

# Introduzione

## Contatti

Prof. Antonio Calzagni  
(antonio.calzagni@unipd.it)

Becheca: unipd.it/acalzagni

Ricev: Giovedì 10:30-12:30

Dott.ssa Ambra Perugini  
(ambra.perugini@phd.unipd.it)

## Organizzazione del corso

50% didattica frontale

50% laboratorio con R

20% didattica interattiva

## Contenuti

(A) Richiami e fondamenti

(B) Analisi descrittiva univariata

(C) CFA + invarianza

(D) Complementi (clustering, PCA)

## Materiale didattico

① Corbetta P. Metodi di analisi multivariata per le scienze sociali. il Mulino

• Economia: QUANT. 6.2.2008

• Statistica: SA. 442-247

② Pace, Salvati. Introduzione alla stat. (Vol. 1). Cedam

• Statistica: DID. X. 3-44

③ Fabbris. Statistica Multivariata. McGraw-Hill

• Statistica: DID. X. 3-47

• Psicologia: 22.428, 5-40

④ Barbamelli, Natali. I test psicologici

• Psicologia: BD. 22.205, 5-00

⑤ Appunti, slides, scripts di R

## Esame

① Prova scritta (60 minuti), 5 domande di analisi dei dati

② Ciascun quesito  $q_j$  vale fino a 1 punto ( $q_j \in [0, 1]$ )

③ Voto finale:  $(\sum_{j=1}^5 q_j) \times 6.2 = \text{voto finale}$

3/5 necessario per 18/30

④ Guida all'Esame → Moodle



- Corbett, App. B
- Pace e Salvem, capp. 12-13

## Definizioni iniziali

Una disposizione di  $n \times m$  oggetti in forma tabellare  $(X_{n \times m})$  è detta matrice.

$$X_{n \times m} = \begin{pmatrix} x_{11} & \dots & x_{1j} & \dots & x_{1m} \\ \vdots & & \vdots & & \vdots \\ x_{i1} & \dots & x_{ij} & \dots & x_{im} \\ \vdots & & \vdots & & \vdots \\ x_{n1} & \dots & x_{nj} & \dots & x_{nm} \end{pmatrix}$$

$(i, j)$

$i \in \{1, 2, \dots, n\}$   
 $j \in \{1, 2, \dots, m\}$

indici per scorrere gli elementi della matrice/vettore.

Consideriamo la prima colonna ( $j=1$ ):

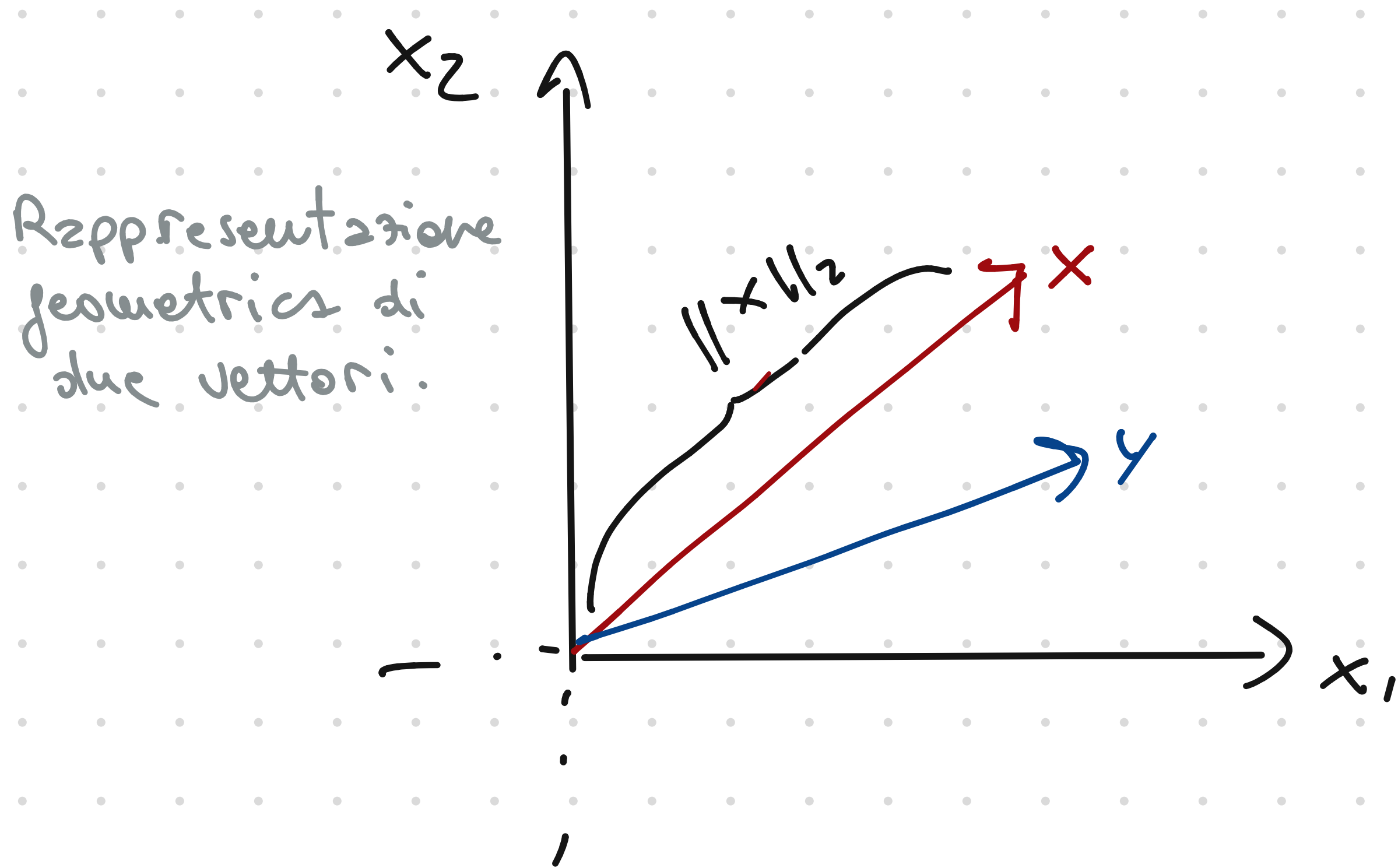
$$X_{n \times 1} = (x_1, \dots, x_i, \dots, x_n)$$

$X_{n \times 1}$  è un vettore colonna (di  $n$  elementi).

Consideriamo la  $n$ -esima riga ( $i=n$ ):

$$X_{1 \times m} = (x_1, \dots, x_j, \dots, x_m)$$

$X_{1 \times m}$  è un vettore riga (di  $m$  elementi)



$$x = (x_1, x_2)$$

$$y = (y_1, y_2)$$

Consideriamo un vettore

$$x \in \mathbb{R}^m$$

$$x = (x_1, \dots, x_i, \dots, x_m)$$

La notazione  $x \in \mathbb{R}^m$  indica che il vettore  $x$  è  $m$ -dimensionale, ha cioè  $m$  elementi di tipo reale.

La notazione  $X \in \mathbb{R}^{n \times m}$  indica che la matrice  $X$  ha  $n$  righe e  $m$  colonne ed è popolata da  $n \times m$  numeri reali.

L'insieme in cui  $x, X, x$  possono 'vivere' può essere  $\mathbb{R}, \mathbb{N}, \mathbb{Q}, \dots$ . Noi ci concentreremo esclusivamente su  $\mathbb{R}$ .

Chiameremo norma la quantità:

$$\|x\|_p = \left( \sum_{i=1}^m |x_i|^p \right)^{1/p}$$

$$p=1: \|x\|_1 = \sum_{i=1}^m |x_i|$$

$$p=2: \|x\|_2 = \left( \sum_{i=1}^m x_i^2 \right)^{1/2} = \sqrt{\sum_{i=1}^m x_i^2} \quad (\text{distanza euclidea dall'origine}) ; \|x-y\|_2 \quad (\text{distanza euclidea tra } x \text{ e } y)$$

Caso che utilizzeremo in questo corso

$$p=\infty: \|x\|_\infty = \max_{i \in I} x_i, \quad I = \{1, \dots, m\} \quad (\text{insieme degli indici})$$

$$p=0: \|x\|_0 = \text{Card}(x) \quad \text{'cardinalità' di } x \text{ anche denotata come } \#x$$

Considerando il caso  $p=2$ , otteniamo il seguente risultato notevole utile per il calcolo:

$$(\text{devianza}) \sum_{i=1}^m (x_i - \bar{x})^2 = \|x - \bar{x}\|_2^2 = \|x\|^2 + \|\bar{x}\|^2 - 2 \sum_{i=1}^m x_i \bar{x} \rightarrow \text{Prodotto Scalare}$$

$$(\text{varianza}) \frac{1}{m} \sum_{i=1}^m (x_i - \bar{x})^2 = \frac{1}{m} \|x - \bar{x}\|_2^2$$

media( $x$ )

Nota: Un vettore è un vettore di norma unitaria, ossia  $\tilde{x} = \frac{x}{\|x\|}$



Due vettori  $x \in \mathbb{R}^m$ ,  $y \in \mathbb{R}^m$  si dicono ortogonali ( $x \perp y$ ) se:

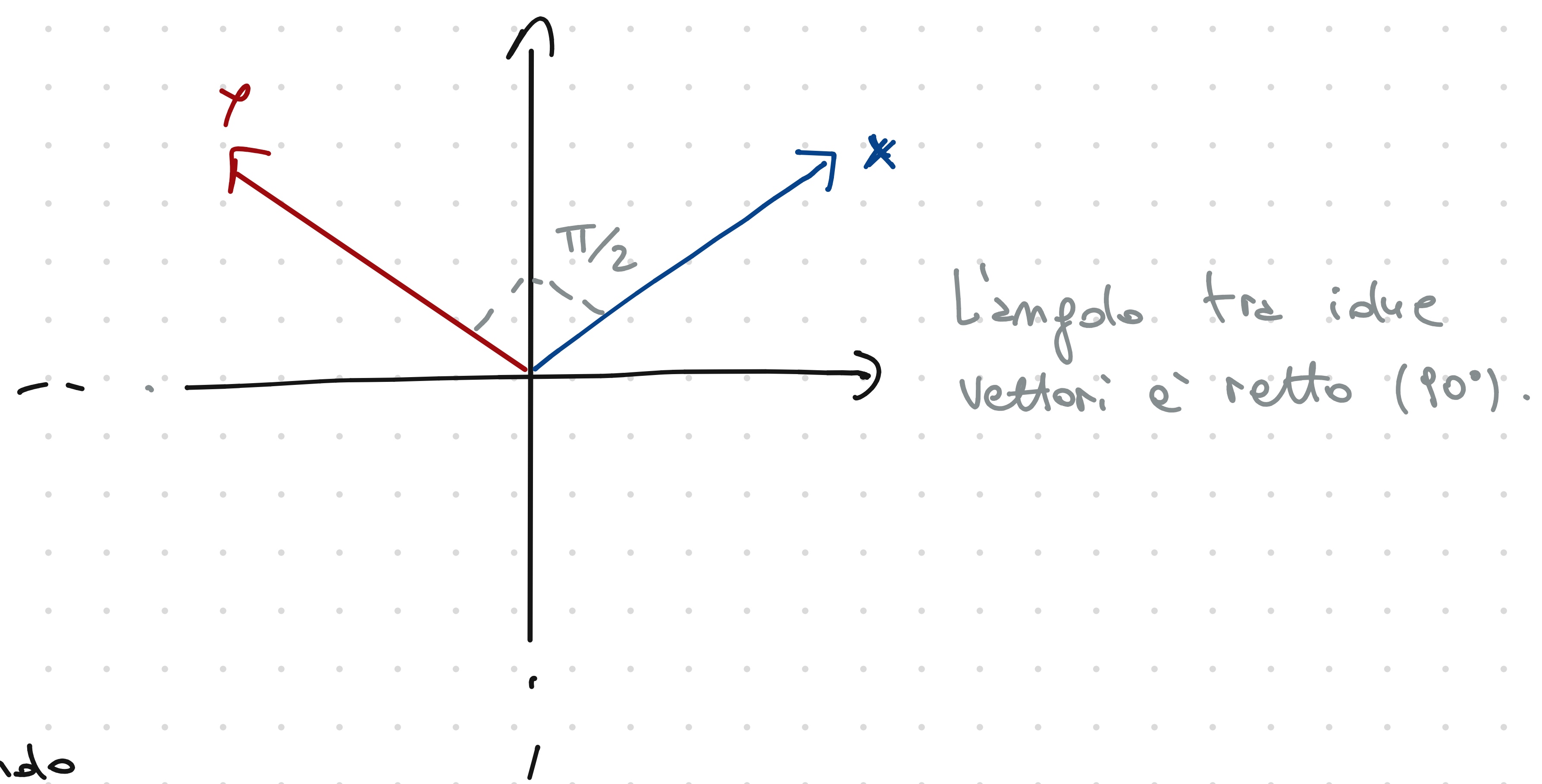
$$\left( \sum_{i=1}^m x_i \cdot y_i \right) = 0 \quad (\text{prodotto scalare nullo})$$

Esempio:

$$y = (-\sqrt{2}, \sqrt{2})$$

$$x = (\sqrt{2}, \sqrt{2})$$

$$\begin{aligned} \sum_{i=1}^m x_i y_i &= -\sqrt{2} \sqrt{2} + \sqrt{2} \sqrt{2} \\ &= -2 + 2 \\ &= 0 \end{aligned}$$



Il prodotto scalare può essere altresì riscritto utilizzando l'angolo tra  $x$  e  $y$ :

$$\underbrace{\langle x, y \rangle}_{\text{notazione del prodotto scalare}} = \sum_{i=1}^m x_i \cdot y_i = \|x\| \cdot \|y\| \cos \theta$$

con  $\theta = \arccos(\langle x, y \rangle / (\|x\| \cdot \|y\|))$ . Note:  $\theta$  è espresso in radianti.

Nell'esempio,  $\theta = \arccos(0) \approx 1.57 = \pi/2$  e  $\|x\| \cdot \|y\| \cdot \underbrace{\cos(\arccos(0))}_{=0} = 0$ .

Applet online:

