{matrix}$$

L'informazione (univariata) relativa alla  $j$ -esima variabile  $X_j$  è contenuta nella  $j$ -esima colonna:

$$\begin{matrix} j\text{-esima variabile (colonna)} \\ X_j = \{x_{1j}; x_{2j}; \dots; x_{ij}; \dots; x_{nj}\}; \quad j = 1, 2, \dots, p \end{matrix}$$

## 7.1 Dati mancanti

Non verranno prese in considerazione in questo momento le problematiche derivanti da matrici di dati incomplete, ossia in cui alcune delle osservazioni  $x_{ij}$  relative ad uno o più casi ed ad una o più variabili sono mancanti.

Alcune di queste problematiche verranno riprese più avanti, in particolare nel corso di esercitazioni e nei laboratori.

Purtroppo, pur sembrando una banalità la gestione dei valori mancanti (*missing data*), è molto complessa e può essere affrontata soltanto con un bagaglio di consocenze teoriche e pratiche molto profondo. Tuttavia, trattandosi di un problema ricorrente nell'analisi di dati reali, fornirò al momento opportuno qualche tecnica elementare per affrontarlo.

E' importante adesso ricordare alcune cose:

- Gli insiemi di dati *completi* sono di solito utilizzati per comodità negli esempi didattici. La realtà ahinoi è più cruda.
- E' improbabile trovare nella realtà insiemi di dati *completi*. I dati mancanti sono sempre in agguato.
- Diffidate di dati grezzi *completi*. E' possibile che qualcuno abbia già stimato in qualche modo i valori mancanti.
- E' importante capire se i dati mancanti sono mancanti per caso (*MAR=missing at random*) oppure in funzione dei valori di qualche variabile (argomento che non verrà approfondito in queste righe).
- R indica i dati mancanti come NA (ossia: not available). Quando si utilizza R (o qualsiasi altro linguaggio) accertarsi sempre se il proprio data frame contiene dati mancanti (per esempio con l'istruzione `is.na()`) e controllare come le diverse funzioni trattano i dati mancanti. .

## 8 I momenti primi e secondi (multivariati) di una variabile statistica multipla

Ritornando alle ordinarie matrici di dati a due vie, che rappresentano le  $n$  rilevazioni di  $p$  variabili, la media aritmetica di ciascuna variabile è data da:

$$M_j = \frac{\sum_{i=1}^n x_{ij}}{n} \quad j = 1, 2, \dots, p$$

Il vettore delle medie è costituito dalle  $p$  medie aritmetiche:

$$M(\mathbf{X}) = \begin{pmatrix} M_1 \\ M_2 \\ \vdots \\ M_j \\ \vdots \\ M_p \end{pmatrix}$$

Se consideriamo una rappresentazione geometrica delle  $n$  unità statistica, la nostra matrice dei dati costituisce l'insieme delle coordinate di  $n$  punti in uno spazio  $p$ -dimensionale.

Il punto di coordinate  $M(\mathbf{X})$  è detto *centroide* dell'insieme multivariato di dati.

E' facile vedere che in notazione matriciale possiamo esprimere  $M(\mathbf{X})$  mediante la relazione:

$$M(\mathbf{X}) = \mathbf{X}^T \mathbf{1}_n / n$$

Abbiamo ancora indicato con  $\mathbf{1}_n$  un vettore colonna di  $n$  elementi tutti uguali ad 1.

$$\mathbf{1}_k^T = (1, \dots, 1, \dots, 1), \quad k \text{ volte}$$

Per i momenti del secondo ordine si ha:

la varianza della singola variabile  $X_j$ :

$$\sigma_j^2 = \sum_{i=1}^n (x_{ij} - M_j)^2 / n \quad j = 1, 2, \dots, p$$

la covarianza fra la variabile  $X_j$  e la variabile  $X_k$ :

$$\sigma_{jk} = \sum_{i=1}^n (x_{ij} - M_j)(x_{ik} - M_k) / n \quad j = 1, 2, \dots, p \quad k = 1, 2, \dots, p$$



E' noto che tali relazioni riguardanti momenti secondi centrali, sono esprimibili in termini dei momenti primi e secondi con origine lo zero:

$$\sigma_j^2 = \sum_{i=1}^n x_{ij}^2/n - M_j^2 \quad j = 1, 2, \dots, p$$

$$\sigma_{jk} = \sum_{i=1}^n x_{ij}x_{ik}/n - M_jM_k \quad j = 1, 2, \dots, p; k = 1, 2, \dots, p;$$

Bozze MARCELLO CHIODI 2020

## 8.1 La matrice di varianze e covarianze

Avendo richiamato la definizione ed il calcolo delle medie, delle varianze e delle covarianze, possiamo definire la matrice di varianze e covarianze:

### Matrice di varianze e covarianze

$$V[\mathbf{X}] = \begin{pmatrix} \sigma_1^2 & \dots & \sigma_{1i} & \dots & \sigma_{1p} \\ \dots & \dots & \dots & \dots & \dots \\ \sigma_{1i} & \dots & \sigma_i^2 & \dots & \sigma_{ip} \\ \dots & \dots & \dots & \dots & \dots \\ \sigma_{1p} & \dots & \sigma_{ip} & \dots & \sigma_p^2 \end{pmatrix}$$

Per gli elementi sulla diagonale principale di  $V[\mathbf{X}]$ , ossia per le varianze delle singole componenti, invece della notazione  $\sigma_{ii}$  si impiega la notazione  $\sigma_i^2$  per uniformità col simbolismo nel caso univariato.

Userò quasi sempre il simbolo  $V(\cdot)$  con l'intesa che se l'argomento è una matrice di dati indica una matrice di varianze e covarianze campionaria; se l'argomento è una variabile statistica semplice allora sarà una varianza campionaria; userò lo stesso simbolo anche per matrici di varianze e covarianze di variabili aleatorie

## 8.2 La matrice di correlazione

Si può definire la matrice di correlazione di elemento generico:

$$r_{ij} = \{R(\mathbf{X})\}_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$$

che, ovviamente, è simmetrica ed ha elementi diagonali tutti uguali ad uno:

matrice  
di corre-  
lazione

### Matrice di correlazione empirica di $p$ variabili statistiche

$$r_{ij} = \{R(\mathbf{X})\}_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$$

$$R(\mathbf{X}) = \begin{pmatrix} 1 & \dots & r_{1i} & \dots & r_{1p} \\ \dots & \dots & \dots & \dots & \dots \\ r_{1i} & \dots & 1 & \dots & r_{ip} \\ \dots & \dots & \dots & \dots & \dots \\ r_{1p} & \dots & r_{ip} & \dots & 1 \end{pmatrix}$$

Misura le correlazioni **lineari** fra le coppie di variabili.  
E' essenziale anche come strumento esplorativo.

**La matrice di correlazione è uguale alla matrice di varianze e covarianze delle corrispondenti variabili standardizzate**

Discutere in aula del significato di  $r$  ed  $r^2$

Indice di interdipendenza (misura simmetrica)

### 8.3 esempio

[esempio da fare con file esempiocor Rmd](#)



Figura 1: matrice di grafici di 4 variabili



Figura 2: Matrice di correlazione delle 4 variabili dell'esempio dei neonati

## 9 La matrice degli scarti

E' utile spesso fare riferimento alla matrice degli scarti  $\mathbf{Z}$ , il cui generico elemento è definito da:

$$z_{ij} = x_{ij} - M_j \quad i = 1, 2, \dots, n \quad j = 1, 2, \dots, p$$

Si faccia attenzione al fatto che lo scarto va effettuato rispetto alla media della colonna corrispondente

$$\mathbf{Z}_{[n \times p]} = \left( \begin{array}{ccc|ccc} x_{11} - M_1 & x_{12} - M_2 & \dots & x_{1j} - M_j & \dots & x_{1p} - M_p \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \hline x_{i1} - M_1 & x_{i2} - M_2 & \dots & x_{ij} - M_j & \dots & x_{ip} - M_p \\ \hline \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \hline x_{n1} - M_1 & x_{n2} - M_2 & \dots & x_{nj} - M_j & \dots & x_{np} - M_p \end{array} \right) \begin{array}{c} \mathbf{U}_1 \\ \dots \\ \dots \\ \hline \mathbf{U}_i \\ \hline \dots \\ \dots \\ \hline \mathbf{U}_n \end{array}$$

Indichiamo ciascuna colonna con  $\mathbf{z}_j$ . Evidentemente le nuove variabili  $Z_j$  risultano a media nulla:

$$M\{Z_1, Z_2, \dots, Z_j, \dots, Z_p\} = \{0, 0, \dots, 0, \dots, 0\} = \mathbf{0}_p^T$$

Adesso possiamo esprimere in modo compatto la generica covarianza  $\sigma_{jk}$  (anzi la **codevianza**) in funzione delle colonne  $\mathbf{z}_j$  e  $\mathbf{z}_k$ :

$$\begin{aligned} n \sigma_{jk} &= \sum_{i=1}^n (x_{ij} - M_j)(x_{ik} - M_k) = \sum_{i=1}^n z_{ij} z_{ik} = \\ &= (z_{1j}, \dots, z_{ij}, \dots, z_{nj}) \begin{pmatrix} z_{1k} \\ \vdots \\ z_{ik} \\ \vdots \\ z_{nk} \end{pmatrix} = \mathbf{z}_j^T \mathbf{z}_k \end{aligned}$$