<https://fslstad.com/dotproduct>

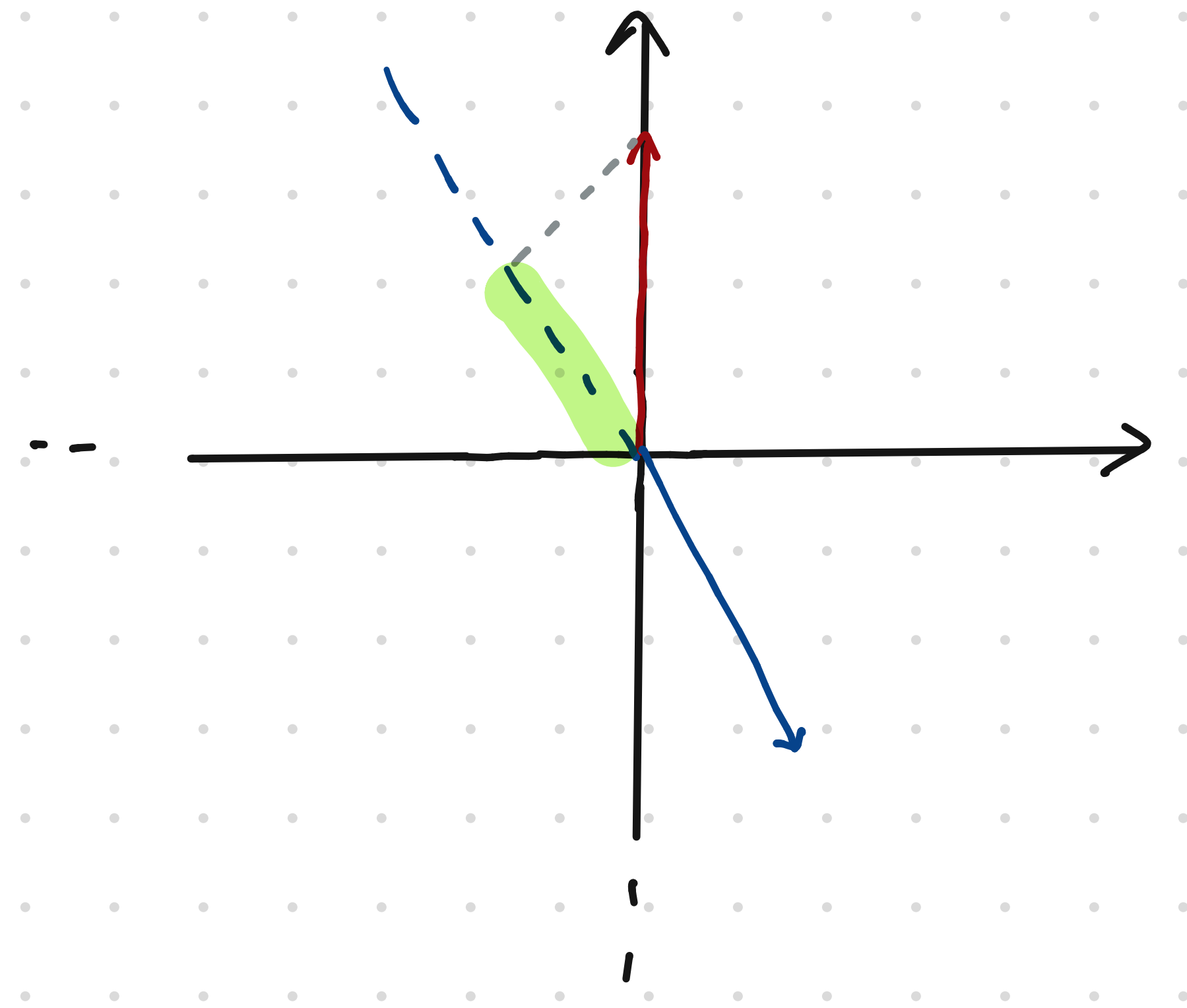
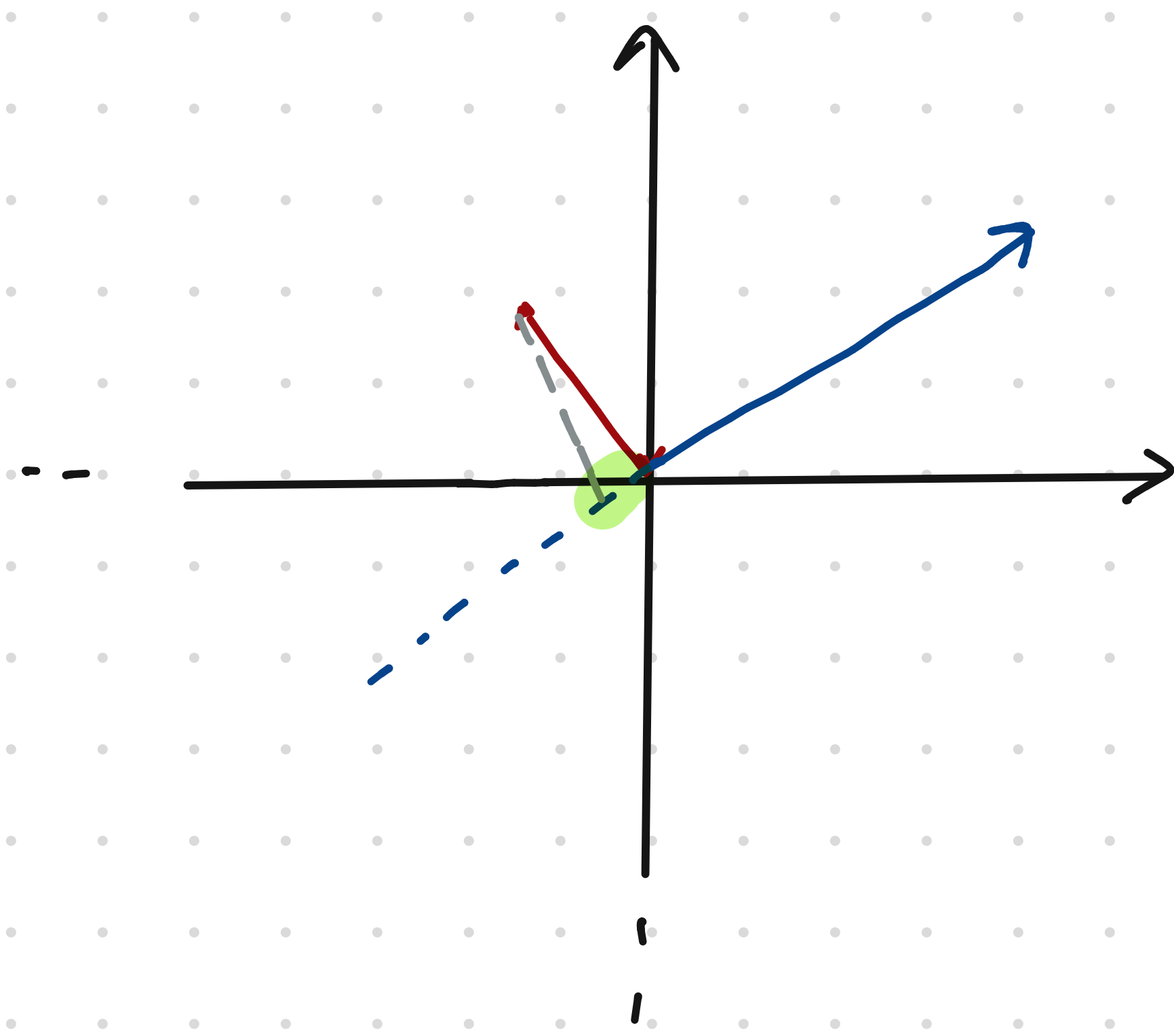
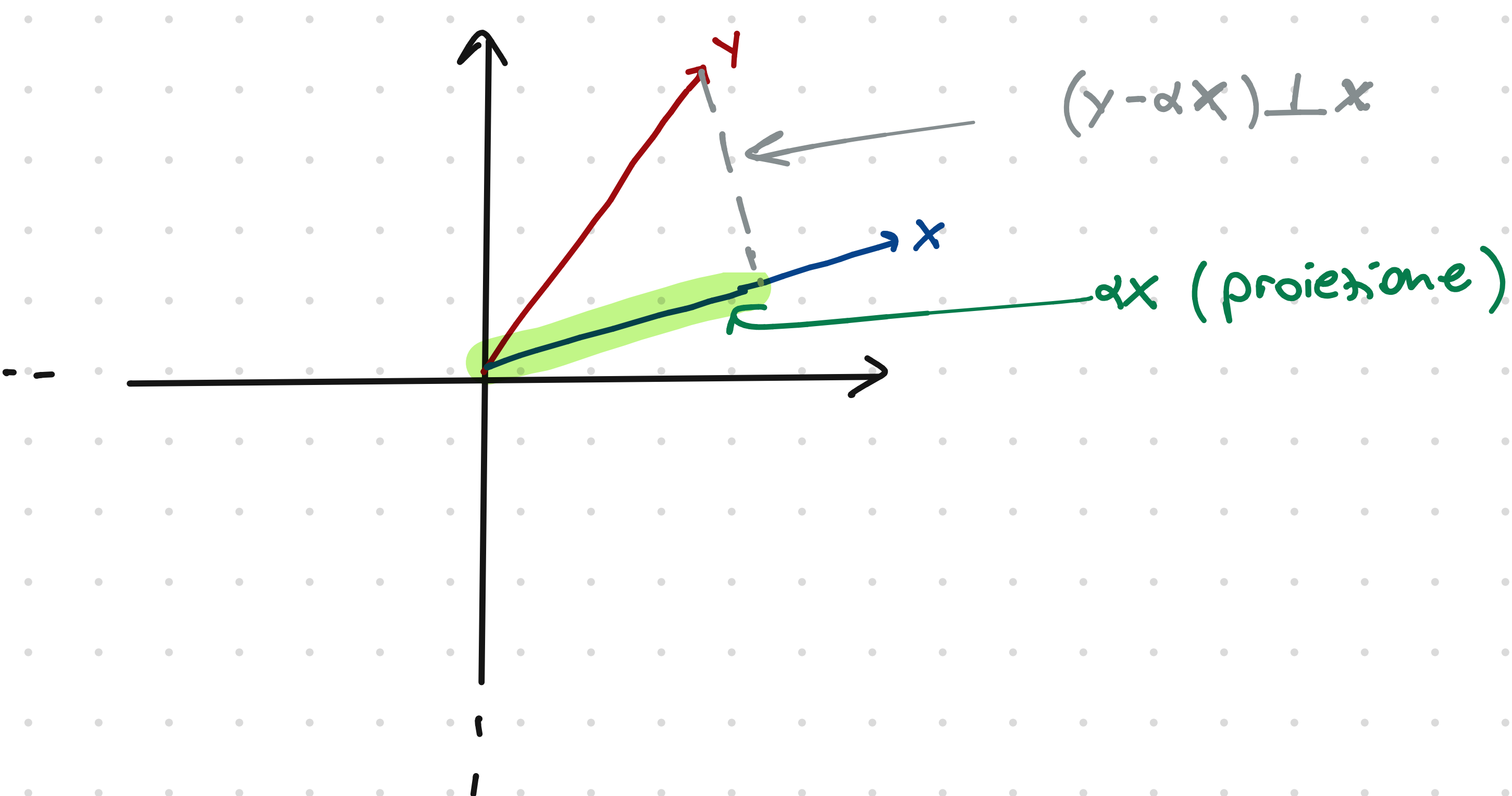
L'angolo tra  $x$  e  $y$  è pari a  $\pi/2$  (radianti). Per convertirlo in gradi applichiamo la seguente conversione:

$$\text{gradi} = \theta \times \frac{180}{\pi}$$

Nell'esempio,  $\arccos(0) \times 180/\pi = 90$ .

La proiezione di  $y$  su  $x$  è il vettore  $x \left( \underbrace{\langle x, y \rangle / \|y\|^2}_d \right)$  ed è denotata come  $\text{proj}_y(x)$ .

Esempi:



Esempio notevole utile per il calcolo:

Si consideri il coeff. di correlazione  $r_{xy}$  tra due variabili  $x$  e  $y$ . Questo può essere riscritto in termini di angolo tra i vettori associati alle due variabili, come segue:

$$\begin{aligned} r_{xy} = \text{Cor}(x, y) &= \frac{\text{Cov}(x, y)}{\text{VAR}(x)^{\frac{1}{2}} \cdot \text{VAR}(y)^{\frac{1}{2}}} \\ &= \frac{\sum_{i=1}^m (x_i - \bar{x})(y_i - \bar{y})}{\|x - \bar{x}\| \cdot \|y - \bar{y}\|} \leftarrow (x - \bar{x})^T (y - \bar{y}) \\ &= \frac{\tilde{x}^T \tilde{y}}{\|\tilde{x}\| \cdot \|\tilde{y}\|}, \text{ dove in generale: } \tilde{z} = (z - \bar{z}) \\ &= d \end{aligned}$$

$\theta = \arccos(d)$  è l'angolo tra  $x$  e  $y$ . Nel caso di incorrelazione, ossia  $x \perp y$ ,  $\theta = \frac{\pi}{2} \Rightarrow \Rightarrow \text{Cor}(x, y) = 0$ .



Consideriamo una matrice  $X_{n \times m}$ :

$$\begin{pmatrix} x_{11} & \dots & x_{1j} & \dots & x_{1m} \\ \vdots & & & & \\ x_{i1} & \dots & x_{ij} & \dots & x_{im} \\ \vdots & & & & \\ x_{m1} & \dots & x_{mj} & \dots & x_{mm} \end{pmatrix}$$

Chiameremo matrice triangolare inferiore di  $X$  ( $\text{tril}(X)$ ) il seguente sottoinsieme:

$$\text{tril}(X) = \{ x_{ij} : i \geq j, i=1, \dots, m, j=1, \dots, m \}$$

Chiameremo matrice triangolare superiore di  $X$  ( $\text{triu}(X)$ ) il sottoinsieme:

$$\text{triu}(X) = \{ x_{ij} : i \leq j, i=1, \dots, m, j=1, \dots, m \}$$

La matrice  $X$  è detta simmetrica se:

$$\text{tril}(X) = \text{triu}(X)$$

Nota: l'uguaglianza vale elemento per elemento

Esempio:

$$X_{3 \times 3} = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 1 & 3 \\ 2 & 3 & 1 \end{pmatrix}$$

Si nota che:

$$x_{21} = x_{12}$$

$$x_{31} = x_{13}$$

$$x_{23} = x_{32}$$

Una matrice  $X_{n \times m}$  si dice:

- quadrata:  $n=m$
- nulla:  $x_{ij}=0, i=1 \dots m, j=1 \dots m$  (si denota come  $0_{n \times m}$ )
- unitaria:  $x_{ij}=1, i=1 \dots m, j=1 \dots m$  (si denota come  $1_{n \times m}$ )
- diagonale:  $x_{ij}=0, i \neq j$
- identità:  $x_{ij}=1, i=j$  e  $x_{ij}=0, i \neq j$  (si denota come  $I_{n \times m}$ )

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

nulla                      unitaria

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

diagonale                      identità

La diagonale di  $X_{n \times m}$  è il seguente vettore:

$$\text{diag}(X) = \{ x_{ij} : i=j, i=1, \dots, m, j=1 \dots m \}$$

La traccia di  $X_{n \times m}$  è la seguente quantità (scalare):

$$\text{tr}(X) = \sum_{i=1}^m \text{diag}(X)_i$$

## Operazioni tra matrici (di base)

Dato  $X_{n \times m}$  si dice trasposta la matrice  $X^T$  che si ottiene scambiando righe e colonne di  $X$ :

$$X_{3 \times 3} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}, \quad X_{3 \times 3}^T = \begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{pmatrix}$$

### • Somma

Due matrici  $X_{n \times m}, Y_{p \times h}$  sono sommabili (elemento per elemento) se  $p=m, h=m$ :

$$Z_{n \times m} = X_{n \times m} + Y_{p \times h}$$

In generale,  $z_{ij} = x_{ij} + y_{ij}$  per ogni  $i=1, \dots, m$  e  $j=1, \dots, m$

le dimensioni delle matrici da sommare devono essere uguali!

Esempio:

$$\begin{pmatrix} 2 & 5 & 8 \\ 3 & 6 & 9 \\ 4 & 7 & 10 \end{pmatrix} = \begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{pmatrix} + \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

$Z_{3 \times 3}$                        $X_{3 \times 3}$                        $1_{3 \times 3}$



## • Prodotto per uno scalare

Sia  $X_{n \times m}$  una matrice e  $b \in \mathbb{R}$  uno scalare. Il loro prodotto è eseguito moltiplicando  $b$  per ciascun elemento di  $X$ , ossia:

$$Z_{n \times m} = X_{n \times m} \cdot b$$

In generale,  $z_{ij} = x_{ij} \cdot b$ , per ogni  $i=1..n$ ,  $j=1..m$ .

Esempio

$$\begin{pmatrix} 2 & 4 & 6 \\ 8 & 10 & 12 \\ 2 & 4 & 6 \end{pmatrix}_Z = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 1 & 2 & 3 \end{pmatrix}_X \cdot 2_b$$

## • Prodotto elemento per elemento (di Hadamard)

Siano  $X_{n \times m}$  e  $Y_{n \times m}$  matrici. Il prodotto elemento per elemento "o" è ottenuto come segue:

$$Z_{n \times m} = X_{n \times m} \circ Y_{n \times m}$$

In generale,  $z_{ij} = x_{ij} \cdot y_{ij}$ .

Esempio

$$\begin{pmatrix} 0 & 2 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}_Z = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix}_X \circ \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}_Y$$

Prodotti e trasposte

$$X_{2 \times 3} Y_{2 \times 3}^T = Z_{2 \times 2}$$

$$X_{4 \times 4} Y_{4 \times 1} = Z_{4 \times 1}$$

$$X_{1 \times 3} Y_{3 \times 1} = Z_{1 \times 1}$$

$$X_{3 \times 1} Y_{3 \times 1}^T = Z_{3 \times 3}$$

$$X_{1 \times 2} Y_{2 \times 2}^T = Z_{1 \times 2}$$

## • Prodotto riga-colonna

Siano  $X_{m \times m}$  e  $Y_{p \times h}$  matrici. Questo prodotto generalizza il prodotto scalare e può essere eseguito se  $m=p$ :

$$Z_{m \times h} = X_{m \times m} Y_{p \times h}, \text{ se } m=p$$

Se vale  $m=p$ , la matrice risultante  $Z$  avrà dimensioni  $m \times h$ .

In generale, il prodotto riga-colonna è calcolato come segue:

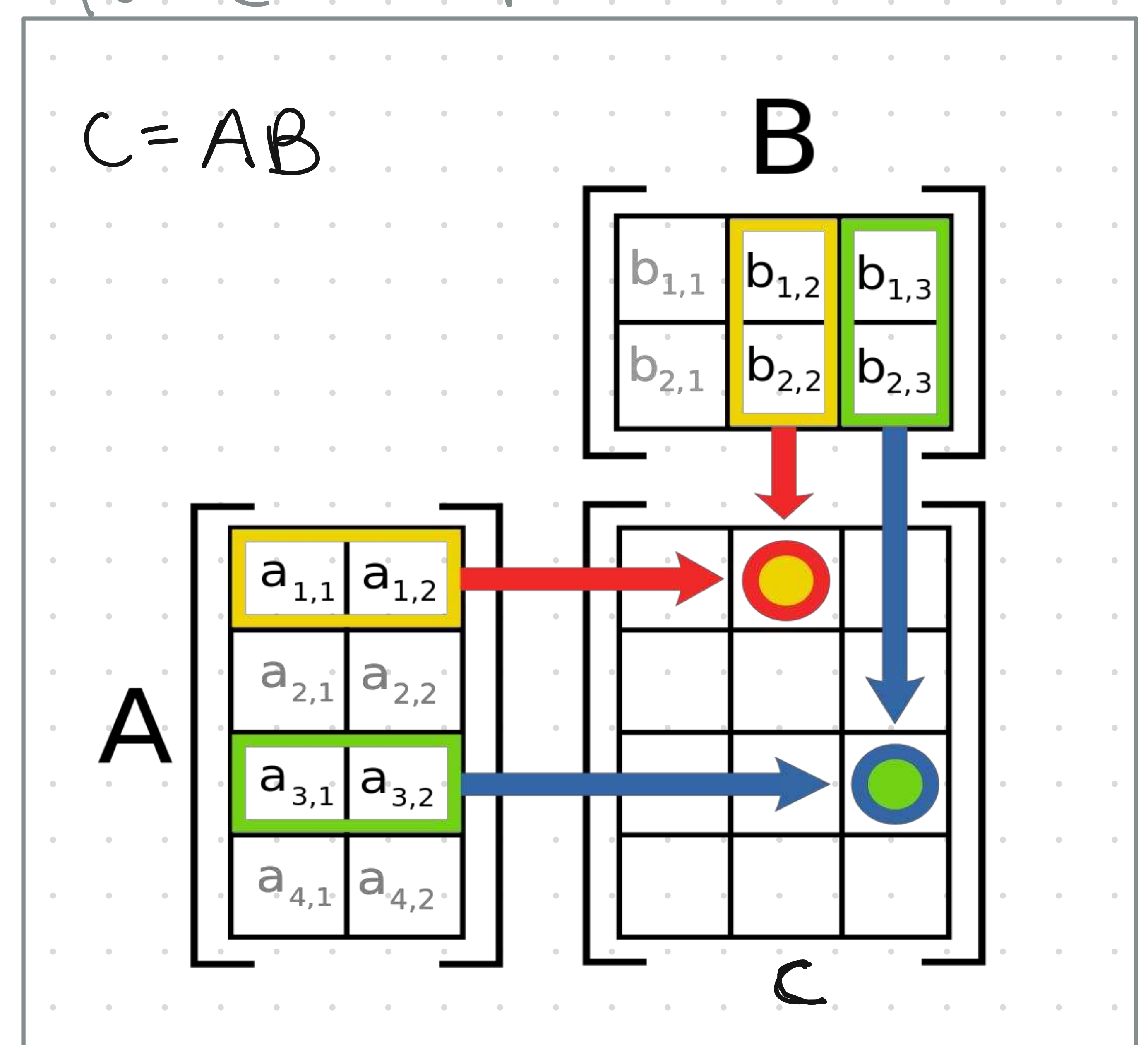
$$\begin{aligned} i &= 1..m \\ l &= 1..h \\ j &= 1..m \\ z_{il} &= \langle x_{i\cdot}, y_{\cdot l} \rangle \\ &= \sum_{k=1}^m x_{i,k}^{(k)} \cdot y_{k,l}^{(k)} \end{aligned}$$

Esempio

$$\begin{pmatrix} 3 & 11 \\ 3 & 0 \\ 3 & 6 \end{pmatrix}_{Z_{3 \times 2}} = \begin{pmatrix} 1 & 2 & 1 \\ 3 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}_{X_{3 \times 3}} \begin{pmatrix} 1 & 0 \\ 0 & 5 \\ 2 & 1 \end{pmatrix}_{Y_{3 \times 2}}$$

$$\begin{aligned} x_{1\cdot} &, y_{\cdot 1} \\ x_{2\cdot} &, y_{\cdot 1} \\ x_{3\cdot} &, y_{\cdot 2} \end{aligned}$$

Fonte: Wikipedia



$$\begin{aligned} z_{11} &= \sum_{i=1}^3 x_{1,i}^{(i)} \cdot y_{i,1}^{(i)} = \langle x_{1\cdot}, y_{\cdot 1} \rangle \\ &= (x_{1,1}^{(1)} y_{1,1}^{(1)}) + (x_{1,2}^{(2)} y_{2,1}^{(2)}) + (x_{1,3}^{(3)} y_{3,1}^{(3)}) \\ &= (1 \cdot 1) + (2 \cdot 0) + (1 \cdot 2) \\ &= 3 \end{aligned}$$

$$z_{21} = \langle x_{2\cdot}, y_{\cdot 1} \rangle = 3$$

$$z_{12} = \langle x_{1\cdot}, y_{\cdot 2} \rangle = 11$$

$$\vdots$$

$$z_{32} = \langle x_{3\cdot}, y_{\cdot 2} \rangle = 6$$



## Casi particolari (del prodotto)

5

①

Consideriamo un vettore reale  $x_{m \times 1}$  e un vettore unitario  $1_{m \times 1}$ . Vale quanto segue:

$$1^T x = \sum_{i=1}^m x_i$$

Esempio

$$\begin{pmatrix} 1 & 1 & 1 \end{pmatrix}_{1 \times 3} \begin{pmatrix} 1 \\ 3 \\ 1 \end{pmatrix}_{3 \times 1} = 5$$

Dati una qualsiasi matrice  $x_{m \times m}$ , allora:

$$\begin{aligned} x_{m \times m}^T 1_{m \times 1} &= \sum_{i=1}^m x_{ij}, \quad \text{per ogni } j=1 \dots m \\ x_{m \times m} 1_{m \times 1} &= \sum_{j=1}^m x_{ij}, \quad \text{per ogni } i=1 \dots m \end{aligned}$$

Esempio

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 0 \end{pmatrix}_{3 \times 3} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}_{3 \times 1} = \begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix}_{3 \times 1}$$

Notz: Nel primo caso, il prodotto equivale alla somma degli elementi del vettore  $x$ . Nel secondo caso, invece, il prodotto consente di effettuare, rispettivamente, la somma lungo le colonne (column-wise sum) e la somma lungo le righe (row-wise sum). In entrambi i casi, il prodotto con un vettore  $1$  può essere usato per fare la somma lungo una dimensione della matrice coinvolta.

②

Consideriamo ancora un vettore  $x_{m \times 1}$  e il vettore  $1_{m \times 1}$ . Allora:

$$x_{m \times 1} \cdot 1^T = \begin{pmatrix} x_1 & \dots & x_1 \\ x_2 & \dots & x_2 \\ \vdots & & \vdots \\ x_m & \dots & x_m \end{pmatrix}$$

Esempio

$$\begin{pmatrix} 1 \\ 3 \\ 1 \end{pmatrix}_{3 \times 1} \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}_{1 \times 3} = \begin{pmatrix} 1 & 1 & 1 \\ 3 & 3 & 3 \\ 1 & 1 & 1 \end{pmatrix}_{3 \times 3}$$

Notz: Il prodotto consente di replicare il vettore  $x$  lungo le colonne (di replicarlo  $n$  volte).

③

Consideriamo un vettore  $x_{m \times 1}$ . Allora:

$$x_{m \times 1}^T x_{m \times 1} = \sum_{i=1}^m x_i^2$$

Esempio

$$\begin{pmatrix} 1 & 3 & 1 \end{pmatrix}_{1 \times 3} \begin{pmatrix} 1 \\ 3 \\ 1 \end{pmatrix}_{3 \times 1} = 14$$

④ Consideriamo  $x_{m \times m}$  e la matrice identità  $I_{m \times m}$ . Allora:

$$x_{m \times m} \circ I_{m \times m} = [\text{diag}(x)]_{m \times m}$$

Notz: Moltiplicare una matrice con la identità elemento per elemento restituisce la matrice diagonale (contenente la diagonale di  $x$ ).



# Proprietà notevoli delle operazioni

5

## • Somma $A_{n \times m} + B_{n \times m}$

- 1) Associativa:  $A + (B + C) = (A + B) + C$
- 2) Commutativa:  $A + B = B + A$
- 3) Elemento neutro:  $A + O_{n \times m} = A$
- 4) Distributiva:  $(A + B)\alpha = \alpha A + \alpha B$

## • Prodotto riga-colonna $A_{n \times m} B_{m \times n}^T$

- 1) Associativo:  $A(BC) = (AB)C$
- 2) Distributivo:  $(A + B)C = AC + BC$
- 3)  $B I_{m \times m} = B$
- 4)  $B O_{m \times m} = O_{m \times m}$

## • Trasposizione $A^T$

- 1)  $(A^T)^T = A$
- 2)  $(A + B)^T = A^T + B^T$
- 3)  $(A\alpha)^T = \alpha A^T$
- 4)  $(AB)^T = B^T A^T$

## • Traccia $\text{tr}(A) = \sum_{i=1}^n \text{diag}(A)_i$

- 1)  $\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$
- 2)  $\text{tr}(A\alpha) = \alpha \text{tr}(A)$
- 3)  $\text{tr}(AB) = \text{tr}(BA)$

## Determinante

Il determinante di una matrice quadrata  $X_{n \times n}$  ( $\det(X)$ ) è un numero reale che fornisce informazioni su  $X$ . Il calcolo di  $\det(X)$  è effettuato mediante algoritmi iterativi e nei casi di matrici di piccole dimensioni può essere calcolato come segue:

### • Caso $n=2$

$$X_{2 \times 2} = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix}, \quad \det(X) = \prod_{i=1}^2 \text{diag}(X)_i - \prod_{i=1}^2 \overline{\text{diag}(X)_i} = (x_{11} \cdot x_{22}) - (x_{12} \cdot x_{21})$$

matrici quadrate ottenute rimuovendo la prima riga

### • Caso $n=3$ (uso del metodo di Laplace o dei minori)

$$X_{3 \times 3} = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix}, \quad \det(X) = x_{11} \underbrace{\begin{vmatrix} x_{22} & x_{23} \\ x_{32} & x_{33} \end{vmatrix}}_{M_1} - x_{12} \underbrace{\begin{vmatrix} x_{21} & x_{23} \\ x_{31} & x_{33} \end{vmatrix}}_{M_2} + x_{13} \underbrace{\begin{vmatrix} x_{21} & x_{22} \\ x_{31} & x_{32} \end{vmatrix}}_{M_3}$$
$$= x_{11} \det(M_1) - x_{12} \det(M_2) + x_{13} \det(M_3)$$

Alcune proprietà del determinante sono le seguenti:

p1) Se  $X$  contiene informazioni ridondanti sulle righe o colonne (ossia, se vettori riga o colonne sono linearmente dipendenti) si ha che  $\det(X) = 0$

Es:  $X = \begin{pmatrix} 2 & 4 \\ 3 & 6 \end{pmatrix}$ ,  $(x_{12}, x_{22}) = (x_{11}, x_{12})\alpha$  (in questo caso  $\alpha=2$ )  
 $\det(X) = 12 - 12 = 0$

p2) Se  $\det(X) \neq 0$ ,  $X$  è detta non singolare ed esiste l'inversa di  $X$  (denotata come  $X^{-1}$ )

p3) Se vale p2,  $X^{-1}$  esiste e dunque  $XX^T = I$  oppure  $XX^{-1} = I$

Es:  $X = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ ,  $XX^T = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$



## Inversa di $X_{n \times n}$

⑥

$X_{n \times n}$  ammette inversa ( $X^{-1}$ ) se:  $XX^{-1} = I$  (solo se  $\det(X) \neq 0$ )

• Caso  $n=2$

$$X_{2 \times 2} = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix}, \quad X^{-1} = \frac{1}{\det(X)} \begin{pmatrix} x_{22} & -x_{12} \\ -x_{21} & x_{11} \end{pmatrix}$$

$\underbrace{\begin{pmatrix} x_{22} & -x_{12} \\ -x_{21} & x_{11} \end{pmatrix}}_{\text{cof}(X)^T}$

• Caso  $n > 2$

Si usano algs numerici

Alcune proprietà:

1)  $(XY)^{-1} = Y^{-1}X^{-1}$

2)  $(X+Y)^{-1} = X^{-1} + Y^{-1}$

3)  $(X^T)^{-1} = (X^{-1})^T$

4)  $\det(X^{-1}) = \det(X)^{-1} = 1/\det(X)$

## Rango

Il rango  $r(X) \in [0, \min(n, m)]$  di una matrice  $X_{n \times m}$  è dato dall'ordine massimo dei minori di  $X$  aventi determinante non nullo. Il suo calcolo può essere fatto in diversi modi, uno di questi è il c.d. metodo dei minori (metodo iterativo/ricorsivo):

Dato  $X_{n \times p}$ :

①  $r = \min(n, m)$

② Esiste un minore (di  $X$ ) di ordine  $r$  con  $\det(\cdot) \neq 0$ ?

2.1 Se Sì,  $r(X) = r$ . Fine

2.2 Se No, go to ③

③  $r = (r - 1)$

④ go to ②

⑤ Se  $r=1 \Rightarrow r(X)=1$ . Fine

Esempio:

$$X_{2 \times 3} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 3 \end{pmatrix} \begin{matrix} 2 \\ 4 \end{matrix}$$

$\underbrace{\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 3 \end{pmatrix}}_{X'}$

•  $r = \min(2, 3) = 2$

•  $\det(X') = 3 \neq 0$

•  $r(X) = 2$

Nota: si può procedere esaustivamente estraendo tutte le sottomatrici quadrate e calcolando il determinante di queste. Si noterà che  $r(X)=2$ .

In generale, se  $\det(X) \neq 0$  allora  $r(X) = n$  (matrici non singolare). Nel caso di matrici singolari, invece,  $r(X) < n$ .

## Matrici dei dati

Rif. Fabbris, 2.1-2.3

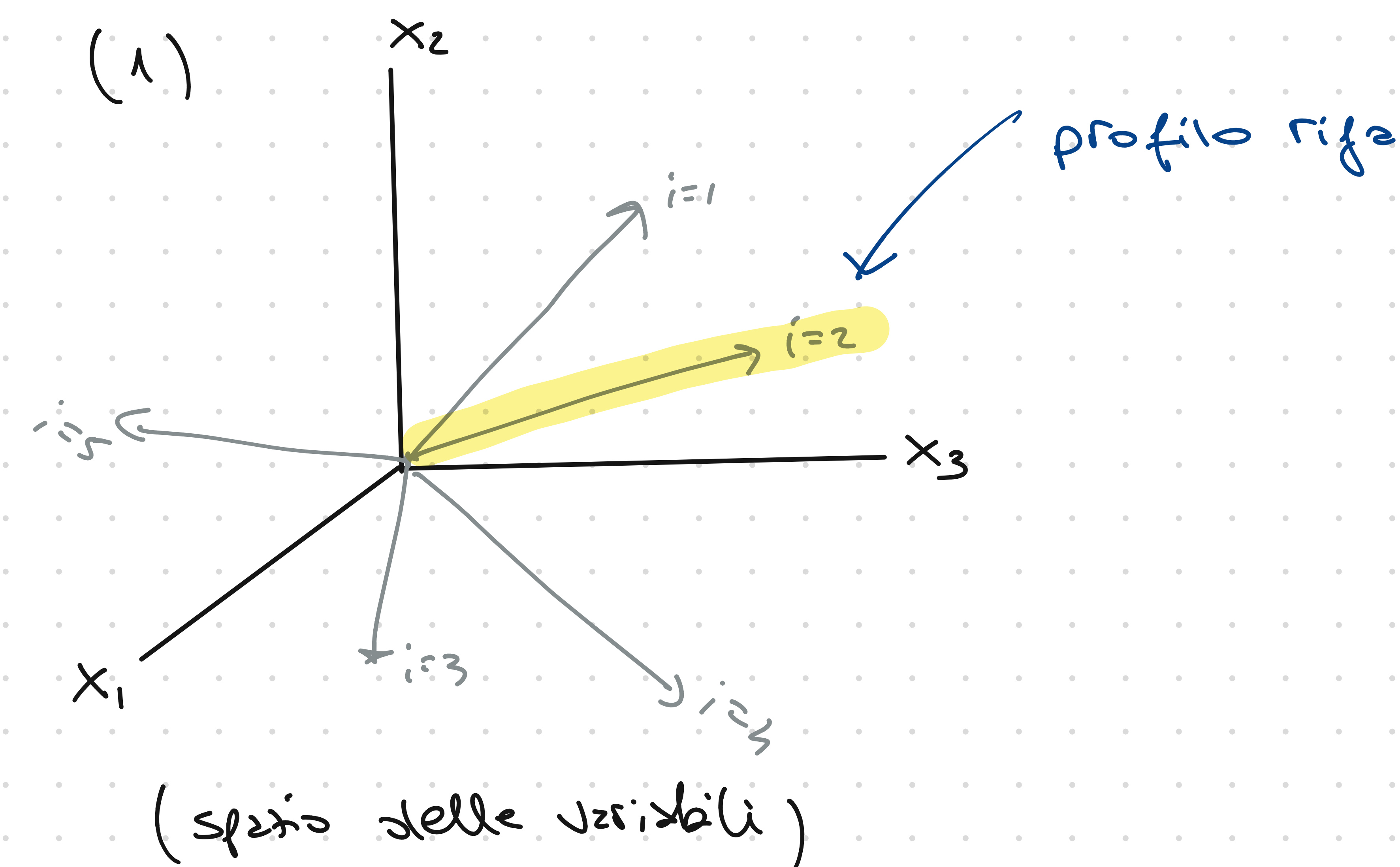
Una matrice dei dati  $X_{n \times p}$  è detta matrice unità-variabili se è una disposizione rettangolare di  $n$  unità stats (es: individui, regioni, città) e  $p$  variabili (raccelte sulle unità) come di seguito (matrice a due vie):

$$X_{n \times p} = \begin{pmatrix} x_{11} & \dots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \dots & x_{np} \end{pmatrix}$$