Con questa posizione di comodo, è facile ora vedere che la matrice di varianze e covarianze  $p \times p$  delle variabili  $X_j$  (o delle variabili  $Z_j$ ) è espressa in forma matriciale compatta:

$$V[\mathbf{X}] = V[\mathbf{Z}] = \mathbf{Z}^T \mathbf{Z} / n$$

Si può anche vedere che:

$$\mathbf{Z} = \mathbf{X} - \mathbf{1}_n M(\mathbf{X})^T = \mathbf{X} - \mathbf{1}_n \mathbf{1}_n^T \mathbf{X} / n = (\mathbf{I} - \mathbf{1}_n \mathbf{1}_n^T / n) \mathbf{X}$$

$$\begin{aligned} V[\mathbf{X}] &= V[\mathbf{Z}] = [\mathbf{X}^T - M(\mathbf{X}) \mathbf{1}_n^T][\mathbf{X} - \mathbf{1}_n M(\mathbf{X})^T] / n = \\ &= \mathbf{X}^T \mathbf{X} / n - M(\mathbf{X}) M(\mathbf{X})^T \end{aligned}$$

ricordando, per l'ultimo passaggio, che:

$$[\mathbf{X}^T - M(\mathbf{X}) \mathbf{1}_n^T][\mathbf{1}_n M(\mathbf{X})^T] / n = 0 \quad \text{e} \quad M(\mathbf{X}) \mathbf{1}_n^T \mathbf{X} / n = M(\mathbf{X}) M(\mathbf{X})^T.$$

Oppure, dalla relazione prima vista:

$$\mathbf{Z} = (\mathbf{I}_n - \mathbf{1}_n \mathbf{1}_n^T / n) \mathbf{X},$$

si ha:

$$\mathbf{Z}^T \mathbf{Z} = \mathbf{X}^T (\mathbf{I}_n - \mathbf{1}_n \mathbf{1}_n^T / n)^T (\mathbf{I}_n - \mathbf{1}_n \mathbf{1}_n^T / n) \mathbf{X};$$

e considerando che la matrice  $(\mathbf{I}_n - \mathbf{1}_n \mathbf{1}_n^T / n)$  è simmetrica e *idempotente*, si ha infine:

$$\begin{aligned} nV[\mathbf{X}] &= nV[\mathbf{Z}] = \mathbf{Z}^T \mathbf{Z} = \mathbf{X}^T (\mathbf{I}_n - \mathbf{1}_n \mathbf{1}_n^T / n)^T (\mathbf{I}_n - \mathbf{1}_n \mathbf{1}_n^T / n) \mathbf{X} = \\ &= \mathbf{X}^T (\mathbf{I}_n - \mathbf{1}_n \mathbf{1}_n^T / n) \mathbf{X}; \end{aligned}$$

Si vedrà a proposito anche l'espressione della devianza residua nell'analisi dei modelli lineari, che è formalmente analoga a questa espressione.

Come si vede, si ottengono risultati già noti nel caso a una e due variabili sui momenti primi e secondi; la notazione matriciale permette di ottenere risultati anche mnemonicamente simili a quelli più che noti del caso univariato.

E' appena il caso di osservare che mentre la notazione matriciale fornisce espressioni compatte ed è inoltre implementabile facilmente negli ambienti di programmazione che supportano operazioni matriciali, difficilmente fornisce gli algoritmi più efficienti per il calcolo dei momenti multivariati.

Raccomando sempre in R di utilizzare le funzioni già disponibili per `data.frame` o per matrici quali `cov()`, `cor()`

## 9.1 I momenti di combinazioni lineari di variabili statistiche multiple

Per i momenti di combinazioni lineari di una variabile multipla valgono ovviamente relazioni del tutto analoghe a quelle valide per combinazioni lineari di vettori di variabili aleatorie:

Costruiamo una nuova variabile statistica a  $k$  componenti, mediante una qualsiasi trasformazione lineare delle variabili  $X_j$ , colonne della matrice dei dati  $\mathbf{X}$  :

$$\mathbf{Y} = \mathbf{X}\mathbf{A}^T + \mathbf{1}_n\mathbf{c}^T$$

La matrice  $\mathbf{A}_{[k \times p]}$  ha  $k$  righe e  $p$  colonne e per il resto è qualsiasi, nel senso che il suo rango può anche essere inferiore a  $\min(k, p)$ .

Il vettore  $\mathbf{c}_{[k \times 1]}$  ha  $k$  elementi.