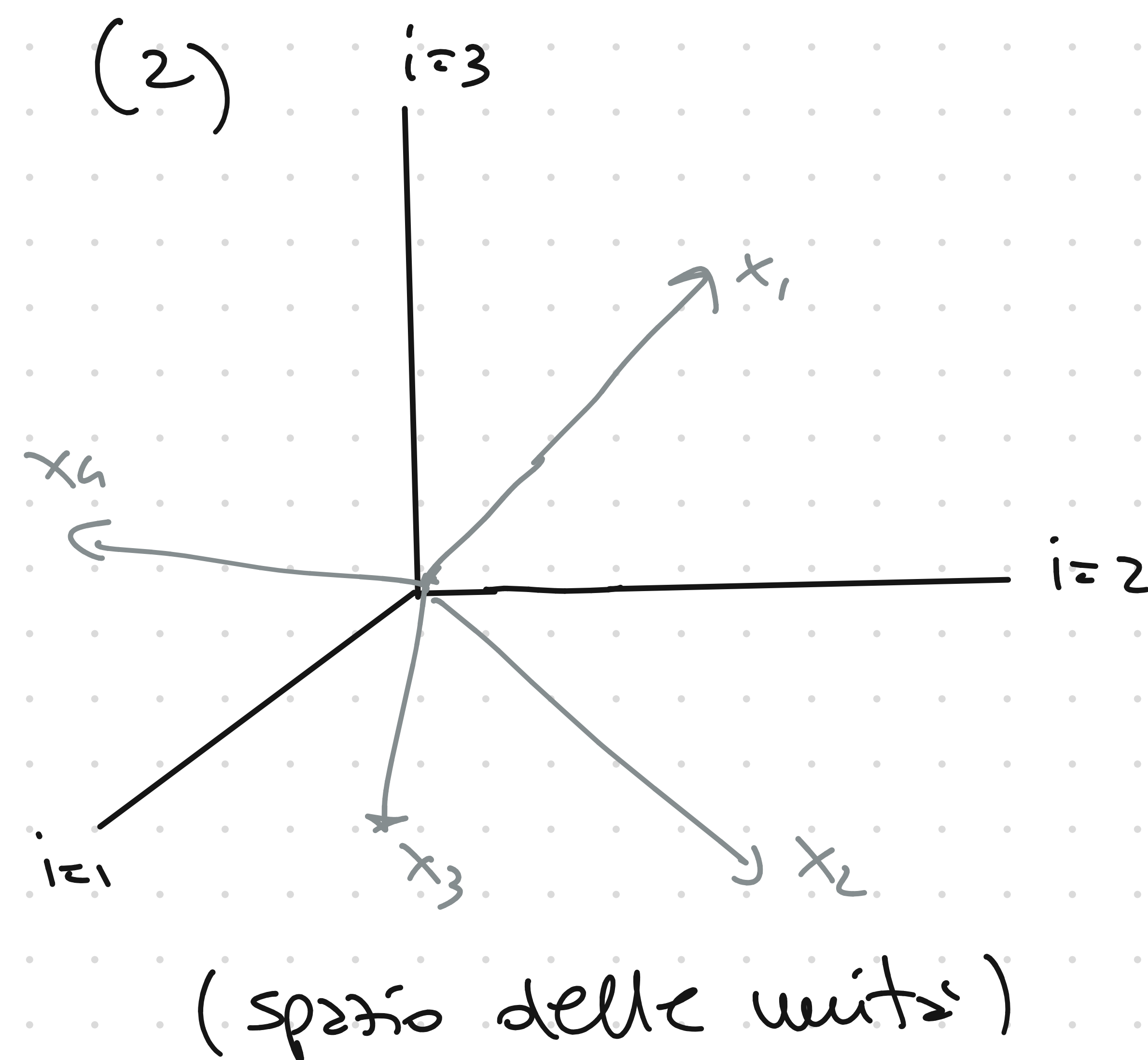
Le righe corrispondono alle unità e le colonne alle variabili.

È possibile avere due rapp.mi grafiche di  $X$  come di seguito:





Esempio con  $p=3$



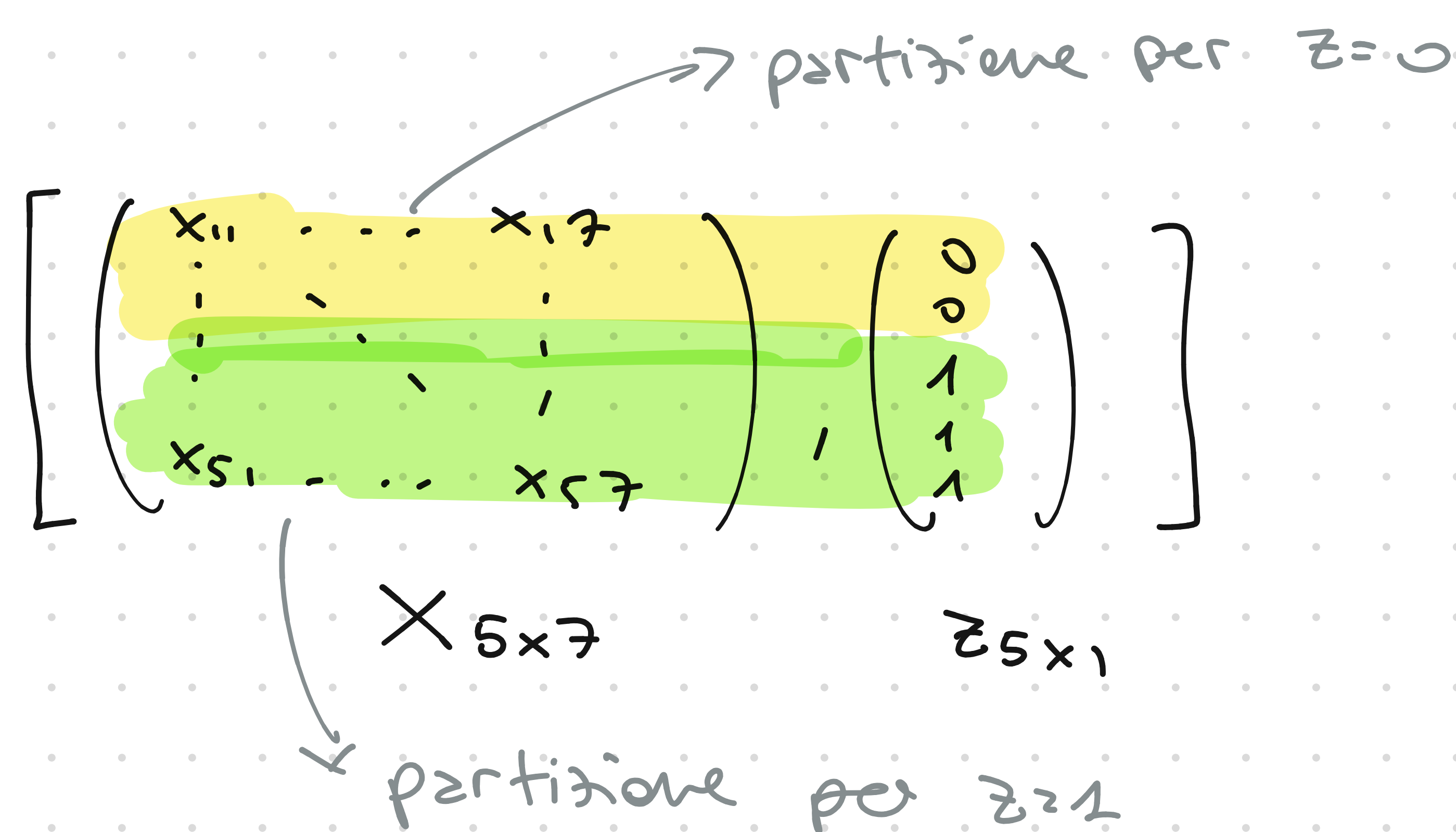
Esempio con  $m=3$

Spesso durante la raccolta di dati, oltre alla matrice  $X_{m \times p}$  contenente le vars di misurazione, sono disponibili variabili concomitanti (es: genere, livello di istruzione) spesso di tipo categoriale.

Quando queste (es:  $z_1, \dots, z_k$ ) sono usate per aggiungere informazione su  $X_{m \times p}$ , si ha come risultato che  $X_{m \times p}$  viene partitionata (stratificata) secondo i livelli (o le combinazioni tra livelli) di  $z_1, \dots, z_k$ .

Esempio

$z$ : variabile  
categoriale  $z$   
due livelli  
(distorsioni)



In questo caso,  $X_{5 \times 7}$  è partitionata in due blocchi condizionatamente ai valori di  $z$ :

$$X^* = \begin{pmatrix} X_{3 \times 7}^{(z=0)} \\ X_{2 \times 7}^{(z=1)} \end{pmatrix}$$

Dati  $X_{m \times p}$ , possiamo calcolare alcune statistiche:

(media)  $\bar{x}_j = \frac{1}{m} \sum_{i=1}^m x_{ij} \quad \forall j \in \mathcal{J}$

(varianza)  $s_j^2 = \frac{1}{m} \sum_{i=1}^m (x_{ij} - \bar{x}_j)^2$

(cov)  $s_{jh} = \frac{\sum_{i=1}^m (x_{ij} - \bar{x}_j)(x_{ih} - \bar{x}_h)}{\sqrt{s_j^2} \cdot \sqrt{s_h^2}} \quad \forall \begin{matrix} j \neq h \\ j \in \mathcal{J} \\ h \in \mathcal{J} \end{matrix}$

(cor)  $r_{jh} = s_{jh} (\sqrt{s_j^2} \cdot \sqrt{s_h^2})^{-1}$

$$\bar{x}_{p \times 1} = X_{m \times p}^T \cdot \left( \mathbf{1}_{m \times 1} \right) \frac{1}{m}$$

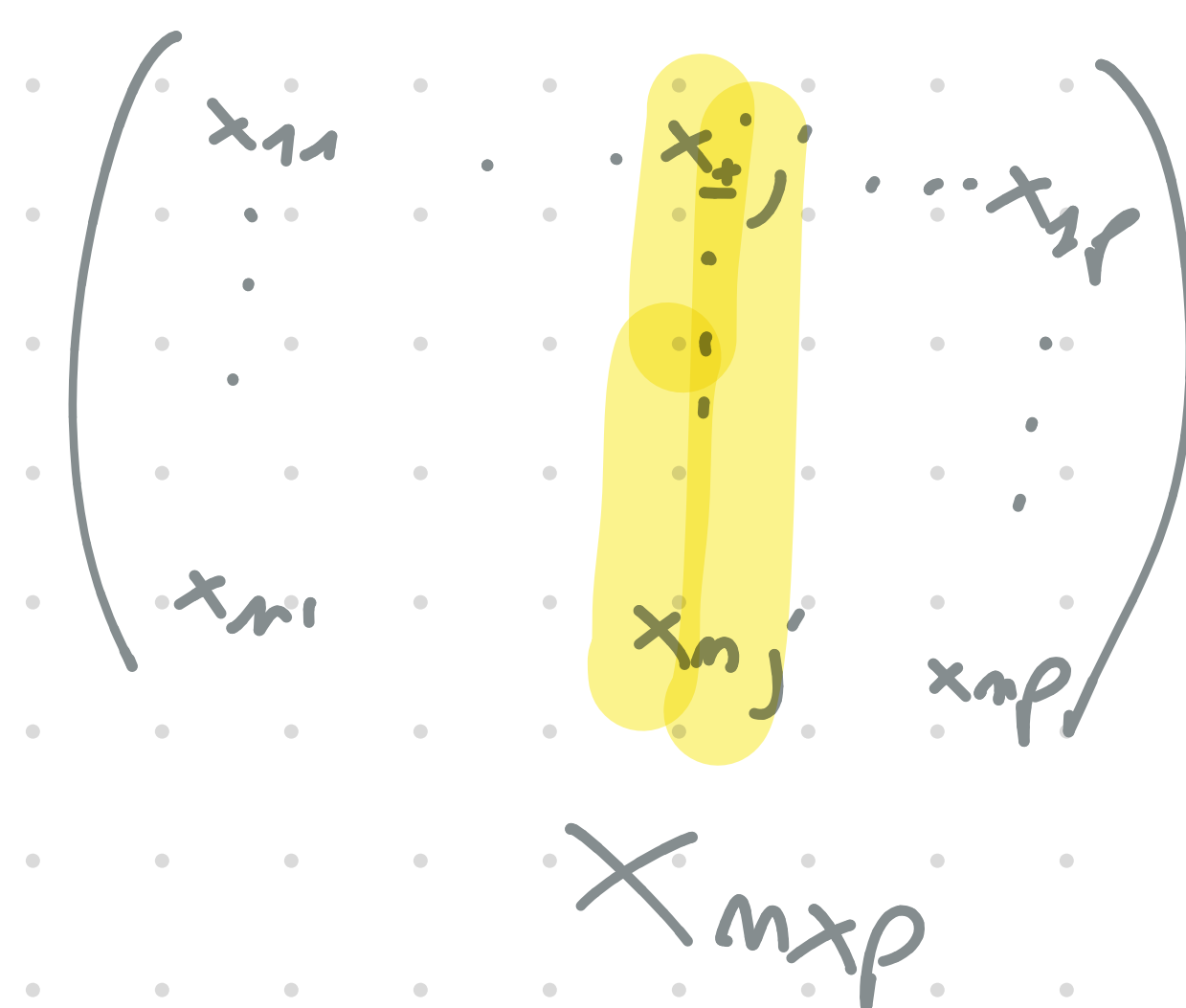
$$S_{p \times 1}^2 = \text{diag} \left( X^T X \frac{1}{m} - \bar{x} \bar{x}^T \right)$$

$$S_{p \times p} = X^T X \frac{1}{m} - \bar{x} \bar{x}^T$$

$$R_{p \times p} = \frac{1}{m} \tilde{X}^T \tilde{X}$$

$$\tilde{X}_{m \times p} = (X_{m \times p} - \mathbf{1}_{m \times 1} \bar{x}^T) \tilde{S}_{p \times p}^{-1}$$

$$\tilde{S}_{p \times p} = \left( \left( \frac{1}{m} X^T X - \bar{x} \bar{x}^T \right) \circ I_{p \times p} \right)^{\frac{1}{2}}$$



## Matrice di covarianza

$$S_{p \times p} = \begin{pmatrix} s_1^2 & \dots & s_{1p} \\ \vdots & \ddots & \vdots \\ s_{p1} & \dots & s_p^2 \end{pmatrix}$$

1)  $\text{diag}(S) = (s_1^2 \dots s_p^2)$  varianze

2) simmetrica

3)  $\text{tril}(S) = (s_{12} \dots s_{pq})$  covarianze - pfq

4)  $r(S) = p$ , autovalori positivi

5) invariante per traslazioni di colonna

$$s_{jh} = s_{hj} \in (-\infty, \infty) \\ \in \mathbb{R}$$



## Matrice di Correlazione

⑧

$$R_{p \times p} = \begin{pmatrix} 1 & \dots & r_{hj} \\ \vdots & \ddots & \vdots \\ r_{jh} & \dots & 1 \end{pmatrix}$$

Valgono le (1)-(5) precedenti.

$$r_{jh} = r_{hj} \in [-1, 1]$$

## Matrice di distanza

$$D_{p \times p} = \begin{pmatrix} 0 & \dots & d_{hj} \\ \vdots & \ddots & \vdots \\ d_{jh} & \dots & 0 \end{pmatrix}$$

$$1) \text{diag}(D) = 0_{p \times 1}$$

$$2) D \text{ e' simmetrica}$$

$$3) \text{tril}(D) = (d_{12}, \dots, d_{pq}), p \neq q \text{ distanze}$$

$$4) r(D) = p$$

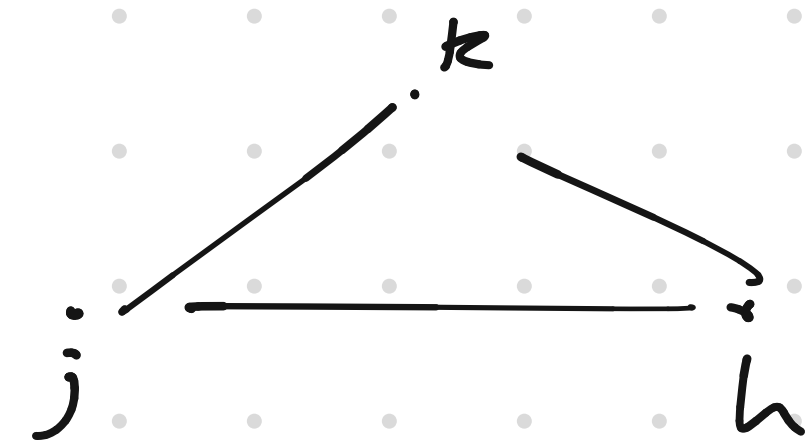
La distanza  $d_{jh}$  ( $j \neq h$ ) gode delle seguenti proprietà:

$$P_1) d_{jh} = \text{dist}(x_j, x_h) \geq 0$$

$$P_2) d_{jh} = 0 \text{ se } j=h$$

$$P_3) d_{jh} = d_{hj}$$

$$P_4) d_{jh} \leq d_{jk} + d_{kh}$$



$$P_1-P_3, P_4: d_{jh} \text{ e' detta metrica}$$

$$P_1-P_3: d_{jh} \text{ e' detta semi-metrica}$$

$$P_4: d_{jh} \leq \max\{d_{jk}, d_{kh}\} \text{ e' detta ultra-metrica}$$

## Esempi di distanza

• Euclidea:

$$d_{jh} = \text{dist}(x_j, x_h) = \left( \sum_{i=1}^m (x_{ji} - x_{hi})^2 \right)^{\frac{1}{2}} = \left( (x_j - x_h)^T (x_j - x_h) \right)^{\frac{1}{2}} = \|x_j - x_h\|_2$$

Nota: se  $x_j$  e  $x_h$  sono scalate e centrate,  $d_{jh} \propto \sqrt{2(1 - \text{cor}(x_j, x_h))}$

"tanto più forte la correlazione tra  $x_j$  e  $x_h$  tanto minore la distanza tra  $x_j$  e  $x_h$ "

• Manhattan:

$$d_{jh} = \sum_{i=1}^m |x_{ji} - x_{hi}| = \|x_j - x_h\|_1 \quad (\text{distanza del taxi})$$

"media degli scarti assoluti"

• Chebyshev:

$$d_{jh} = \sum_{i=1}^m \max\{|x_{ji} - x_{hi}|\} = \|x_j - x_h\|_\infty \quad \text{"scostamento maggiore assoluto tra due oggetti"}$$

• Jaccard:

$$d_{jh} = \frac{m_{01} + m_{10}}{m_{01} + m_{10} + m_{11}} \quad \text{con}$$

		$x_h$	
		0	1
$x_j$	0	$m_{00}$	$m_{01}$
	1	$m_{10}$	$m_{11}$

si usa quando  $x_j$  e  $x_h$  sono vettori booleani, ossia:  $x_j \in \{0, 1\}^m$   
 $x_h \in \{0, 1\}^m$

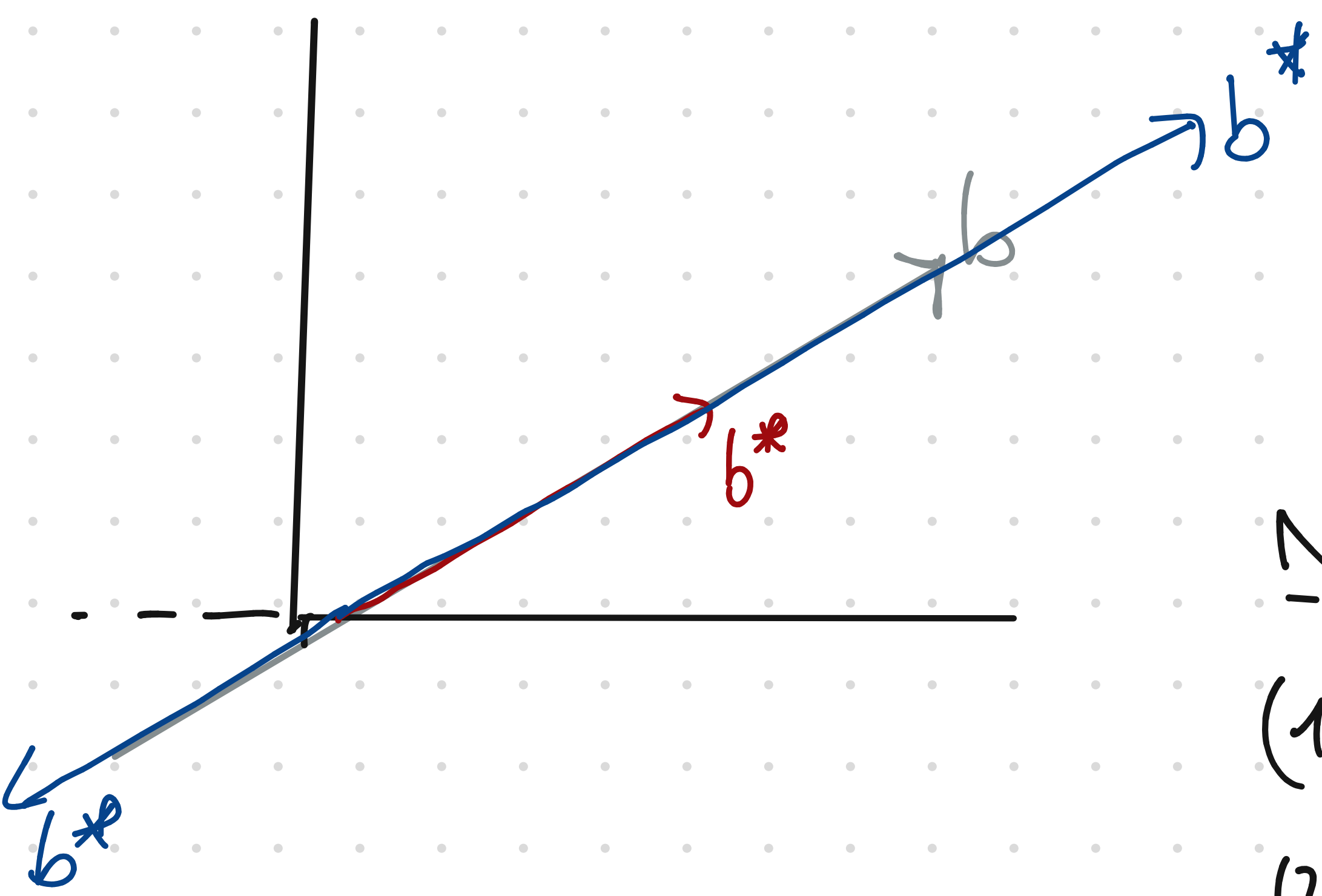


## Autovettori e autovalori (introduzione non tecnica)

Consideriamo una matrice quadrata  $X_{n \times n}$  e un vettore  $b_{n \times 1}$ :

$$X_{2 \times 2} = \begin{pmatrix} 1 & 2 \\ 1 & 0 \end{pmatrix}, \quad b_{2 \times 1} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

Il vettore risultante  $b_{n \times 1}^* = Xb$  è detto autovettore di  $X$  (eigenvector) se vale quanto segue  $b^* = b\lambda$ ,  $\lambda \neq 0$ . Lo scalare  $\lambda$  è detto autovalore (eigenvalue).



Nell'esempio,  $Xb = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$  non ha cambiato direzione, dunque è un autovettore di  $X$ .  $\underbrace{\begin{pmatrix} 4 \\ 2 \end{pmatrix}}_{b^*}$

Note:

- (1)  $\lambda < 0$  è comunque un autovalore di  $X$  (il vettore  $b$  cambia verso)
- (2) Fissato  $\lambda$ , si hanno molteplici autovettori di  $X$ . Es:  $b = \begin{pmatrix} 3 \\ 1.5 \end{pmatrix}$   
 $\lambda = 2$ ,  $b^* = b\lambda = \begin{pmatrix} 6 \\ 3 \end{pmatrix}$  è ancora un autovettore di  $X$ .
- (3) Per  $X_{n \times n}$ , ci sono  $n$  autovettori e  $n$  autovalori
- (4)  $\text{rank}(X) = r(X)$  è uguale al numero di autovalori di  $X$  non nulli
- (5)  $X$  è detta semi-positiva definita quando  $\lambda_j > 0$  per ogni  $j=1 \dots n$
- (6) Se  $X$  è simmetrica gli autovettori sono ortogonali.

In generale,  $(X - \lambda I)b = 0$  è detta equazione agli autovettori e autovalori di  $X$ . La sua risoluzione, che coinvolge il polinomio caratteristico  $\det(X - \lambda I)$ , consente di estrarre gli autovettori  $(b_1, \dots, b_m)$  e autovalori  $(\lambda_1, \dots, \lambda_m)$  di  $X$ .

Nella pagina che segue si fornisce un esempio elementare di risoluzione di  $\det(X - \lambda I)$ .



## 2.4 Appendice

### 2.4.1 Calcolo di autovalori e autovettori

Consideriamo la seguente matrice quadrata:

$$\mathbf{A}_{2 \times 2} = \begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix}$$

l'equazione caratteristica associata ad  $\mathbf{A}$  è:

$$\mathbf{A}x = \lambda x$$

dove rispetto ad  $x$ , con qualche passaggio, si ottiene:

$$(\mathbf{A} - \mathbf{I}\lambda)x = 0$$

con  $\mathbf{I}$  matrice identità. Considerando quest'ultima equazione e rispetto alla matrice data in esempio avremo:

$$\begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \cdot \lambda = \begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} = \begin{pmatrix} 2 - \lambda & 1 \\ 3 & 4 - \lambda \end{pmatrix}$$

Per ottenere il polinomio caratteristico associato alla matrice in oggetto occorre calcolare il *determinante* ed eguagliarlo a zero. In breve, il determinante di una matrice quadrata  $\det(\cdot)$  è una funzione che ad ogni matrice associa uno scalare e sintetizza alcune delle più importanti informazioni su una matrice; in termini geometrici esso ci informa sulla trasformazione che un oggetto ha subito in particolare nella conservazione dell'area e dell'orientamento nello spazio. Esso è utilizzato, con diverse interpretazioni, in svariate applicazioni di algebra lineare, ad esempio nella soluzione di sistemi lineari mediante il calcolo del *rango* e permette di capire se un sistema lineare ha soluzioni o meno. Nel caso di matrici quadrate  $2 \times 2$  ed in particolare per la matrice in esempio, il determinante è il seguente:

$$\det(\mathbf{A}) = \begin{vmatrix} 2 & 1 \\ 3 & 4 \end{vmatrix} = (2 \cdot 4) - (3 \cdot 1) = 5$$

ossia la differenza tra il prodotto degli elementi sulla diagonale principale e secondaria. Tale calcolo si estende anche a matrici di ordine superiore con qualche variante che qui comunque non riporteremo. Rispetto al nostro obiettivo possiamo scrivere:

$$\det(\mathbf{A} - \mathbf{I}\lambda) = \begin{vmatrix} 2 - \lambda & 1 \\ 3 & 4 - \lambda \end{vmatrix} = [(2 - \lambda) \cdot (4 - \lambda)] - (3 \cdot 1) = 8 - 2\lambda - 4\lambda + \lambda^2 - 3 = \lambda^2 - 6\lambda + 5$$

il risultato del determinante è il polinomio caratteristico della matrice  $(\mathbf{A} - \mathbf{I}\lambda)$ :

$$\rho(\lambda) = \lambda^2 - 6\lambda + 5$$



Le soluzioni dell'equazione di secondo grado  $\rho(\lambda)$  sono le seguenti:

$$\lambda_1 = 1 \quad \text{e} \quad \lambda_2 = 5$$

che sono gli autovalori della matrice  $\mathbf{A}$ . Per calcolare gli autovettori associati ai due autovalori trovati, possiamo procedere sostituendo  $\lambda_1$  e  $\lambda_2$  nell'equazione  $(\mathbf{A} - \lambda \mathbf{I})x = 0$ , come segue:

$$\text{considerando } \lambda_1 = 1 \quad \left[ \begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right] \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

dove la matrice che viene sottratta ad  $\mathbf{A}$  è la matrice risultante dal prodotto  $\lambda \mathbf{I}$ ; con qualche passaggio otteniamo il seguente sistema:

$$\begin{cases} x_1 + x_2 = 0 \\ 3x_1 + 3x_2 = 0 \end{cases}$$

dove chiaramente le soluzioni sono  $(1, -1)$ . Il *primo autovettore*  $\mathbf{a}$  associato all'autovalore  $\lambda_1 = 1$  è un multiplo non nullo di

$$\mathbf{a} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

$$\text{considerando } \lambda_2 = 5 \quad \left[ \begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix} - \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix} \right] \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

dove operando come prima otteniamo:

$$\begin{cases} -3x_1 + x_2 = 0 \\ 3x_1 - x_2 = 0 \end{cases}$$

le soluzioni sono chiaramente  $(1, 3)$ . Il *secondo autovettore*  $\mathbf{b}$  associato a  $\lambda_2 = 5$  è un multiplo non nullo di:

$$\mathbf{b} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$$



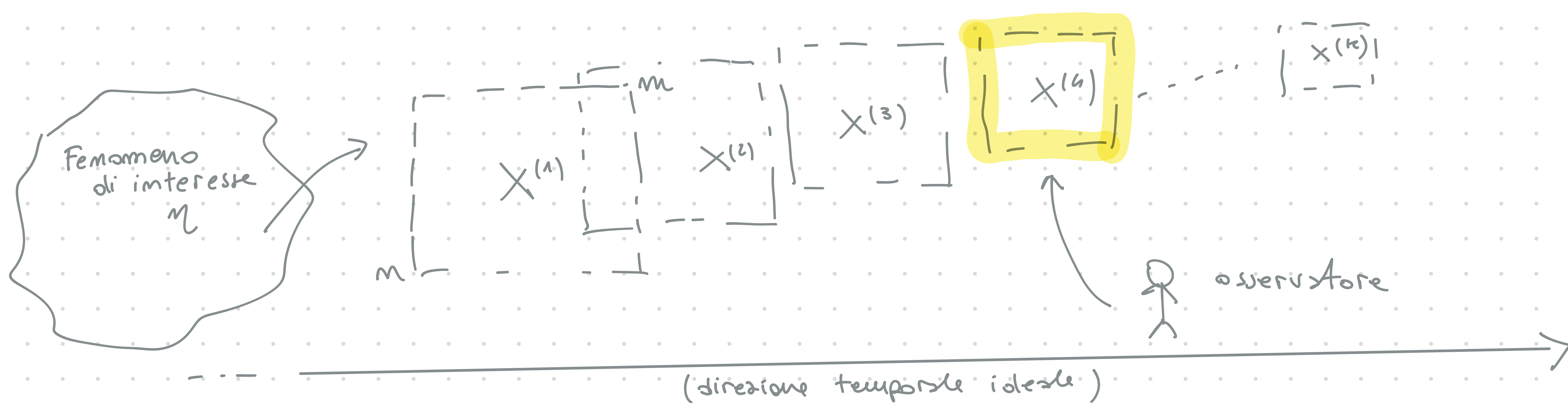
Rif.: Ross, capp. 4-5 (vedi Peruselli)

Note: gli appunti che seguono non sono esaustivi rispetto ai contenuti tipici affrontati in un corso di Inferenza Statistica. Al contrario essi vogliono offrire un compendio elementare di alcuni strumenti inferenziali utili per il corso che seguirà. Si preferirà, perciò, una disposizione "non strutturata" dei concetti offerti. Si rimanda ai riferimenti elencati per una esposizione più strutturata del materiale qui ripreso.

## Introduzione

Nella parte precedente abbiamo introdotto l'algebra lineare quale strumento essenziale al calcolo sviluppato in questo corso. Data una matrice (non necessariamente quadrata)  $X_{n \times m}$  abbiamo assegnato un primo significato ai vettori riga e vettori colonna come, rispettivamente, unità statistiche (es: individui, città) e variabili (su cui le unità sono misurate). Assumiamo così anche un significato peculiare le operazioni effettuate su  $X$ . Ad esempio, supposto  $n=m$ , l'operazione  $\text{dist}(X)$  restituisce la matrice delle distanze tra i profili riga, ad esempio nel caso di individui, la distanza tra l'individuo  $i$  e  $i'$  (con  $i \neq i'$ ). Viceversa, l'operazione  $\text{dist}(X^T)$  restituisce la matrice delle distanze tra le variabili, ad esempio la distanza tra le risposte degli individui alla variabile  $j$  rispetto a  $j'$  (con  $j \neq j'$ ).

Un'ulteriore assunzione di significato è necessaria in questo momento rispetto alla natura stocastica/storica di  $X$ .



Idealmente,  $X_{n \times m}$  è un esito - tra molteplici possibili - del fenomeno off. di studio ( $M$ ) rispetto a un osservatore esterno e a un istante temporale fissato. La matrice  $X_{n \times m}$  contiene gli esiti del processo aleatorio che si vuole studiare/comoscere. Anche se non necessario, assumeremo che i vettori riga di  $X$  sono indipendenti, ossia  $X_{1 \times m}^{(i)} \perp X_{1 \times m}^{(i')}$  ( $i \neq i'$ ). Ciò, richiesto per la semplificazione del modello statistico che utilizzeremo, implica che le unità statistiche sono omogenee rispetto al fenomeno che si vuole indagare (e sono dunque scambiabili/permutabili tra loro). Al contrario, i vettori colonna di  $X$  sono non indipendenti e ciò è richiesto dalla natura dei fenomeni che analizzeremo nel corso. L'interesse dunque risiede nell'indagare  $M$  attraverso una collezione indipendente  $(X_{1 \times m}^{(1)}, \dots, X_{1 \times m}^{(n)})$  di "rilevazioni" (o misurazioni) fatte dall'osservatore su  $M$  attraverso le unità statistiche. In questo contesto,  $x^{(i)}$  è un vettore aleatorio dotato di un certo modello probabilistico  $F_{X^{(i)}}(x; \theta)$  che descrive il suo comportamento stocastico. Naturalmente  $\theta$ , che esprime il modo con cui  $F_X(\cdot)$  "si comporta", deve esprimere anche l'informazione sul fatto che  $(x_1^{(i)}, \dots, x_m^{(i)})$  sono non indipendenti (vedremo in seguito tale aspetto).

La giustificazione di  $m > 1$  (se  $m=1$  avremo una sola misurazione per ciascuna unità stat) risiede nel fatto che per conoscere  $M$  sono necessarie più misurazioni affinché l'errore indotto dalla misurazione e il "rumore di fondo" del processo aleatorio non superi il segnale proveniente da  $M$ .



L'osservatore ha perciò interesse a quantificare  $\eta$  (misurando) e tale quantificazione (2) deve essere tale che le rilevazioni  $x^{(1)}, \dots, x^{(m)}$  siano sufficienti/idonee a fornire una sua stima precisa e attendibile.