La nuova matrice di dati  $\mathbf{Y}$  ha  $n$  righe e  $k$  colonne. Con semplici passaggi si vede come data la matrice  $\mathbf{A}$  e il vettore  $\mathbf{c}$  è possibile ottenere tutti i momenti di  $\mathbf{Y}$  in funzione di quelli di  $\mathbf{X}$  :

$$M[\mathbf{Y}] = M[\mathbf{X}] \mathbf{A}^T + \mathbf{c} \quad (11)$$

$$V[\mathbf{Y}] = V[\mathbf{X}\mathbf{A}^T + \mathbf{1}_n\mathbf{c}^T] = \mathbf{A}V[\mathbf{X}]\mathbf{A}^T \quad (12)$$

---

### argomenti avanzati e complementi (non in programma)

Formule più complesse valgono per i momenti multivariati di ordine superiore al secondo, ma è possibile ricavare tutti i momenti (multivariati) di grado  $m$  di  $\mathbf{Y}$ , sia centrali che non centrali, a partire dalla conoscenza della matrice di trasformazione  $\mathbf{A}$  e dei momenti multivariati di grado  $1, 2, \dots, m$  di  $\mathbf{X}$ .

Come per le variabili statistiche semplici i momenti multivariati di ordine superiore al secondo forniscono degli indici di forma multivariati, degli indicatori di allontanamento dalla multinormalità, indici di non linearità delle regressioni e di eteroscedasticità, ma non ne farò uso in questo corso.

---

### I momenti di combinazioni lineari di variabili statistiche multiple

$$\mathbf{Y} = \mathbf{X}\mathbf{A}^T + \mathbf{1}_n\mathbf{c}^T$$

$$M(\mathbf{X}\mathbf{A}^T + \mathbf{1}_n\mathbf{c}^T) = M(\mathbf{X})\mathbf{A}^T + \mathbf{c} \quad \text{Vettore delle medie}$$

$$V[\mathbf{X}\mathbf{A}^T + \mathbf{1}_n\mathbf{c}^T] = \mathbf{A}V[\mathbf{X}]\mathbf{A}^T \quad \text{Matrice di varianze e covarianze}$$

In particolare se  $k = 1$  allora  $\mathbf{A}$  è un vettore riga  $\mathbf{b}^T$ ,  $\mathbf{c}$  è uno scalare  $c$  e  $\mathbf{Y}$  è una v.c. semplice (ossia scalare) e si ha:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + c$$

e quindi:

$$M(\mathbf{y}) = \mathbf{b}^T M(\mathbf{X}) + c = b_1\mu_1 + b_2\mu_2 + \dots + b_p\mu_p + c$$

$$\begin{aligned} V[\mathbf{y}] &= \mathbf{b}^T V[(\mathbf{X})] \mathbf{b} = \\ &= b_1^2\sigma_1^2 + b_2^2\sigma_2^2 + \dots + b_i^2\sigma_i^2 + \dots + b_p^2\sigma_p^2 + \end{aligned}$$

$$+2b_1b_2\sigma_{12} + \dots + 2b_ib_j\sigma_{ij} + \dots + 2b_{p-1}b_p\sigma_{p-1,p}$$

Dall'espressione precedente si ricava immediatamente una proprietà che sarà molto utile:

### Forme quadratiche e combinazioni lineari

Una forma quadratica con matrice di coefficienti data da una matrice di varianze e covarianze  $V[\mathbf{X}]$  esprime sempre la varianza di una combinazione lineare delle  $\mathbf{X}$  :

$$\mathbf{b}^T V[\mathbf{X}] \mathbf{b} = V[\mathbf{X}\mathbf{b}]$$

### Positività delle matrici di varianza e covarianza

Una matrice di varianze e covarianze è quindi sempre **semidefinita positiva** .  
essendo  $V[\mathbf{Y}] \geq 0$ , in quanto una varianza è sempre non negativa, allora:

$$\mathbf{t}^T V[\mathbf{X}] \mathbf{t} = V[\mathbf{X}\mathbf{t}] \geq 0, \forall \mathbf{t}, \mathbf{t} \neq 0$$



## 9.2 Rango della matrice di varianza e covarianza ( $n \geq p$ )

- Se una variabile statistica è combinazione lineare delle altre  $p-1$ , allora il rango della matrice di varianza e covarianza di  $\mathbf{X}$  risulta uguale a  $p-1$ ; (con  $n \geq p$ )
- in generale il rango di  $V[\mathbf{X}]$  risulta uguale a  $p-v$  se  $v$  componenti sono ottenute attraverso combinazioni lineari (indipendenti) degli elementi di  $\mathbf{X}$ .
- il rango di  $V[\mathbf{X}]$  risulta uguale esattamente a  $p$  (ossia a rango pieno) se e solo se le componenti di  $\mathbf{X}$  sono linearmente indipendenti (con  $n \geq p$ ).

Bozze MARCELLO CHIODI 2020