Un primo modo per fornire una stima di  $\eta^{(i)}$  per ciascuna unità statistica è quella di applicare un operatore di aggregazione su  $x_{1 \times m}^{(i)}$ , ad esempio mediante somma (o media):

$$\hat{\eta}^{(i)} = x_{1 \times m}^{(i)} \cdot 1_{m \times 1} \quad \left( \hat{\eta}_j^{(i)} = \sum_j x_j^{(i)} \right)$$

Naturalmente tale aggregazione ignora la non indipendenza delle rilevazioni  $x_{1,1}^{(i)}, \dots, x_{1,m}^{(i)}$ . Aggiungendo dei pesi non negativi a somma unitaria, porterebbe a un miglioramento della situazione:

$$\begin{aligned} \hat{\eta}^{(i)} &= x_{1 \times m}^{(i)} \cdot w_{m \times 1}^{(i)} & \left( \hat{\eta}_j^{(i)} = \sum_j x_j^{(i)} w_j^{(i)} \right) \\ w^{(i)} 1_m &= 1 & \left( \sum_j w_j^{(i)} = 1 \right) \end{aligned}$$

Tuttavia rimane il problema di come determinare  $w_{m \times 1}^{(i)}$  e fare delle scelte razionali sui pesi (ad es:  $w^{(i)}$  uguale per ogni  $j$ ,  $w^{(i)}$  uguale per ogni  $i$  e  $j$ , ecc). Questo problema misurativo verrà affrontato all'interno di questo corso mediante i c.d. modelli fattoriali.

Negli appunti che seguono verranno fornite informazioni e strumenti essenziali/basilari per costruire/definire al meglio il problema misurativo alla base dei modelli fattoriali.

## Variabili aleatorie

Richiamando alcune nozioni di base, indipendentemente dal contesto descritto sopra, diremo che una variabile aleatoria  $X \sim F_X(x; \theta)$  riassume l'effetto del nostro studio. Più precisamente, un modello statistico su  $X$  è la seguente tripla:

$$M_X = \left\{ \underbrace{F_X(x; \theta)}_{\substack{\text{modello (parametrico) \\ probabilistico su } X}}, \underbrace{\theta \in \Theta \subseteq \mathbb{R}^p}_{\substack{\text{spazio parametrico} \\ p\text{-dimensionale}}}, \underbrace{x \in \mathcal{X}^m}_{\substack{\text{spazio campionario (o dei dati)} \\ n\text{-dimensionale}}} \right\}$$

Ricorderemo che  $X$  può essere discreta o continua.

### Esempio

$X \sim N(x; \mu, \sigma)$	$X \sim \text{Bern}(x; \pi)$
$F_X \hat{=}$ densità di prob. $N(\cdot)$	$F_X \hat{=}$ distribuzione di massa $\text{Bern}(\cdot)$
$\theta = \{\mu, \sigma\} \in \mathbb{R} \times \mathbb{R}^+$ ( $p=2$ )	$\theta = \{\pi\} \in [0, 1]$
$\mathcal{X}^m = \mathbb{R}^m$	$\mathcal{X}^m = \{0, 1\}^m$

Nelle tabelle che seguono si riportano alcuni modelli probabilistici discreti e continui.

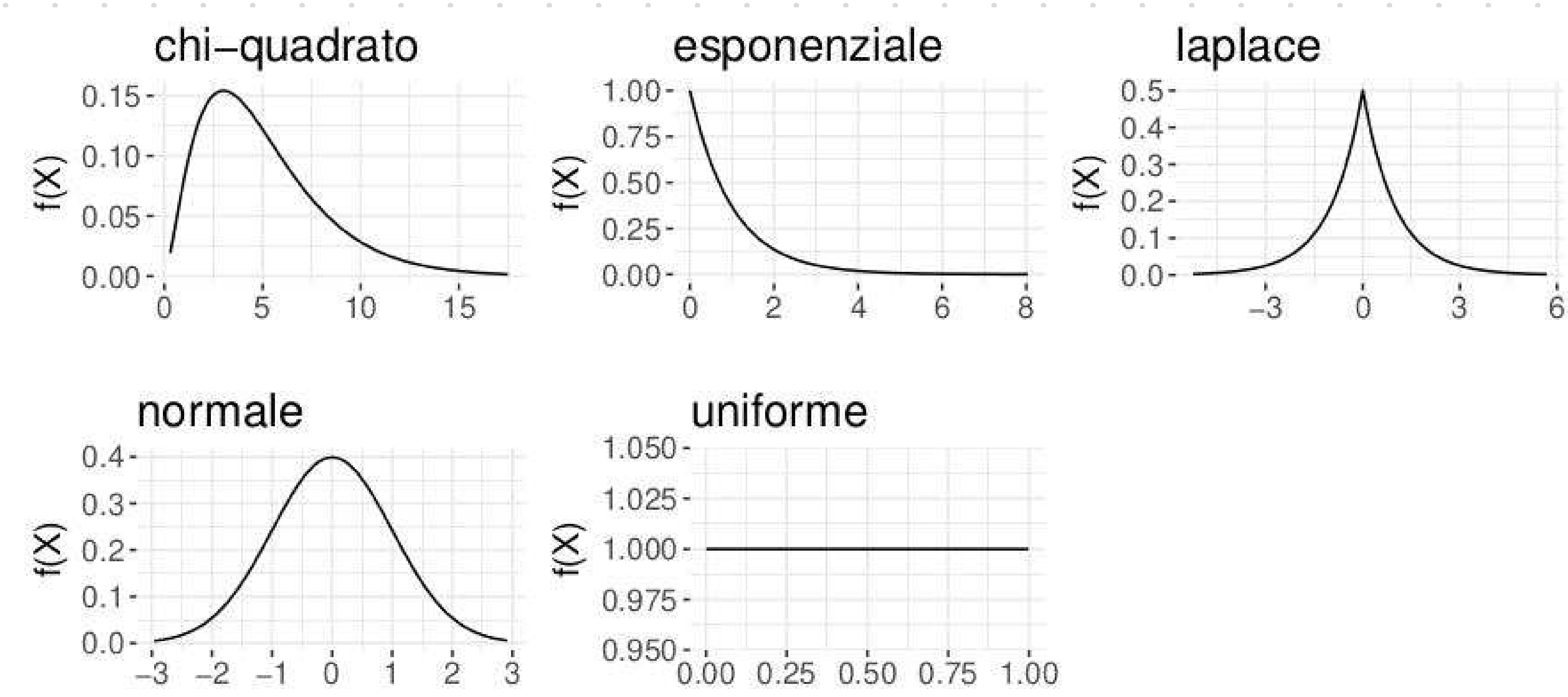
Some univariate **discrete probabilistic models**

Model	Notation	$\text{sup}(X)$	$\theta$	$f_X$
Bernoulli	$Ber(x; \pi)$	$\{0, 1\}$	$\pi \in [0, 1]$	$\pi^x (1 - \pi)^{1-x}$
Binomial	$Bin(x; n, \pi)$	$\mathbb{N}_0$	$n \in \mathbb{N}$ $\pi \in [0, 1]$	$\binom{n}{x} \pi^x (1 - \pi)^{n-x}$
Poisson	$Poi(x; \lambda)$	$\mathbb{N}_0$	$\lambda \in \mathbb{R}^+$	$\frac{\lambda^x}{x!} \exp(-\lambda)$
Geometric	$\mathcal{G}(x; \pi)$	$\mathbb{N}$	$\pi \in [0, 1]$	$\pi (1 - \pi)^{x-1}$
Multinomial	$Multi(x; n, \pi)$	$\mathbb{N}_0$	$n \in \mathbb{N}$ $\pi = (\pi_1, \dots, \pi_K)$ $\pi^T 1_K = 1$	$\binom{n!}{x_1! \dots x_K!} \pi_1^{x_1} \dots \pi_K^{x_K}$



Model	Notation	$\text{sup}(X)$	$\theta$	$f_X$
Normal	$\mathcal{N}(x; \mu, \sigma^2)$	$\mathbb{R}$	$\mu \in \mathbb{R},$ $\sigma \in \mathbb{R}^+$	$(\sigma\sqrt{2\pi})^{-1} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$
Uniform	$\mathcal{U}(x; \alpha, \beta)$	$[\alpha, \beta] \subset \mathbb{R}$	$\alpha \in \mathbb{R}$ $\beta \in \mathbb{R}$ $\alpha < \beta$	$\frac{1}{\beta-\alpha}$
Exponential	$\mathcal{Exp}(x; \lambda)$	$\mathbb{R}^+$	$\lambda \in \mathbb{R}$	$\lambda \exp(-\lambda x)$
Beta	$\mathcal{Beta}(x; \alpha, \beta)$	$[0, 1] \subset \mathbb{R}$	$\alpha \in \mathbb{R}^+,$ $\beta \in \mathbb{R}^+$	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}$
Chi-square	$\chi^2(x; \nu)$	$\mathbb{R}^+$	$\nu \in \mathbb{N}$	$(2^{\nu/2} \Gamma(\nu/2))^{-1} x^{\nu/2-1} \exp(-x/2)$

Alcuni modelli continui sono rappresentati in figura...



Il comportamento medio (stesso, sperato) di un esperimento aleatorio e' la seguente quantita' scalare:

$$E[X] = \sum_{x \in \text{sup}(X)} x \cdot f_X(x; \theta) \quad (\text{caso discreto})$$

$$E[X] = \int_{\text{sup}(X)} x f_X(x; \theta) dx \quad (\text{caso continuo})$$

La varianza e' uno scalare che quantifica la dispersione di un esperimento aleatorio:

$$V[X] = \sum_{x \in \text{sup}(X)} (x - E[X])^2 f_X(x; \theta) \quad (\text{caso discreto})$$

$$V[X] = \int_{\text{sup}(X)} (x - E[X])^2 f_X(x; \theta) dx \quad (\text{caso continuo})$$

Proprietà di  $E[\ ]$  e  $V[\ ]$ :

$E[\alpha] = \alpha$

$V[\alpha] = 0$

$E[X+Y] = E[X] + E[Y]$

$V[X+Y] = V[X] + V[Y] + 2Cov[X, Y]$   
se  $X \perp Y$ ,  $Cov[X, Y] = 0$

$E[X\beta] = \beta E[X]$

$V[X\beta] = V[X]\beta^2$

$E[\alpha + X\beta] = \alpha + \beta E[X]$

$V[\alpha + X\beta] = \beta^2 V[X]$

Quando gli esperimenti casuali sono descritti da piu' variabili aleatorie congiuntamente, nasce nella necessita' di descrivere il comportamento congiunto di tali variabili. Diremo:

$$X = (X_1, \dots, X_m) \sim F_{X_1, \dots, X_m}(x; \theta)$$

dove  $F_{X_1, \dots, X_m}(x; \theta)$  e' la densita' di probabilita' congiunta del vettore aleatorio  $X$   $m$ -dimensionale.



Ponendo  $m=2$  per semplicità di esposizione, la densità marginale:

(4)

$$f_{x_1}(x_1; \theta) = \int_{\text{sup}(x_2)} f_{x_1, x_2}(x; \theta) dx_2$$

esprime l'informazione di  $x_1$  quando  $x_2$  viene 'eliminata' (integrated out). In questo caso, il comportamento di  $x_1$  viene studiato "isolatamente" (senza considerare che  $x_1, x_2$  sono state osservate congiuntamente). Analoghe considerazioni valgono per  $f_{x_2}(x_2; \theta)$ .

La densità condizionale

$$f_{x_1|x_2}(x_1; \theta) = [f_{x_2}(x_2; \theta)]^{-1} f_{x_1, x_2}(x; \theta)$$

esprime l'informazione della variabile aleatoria  $x_1|x_2$ , dove  $x_2$  condiziona il verificarsi di  $x_1$ . E' evidente che in questo caso si assume una asimmetria tra  $x_1$  e  $x_2$ , essendo quest'ultima antecedente (condizionante il comportamento di  $x_1$ ). Analoghe considerazioni valgono per  $x_2|x_1$ .

Le v.z.  $x_1$  e  $x_2$  sono indipendenti ( $x_1 \perp x_2$ ) se  $f_{x_1|x_2}(x_1; \theta) = f_{x_1}(x_1; \theta)$  oppure se  $f_{x_1, x_2}(x; \theta) = f_{x_1}(x_1; \theta) \cdot f_{x_2}(x_2; \theta)$ .

Quando si studiamo oggetti del tipo  $x_1|x_2$  si può essere interessati a conoscerne il comportamento medio. In questo caso, i valori attesi condizionati si generalizzano:

$$\mathbb{E}[x_1 | x_2 = x_2] = \int_{\text{sup}(x_1)} x_1 \cdot \underbrace{f_{x_1|x_2}(x_1; \theta)}_{\text{densità della v.z. condizionata}} dx_1$$

$$\mathbb{V}[x_1 | x_2 = x_2] = \mathbb{E}[(x_1 - \mathbb{E}[x_1 | x_2])^2 | x_2 = x_2]$$

Nello studio della v.z. congiunta  $(x_1, \dots, x_m)$  può essere importante quantificare il grado con cui le m-realizzazioni covariano tra loro (associazione lineare). A questo riguardo, il seguente valore atteso assume rilevanza:

$$\begin{aligned} \text{Cov}[x_h, x_j] &= \mathbb{E}[(x_h - \mathbb{E}[x_h])(x_j - \mathbb{E}[x_j])] \\ &= \mathbb{E}[x_h x_j] - \mathbb{E}[x_h] \mathbb{E}[x_j] \end{aligned}$$

Tale quantità è detta covarianza tra  $x_h$  e  $x_j$ . Poiché è una misura di associazione lineare si hanno i casi seguenti.

$\text{Cov}[x_h, x_j] > 0$ :  $x_h$  e  $x_j$  sono positivamente associati

$\text{Cov}[x_h, x_j] = 0$ :  $x_h$  e  $x_j$  non sono linearmente associati

$\text{Cov}[x_h, x_j] < 0$ :  $x_h$  e  $x_j$  sono negativamente associati

Modello Normale multivariato Rif.: Pize e Salem, cap. 10

Consideriamo un vettore aleatorio  $m$ -dimensionale  $X = (x_1, \dots, x_m)$ . Si dice che  $X$  segue in legge il modello Normale  $m$ -variato se:

$$X \sim N_m(x; \mu, \Sigma)$$

con  $\mu \in \mathbb{R}^m$  (vettore delle medie) e  $\Sigma \in \mathbb{S}^{m \times m}$  (matrice di covarianza  $m \times m$ ), con  $\mathbb{S}^{m \times m}$  detto spazio delle matrici semi-positive definite.



In termini matriciali, il modello e' il seguente:

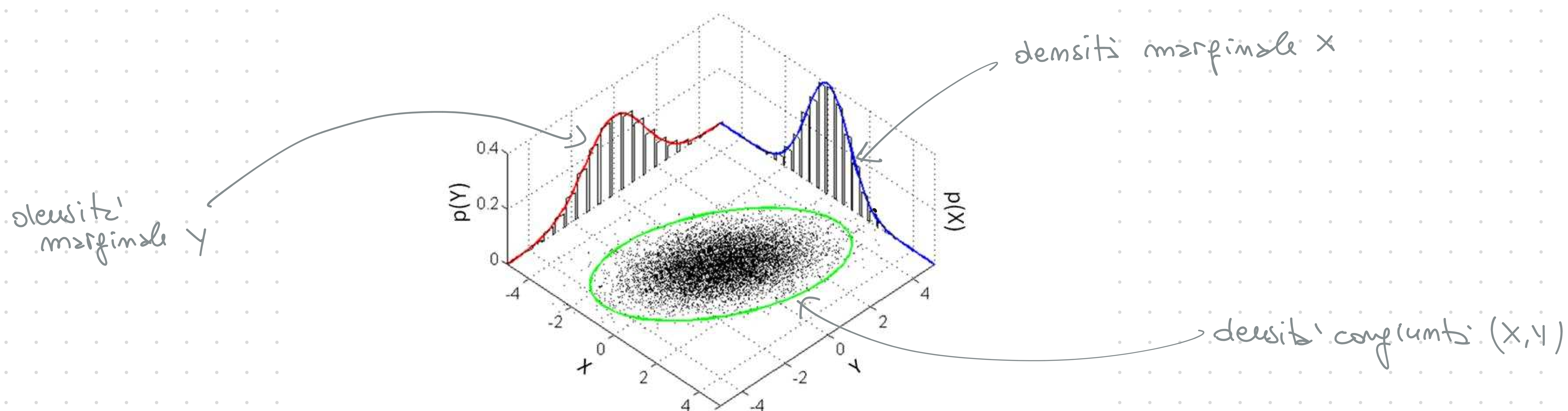
5

$$\underbrace{\begin{pmatrix} x_1 \\ \vdots \\ x_j \\ \vdots \\ x_m \end{pmatrix}}_{X_{m \times 1}} \sim N_m \left( \underbrace{\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_j \\ \vdots \\ \mu_m \end{pmatrix}}_{\mu_{m \times 1}}, \underbrace{\begin{pmatrix} \sigma_{11}^2 & \dots & \sigma_{1j}^2 & \dots & \sigma_{1m}^2 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \sigma_{j1}^2 & \dots & \sigma_{jj}^2 & \dots & \sigma_{jm}^2 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \sigma_{m1}^2 & \dots & \sigma_{mj}^2 & \dots & \sigma_{mm}^2 \end{pmatrix}}_{\Sigma_{m \times m}} \right)$$

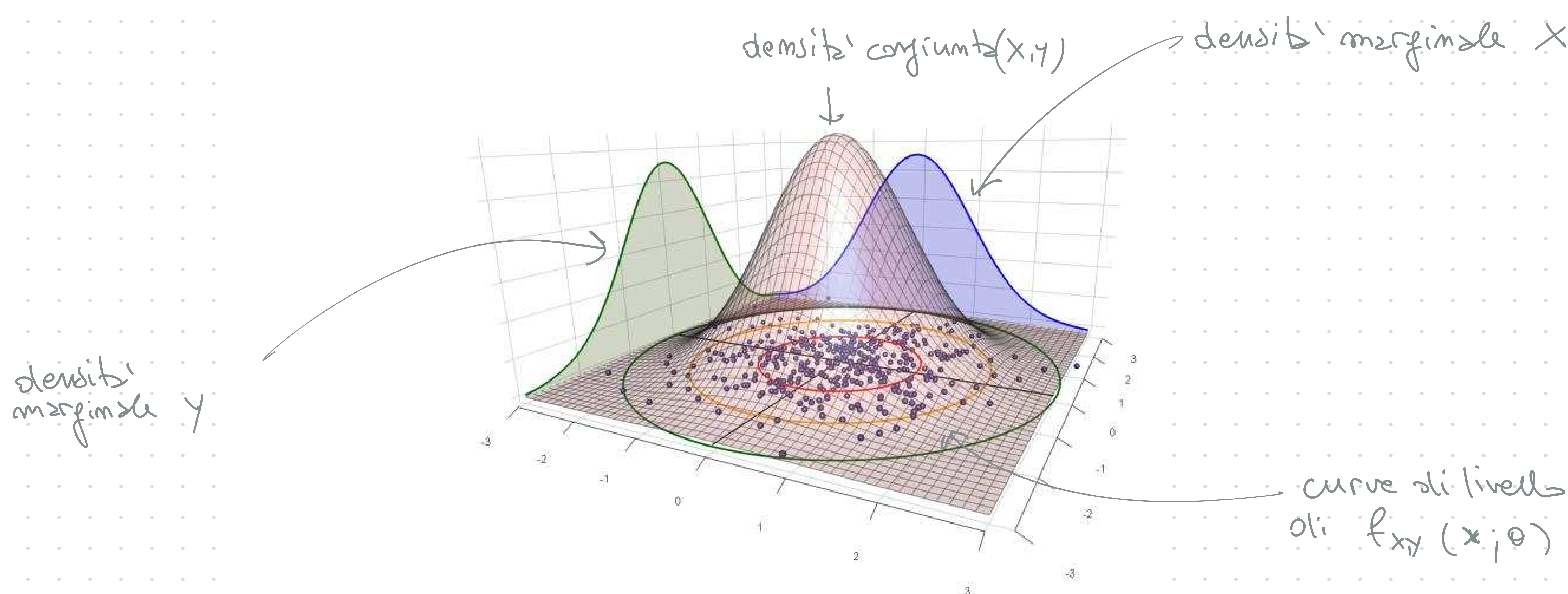
Come detto nelle lezioni precedenti, la matrice  $\Sigma_{m \times m}$ :

- e' simmetrica (diagonale contiene varianze)
- e' semi-positiva definita
- diventa diagonale se, 2 coppie,  $x_j \perp x_h$  con  $j \neq h$

Nel caso semplice  $m=2$  (caso bivariato), la seguente rappresentazione grafica assume rilevanza:



**Note:** Le densità in rosso e blu sono le densità marginali della distribuzione congiunta, quest'ultima rappresentata mediante i punti sul piano cartesiano  $S_x \times X_y$  dei supporti delle singole distribuzioni. La covarianza è rappresentata dall'ellisse di colore verde. Il valore della covarianza in questo caso è negativo.  
Fonte: [https://en.wikipedia.org/wiki/Multivariate\\_normal\\_distribution](https://en.wikipedia.org/wiki/Multivariate_normal_distribution)

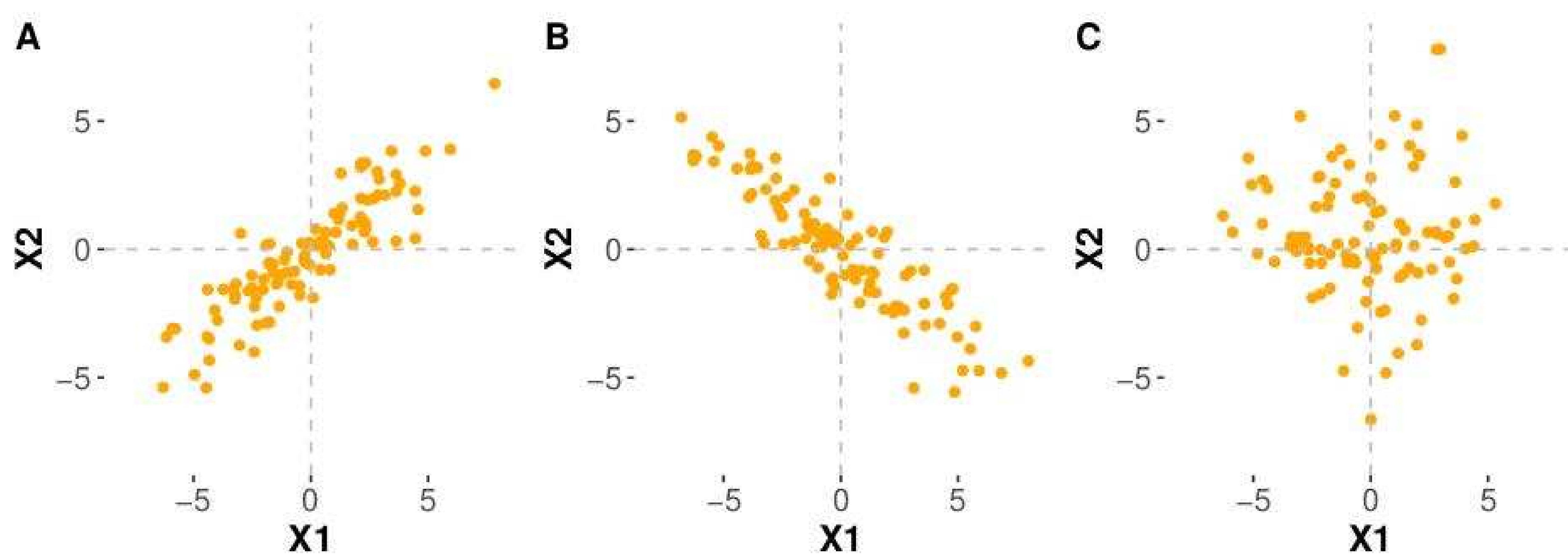


Fonte: [http://ballistipedia.com/index.php?title=Closed\\_Form\\_Precision](http://ballistipedia.com/index.php?title=Closed_Form_Precision)

La densità Normale  $m$ -variata ha la seguente forma:

$$f_{x_1 \dots x_m}(x; \mu, \Sigma) = \frac{1}{\sqrt{\det(\Sigma)}} \exp \left\{ -\frac{1}{2} \underbrace{(x - \mu)^T \Sigma^{-1} (x - \mu)}_{\text{forma quadratica}} \right\}$$





**Caso bivariato:** Rappresentazione grafica dei valori osservati congiuntamente delle variabili  $X_j$  e  $X_p$ .

**A:** Grafico con associazione lineare positiva  $\sigma_{X_j X_p} = 6.003$ .

**B:** Grafico con associazione lineare negativa  $\sigma_{X_j X_p} = -6.003$ .

**C:** Grafico con nessuna associazione lineare  $\sigma_{X_j X_p} \approx 0$ .

In tutti e tre i grafici  $\mu_{X_j} = \mu_{X_p} = 0$  e varianze  $\sigma_{X_j}^2 = 9.140$  e  $\sigma_{X_p}^2 = 5.165$ .

In figura sopra, si notano le associazioni marginali tra  $X_1$  e  $X_2$  con tre casi notevoli:

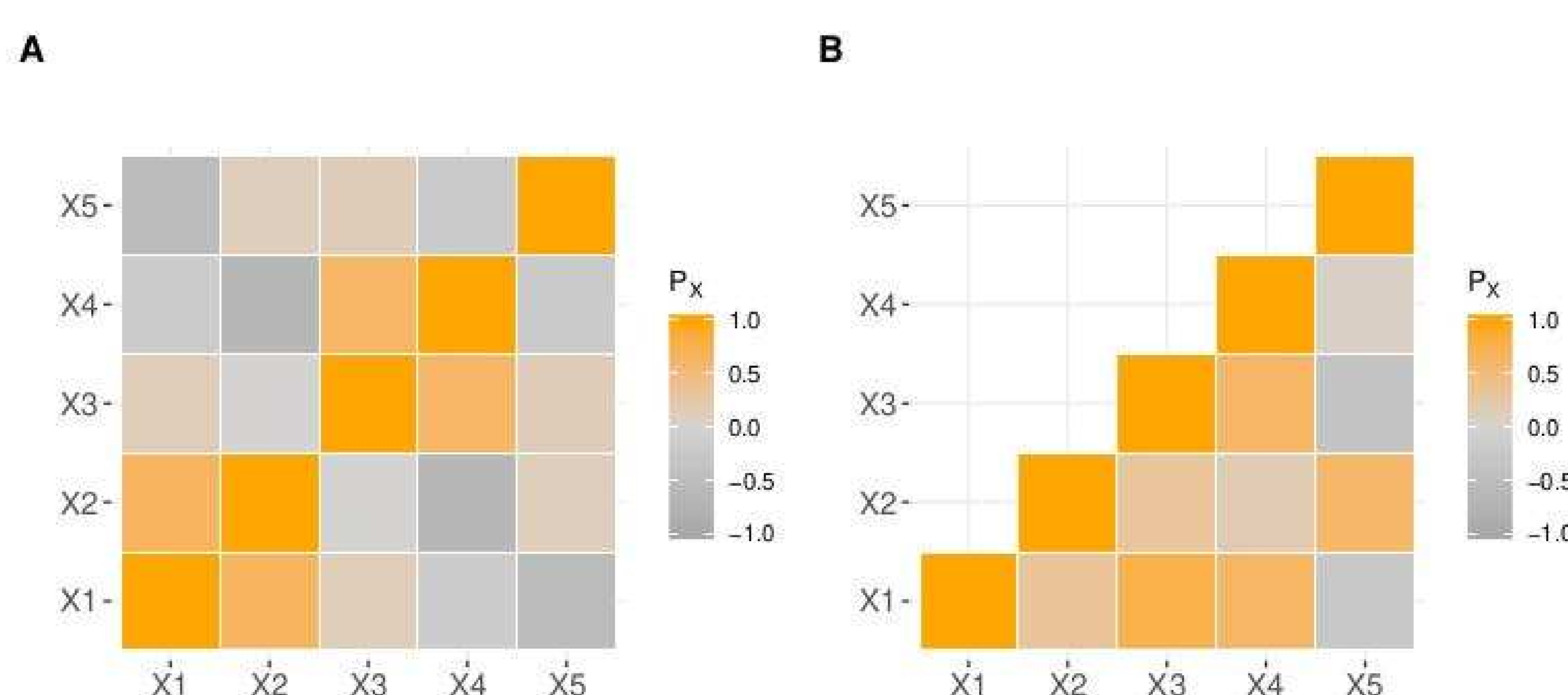
(A) covarianza positiva, (B) covarianza negativa, (C) covarianza nulla.

Spesso nel modello Normale multivariato è utile rappresentare  $\Sigma$  come matrice di correlazione  $R_{m \times m}$ . Il famoso elemento fuori dalla diagonale (che in questo caso è un vettore di 1) è il seguente:

$$\rho_{jh} = \text{Cor}[X_j, X_h] = \frac{\text{Cov}[X_j, X_h]}{\sqrt{\text{Var}[X_j]} \cdot \sqrt{\text{Var}[X_h]}}$$

Le proprietà di  $R$  sono le medesime di  $\Sigma$ .

Una possibile rappresentazione grafica per  $R$  è la seguente:



**Grafico della matrice di correlazione:** Rappresentazione grafica di una matrice di correlazione  $P$  mediante gradiente dei colori:  $\rho_{jp} \approx -1$  (grigio scuro),  $\rho_{jp} \approx 0$  (grigio chiaro),  $\rho_{jp} \approx 1$  (arancione). **A:** matrice di correlazione piena (parte triangolare superiore e inferiore). **B:** matrice di correlazione senza parte triangolare superiore. Si nota che la diagonale è di colore arancione poiché  $\text{diag}[P] = 1$ .

## Risultati notevoli del modello Normale m-variato

L'uso del modello Normale m-variato consente di calcolare con una certa semplicità le quantità che seguono. Consideriamo una collezione di v.z. distribuite come segue:

$$\left( \underbrace{X_1, \dots, X_m}_A, \underbrace{X_{m+1}, \dots, X_p}_B \right) \sim N_p(x; \mu, \Sigma) ; \mu \in \mathbb{R}^p, \Sigma \in \mathcal{S}^{p \times p}$$

Consideriamo due blocchi,  $(X_1, \dots, X_m)$  e  $(X_{m+1}, \dots, X_p)$  dal vettore originario (una partizione). Allora:

Nota: se  $X_A \perp X_B$   
allora  $\Sigma_{AB} = \Sigma_{BA} = 0$

$$\left( \underbrace{(X_1, \dots, X_m)}_A, \underbrace{(X_{m+1}, \dots, X_p)}_B \right) \sim N_p(x; \begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix}, \begin{pmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_{BB} \end{pmatrix})$$

Si ha che la densità marginale di A è  $f_{X_A} \sim N_m(x; \mu_A, \Sigma_{AA})$ . Analogamente per B.

La densità condizionata di A dato B (cioè:  $X_A|X_B$ ) è:

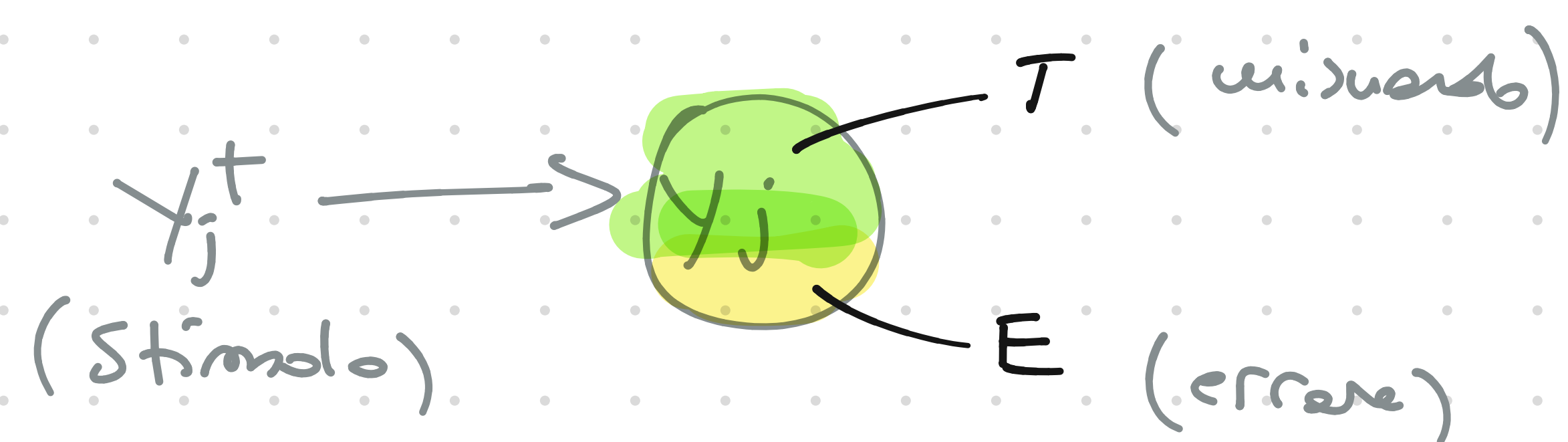
$$f_{X_A|X_B}(x; \underbrace{\mu_A + \Sigma_{AB} \Sigma_{BB}^{-1} (x_B - \mu_B)}_{\mu_{A|B}}, \underbrace{\Sigma_{AA} - \Sigma_{AB} \Sigma_{BB}^{-1} \Sigma_{BA}}_{\Sigma_{A|B}})$$



Rif: Barbarnelli 2.1-2.5, 2.8-2.9

Consideriamo un insieme di variabili discrete (di cardinalità  $m$ ) che codificano item a cui un insieme  $n$  di rispondenti (unità stats) reagisce. La teoria classica dei test (TCT) formula quanto segue:

$$Y_j = T_j + E_j$$



dove:  $Y_j$  è la  $j$ -esima v.z. codificante lo stimolo

$T_j$  è la  $j$ -esima v.z. rappresentante il misurando (cio' che si vuole misurare)

$E_j$  è la  $j$ -esima v.z. codificante l'errore casuale di misurazione

Nota:  $T_1, \dots, T_m$  son v.z. che rappresentano la parte non osservata.

Assunzioni:

- ①  $E[E_j] = 0$  (errore casuale, non sistematico)
- ②  $\text{Cor}[E_j, T_j] = 0$
- ③  $\text{Cor}[E_j, E_{j'}] = 0$  ( $j \neq j'$ )

Risultati notevoli:

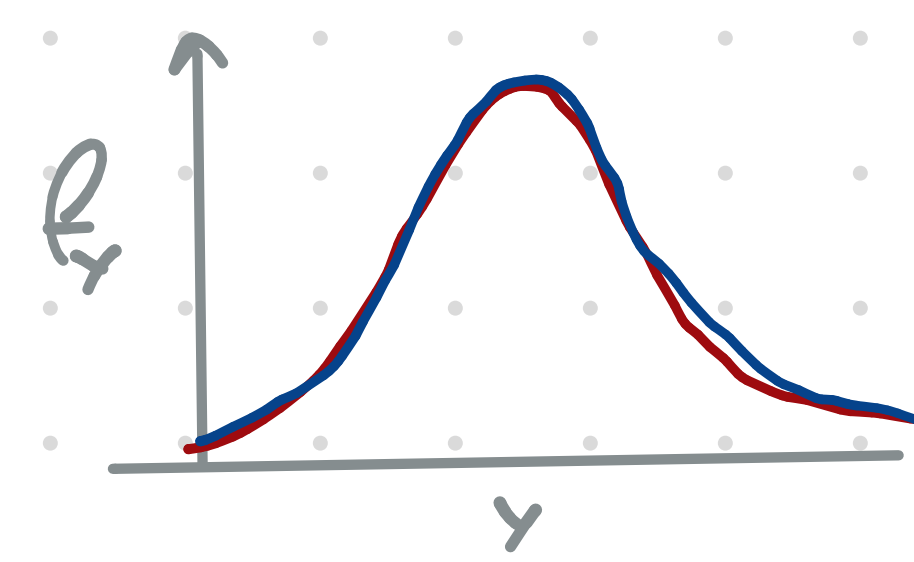
- ④  $E[Y_j] = E[T_j]$
- ⑤  $V[Y_j] = V[T_j] + V[E_j]$
- ⑥  $V[T_j] = \text{Cor}[Y_j, Y_{j'}] V[Y_j]$  ( $j \neq j'$ ,  $Y_j$  e  $Y_{j'}$  parallele)
- ⑦  $V[E_j] = V[Y_j] (1 - \text{Cor}[Y_j, Y_{j'}])$

$Y_j, Y_{j'}$  si dicono parallele:

$$E[Y_j] = E[Y_{j'}]$$

$$V[Y_j] = V[Y_{j'}]$$

Es:  $Y_j \sim N(\mu_0, \sigma_0^2)$   
 $Y_{j'} \sim N(\mu_0, \sigma_0^2)$



Attendibilità

Date due misure  $Y_j$  e  $Y_{j'}$  parallele, l'attendibilità  $\rho_{TY} \in [0, 1]$  è una q.ta' reale che esprime l'intensità con cui  $Y_j$  quantifica  $T$ .  $\rho_{TY}$  è legato alla precisione con cui  $Y_j$  quantifica  $T$ , come segue:

$$\begin{aligned} \text{Cor}[Y_j, Y_{j'}] &= \frac{\text{Cov}[Y_j, Y_{j'}]}{V[Y_j]^{1/2} V[Y_{j'}]^{1/2}}, & \text{dove } Y_j = T + E_j \text{ e } Y_{j'} = T + E_{j'} \text{ (nota: } T \text{ comune a } Y_j \text{ e } Y_{j'}) \\ &= \frac{\text{Cov}[T + E_j, T + E_{j'}]}{V[Y_j]} \\ &= \frac{\text{Cov}[T, T] + \text{Cov}[T, E_j] + \text{Cov}[T, E_{j'}] + \text{Cov}[E_j, E_{j'}]}{V[Y_j]} \\ &= \frac{V[T] + 0 + 0 + 0}{V[Y_j]} \\ &= \frac{V[T]}{V[Y_j]} \end{aligned}$$

$$\rho_{TY} = \frac{V[T]}{V[Y_j]} \quad \left( \text{oppure: } 1 - \frac{V[E_j]}{V[Y_j]} \right)$$

La q.ta'  $\rho_{TY} \in [0, 1]$  indica quanto  $Y$  è preciso nel misurare  $T$ . Quando  $\rho_{TY} \rightarrow 1$  allora la coppia  $(Y_j, Y_{j'})$  è massimamente attendibile nel quantificare  $T$  (massimamente precisa).

Nota: Consideriamo una v.z. reale  $X$  che ha come parametro di locazione vero  $\theta_0$  (media).

Successivamente, denotiamo con  $x = (x_1, \dots, x_n)$  un campione casuale di misurazioni di  $X$ . Dunque:

(i) l'ACCURATEZZA è in relazione alla differenza tra  $E[x]$  e  $\theta_0$

(ii) la PRECISIONE è in relazione alla varianza  $V[x]$ , in particolare  $V[x]^{-1}$



per calcolare  $\hat{P}_Y$  necessitiamo di  $Y_j, Y_{j'}$  parallele (caso poco frequente). Empiricamente possiamo ottenere  $Y_j, Y_{j'}$  parallele agendo sul c.d. disegno della misurazione:

- ① Rilevare  $Y_j, Y_{j'}$  sullo stesso campione in tempi diversi:  $\{(Y_j, Y_{j'})_{t_1}, (Y_j, Y_{j'})_{t_2}\}$
- ② Rilevare  $Y_j, Y_{j'}$  sullo stesso campione con strumenti equivalenti

Per stimare  $\hat{P}_{YT}$  mediante un approccio statistico possiamo usare procedure che richiedono:

- ① Due somministrazioni del test
  - metodo delle forme parallele
  - metodo test-retest
- ② Una sola somministrazione del test
  - metodo split-half
  - metodo basato sulla covarianza tra le misure (coerenza interna)

La tabella che segue riassume alcune procedure per il calcolo dell'attendibilità:

Major Error Source	Reliability Coefficient	Data-Collection Procedure	Statistical Treatment of Data
1. Change in examinee's overtime	1. Stability coefficient	1. Test, wait, retest	1. Compute Pearson product moment coefficient, $\hat{\rho}_{12}$
2. Content sampling from form to form	2. Equivalence coefficient	2. Give form 1, give form 2	2. Compute Pearson product moment coefficient, $\hat{\rho}_{12}$
3. Content sampling, or flawed items	3. Internal consistency coefficient	3. Give one form on one occasion	3a. Divide test into halves; correlate half-test, $\hat{\rho}_{AB}$ ; use Spearman Brown correction: $\hat{\rho}_{XX'} = \frac{2\hat{\rho}_{AB}}{1 + \hat{\rho}_{AB}}$ b. Divide test into halves; use Guttman's or Rulon's formula: $\hat{\rho}_{XX'} = 1 - \frac{\sigma_D^2}{\sigma_X^2}$ c. Compute item variances; compute coefficient alpha: $\hat{\alpha} = \frac{k}{k-1} \left( 1 - \frac{\sum \sigma_i^2}{\sigma_X^2} \right)$

### • Split-half

Consideriamo una collezione di v.z.  $\mathcal{Y} = (Y_1, \dots, Y_m)$  e procediamo individuando una partizione interna di tale collezione,  $\mathcal{Y}_A$  e  $\mathcal{Y}_B$  (esg.  $\mathcal{Y}_A \cup \mathcal{Y}_B = \mathcal{Y}$ ). Denotiamo con  $Y_A^{tot}$  e  $Y_B^{tot}$  le v.z. risultanti dalla somma delle v.z. di  $\mathcal{Y}_A$  e  $\mathcal{Y}_B$ :

$$Y_A^{tot} = \sum_{j \in \mathcal{Y}_A} Y_j, \quad Y_B^{tot} = \sum_{h \in \mathcal{Y}_B} Y_h$$

$$\hat{P}_{YT} = \frac{2 \text{Cor}[Y_A^{tot}, Y_B^{tot}]}{1 + \text{Cor}[Y_A^{tot}, Y_B^{tot}]}$$

$$\left( \text{opp. } \hat{P}_{YT}^2 = \frac{2 P_{AB}^2}{1 + P_{AB}^2} \right)$$

Nota:  $Y_1, \dots, Y_m$  sono parallele.



Esempio (10 partecipanti, 6 item):

		variabili/item $j$								
		item 1	item 2	item 3	item 4	item 5	item 6	$y_{totA}$	$y_{totB}$	$y_{tot}$
unibi stats  $i$	1	0	0	0	0	0	0	0	0	0
	2	0	0	0	0	1	0	1	0	1
	3	1	0	1	1	1	0	3	1	4
	4	1	1	1	1	1	1	3	3	6
	5	1	1	1	1	1	1	3	3	6
	6	0	0	1	0	0	0	1	0	1
	7	0	0	1	1	1	0	2	1	3
	8	0	0	0	1	0	0	0	1	1
	9	1	0	1	1	1	0	3	1	4
	10	0	1	0	1	0	1	0	3	3

$J_A = \{1, 3, 5\}$

$J_B = \{2, 4, 6\}$

$$\hat{\rho}_Y^2 = \frac{2\hat{\rho}_{AB}^2}{1 + \hat{\rho}_{AB}^2} = \frac{2 \cdot 0.34}{1 + 0.34} = 0.507$$

Nel caso in cui  $(y_1, \dots, y_m)$  sono non parallele,  $\hat{\rho}_{YT}$  può essere calcolato tramite indice di Rulon:

$$\hat{\rho}_{YT} = 2 \left( 1 - \frac{\psi[y_A^{Tot}] + \psi[y_B^{Tot}]}{\psi[y^{Tot}]} \right), \quad \text{dove: } y^{Tot} = \sum_{j=1}^m y_j$$

Esempio (10 partecipanti, 6 item):

	item 1	item 2	item 3	item 4	item 5	item 6	$y_{totA}$	$y_{totB}$	$y_{tot}$
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	0	1	0	1
3	1	0	1	1	1	0	3	1	4
4	1	1	1	1	1	1	3	3	6
5	1	1	1	1	1	1	3	3	6
6	0	0	1	0	0	0	1	0	1
7	0	0	1	1	1	0	2	1	3
8	0	0	0	1	0	0	0	1	1
9	1	0	1	1	1	0	3	1	4
10	0	1	0	1	0	1	0	3	3

$$\hat{\rho}_Y^2 = 2 \left( 1 - \frac{\sigma_{yA}^2 + \sigma_{yB}^2}{\sigma_{yTot}^2} \right) = 2 \left( 1 - \frac{1.82 + 1.56}{4.54} \right) = 0.508$$

- Coerenza interna

Consideriamo  $(y_1, \dots, y_m)$  e calcoliamo  $y^{Tot} = \sum_{j=1}^m y_j$  (misura composita). È possibile stimare  $\hat{\rho}_{YT}$  usando il c.d. di Cronbach:

$$\hat{\rho}_{YT} = \frac{m}{m-1} \left( 1 - \frac{\sum_{j=1}^m \psi[y_j]}{\sum_{j=1}^m \psi[y_j] + \sum_{j \neq h} \text{cov}[y_j, y_h]} \right)$$

Nota: la denominazione 'coerenza interna' deriva dall'uso della p.te indicata nel box rosso, ossia la covarianza tra le misurazioni.

In alternativa:

$$\hat{\rho}_{YT} = \frac{m}{m-1} \left( 1 - \frac{\sum_j \text{diag}(S)_i}{\sum_j \text{diag}(S)_i + \sum_j \text{tril}(S)_i} \right), \quad S_{m \times m} = \begin{pmatrix} \sigma_{11}^2 & & \\ \sigma_{12}^2 & \sigma_{22}^2 & \\ \vdots & \vdots & \ddots \\ \sigma_{m1}^2 & \dots & \dots & \sigma_{mm}^2 \end{pmatrix} \quad \text{matrice di coerenza di } Y_{m \times m}$$

Nota:  $y_1, \dots, y_m$  sono v.e. reali non necessariamente parallele.

Quando  $y_1, \dots, y_m$  sono v.e. di tipo dicotomica (ie,  $y_j \in \{0, 1\}$ ), l'attendibilità  $\hat{\rho}_{YT}$  è stimata mediante calcolo offerto da Kuder-Richardson (KR-20):

$$\hat{\rho}_{YT} = \frac{m}{m-1} \left( 1 - \frac{\sum_{j=1}^m \pi_j (1-\pi_j)}{\psi[y^{Tot}]} \right) \quad \text{dove } \pi_j \doteq P[y_j = 1]$$

$y_j \sim \text{Bern}(y; \pi_j)$



Esempio (10 partecipanti, 6 item):

	item 1	item 2	item 3	item 4	item 5	item 6	$y_{totA}$	$y_{totB}$	$y_{tot}$
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	0	1	0	1
3	1	0	1	1	1	0	3	1	4
4	1	1	1	1	1	1	3	3	6
5	1	1	1	1	1	1	3	3	6
6	0	0	1	0	0	0	1	0	1
7	0	0	1	1	1	0	2	1	3
8	0	0	0	1	0	0	0	1	1
9	1	0	1	1	1	0	3	1	4
10	0	1	0	1	0	1	0	3	3
$\pi$	0.4	0.3	0.6	0.7	0.6	0.3			

$$\hat{\rho}^2_{XX'} = \frac{m}{m-1} \left( 1 - \frac{\sum_{j=1}^m \pi_j (1 - \pi_j)}{\sigma^2_{y_{tot}}} \right) = \frac{6}{6-1} \left( 1 - \frac{1.35}{4.54} \right) = 0.703$$

Un altro indice per stimare  $\rho_{YT}$  e' la c.d. correlazione intraclass:

$$\hat{\rho}_{YT} = \sqrt{V[T] \cdot (V[T] + V[E])^{-1}}$$

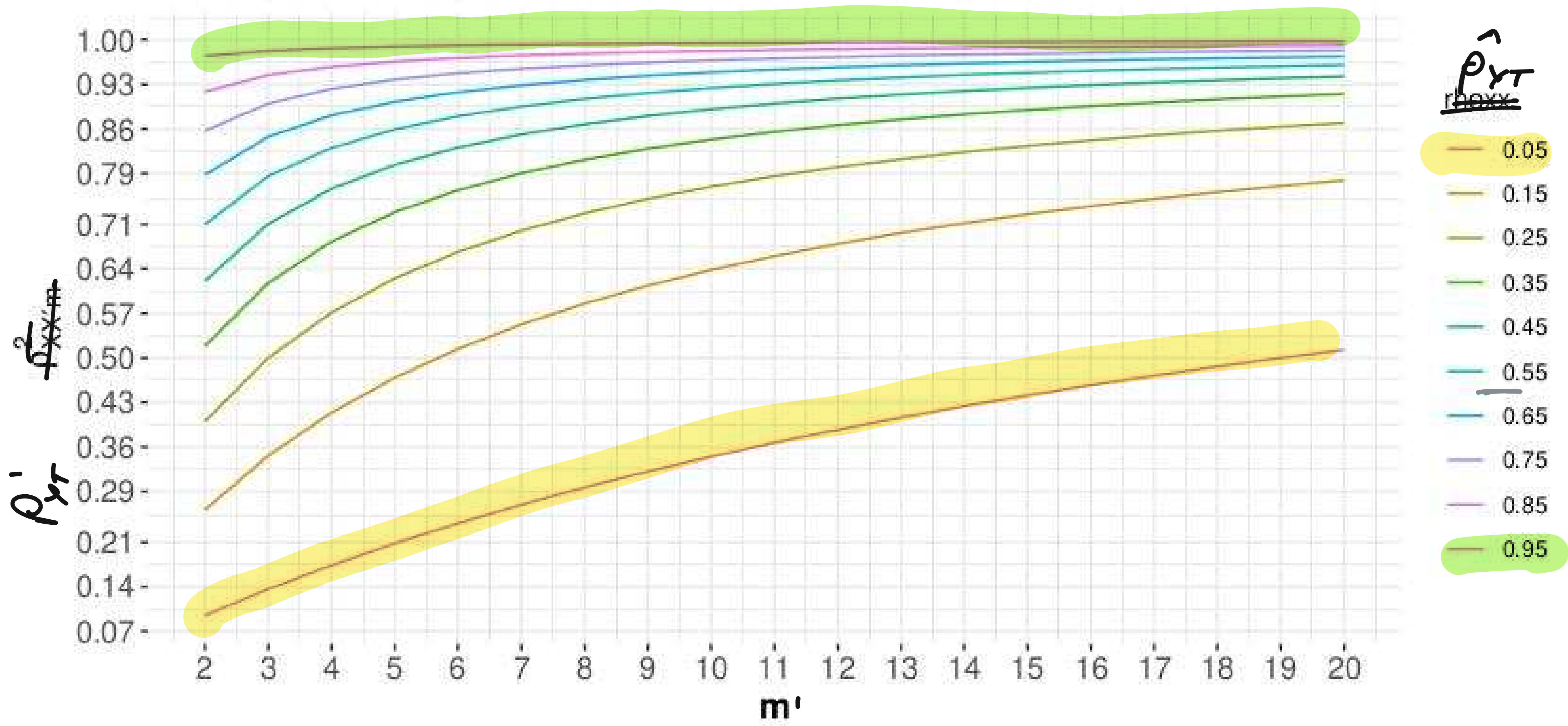
Esempio (10 partecipanti, 6 item):

	item 1	item 2	item 3	item 4	item 5	item 6	$y_{totA}$	$y_{totB}$	$y_{tot}$
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	0	1	0	1
3	1	0	1	1	1	0	3	1	4
4	1	1	1	1	1	1	3	3	6
5	1	1	1	1	1	1	3	3	6
6	0	0	1	0	0	0	1	0	1
7	0	0	1	1	1	0	2	1	3
8	0	0	0	1	0	0	0	1	1
9	1	0	1	1	1	0	3	1	4
10	0	1	0	1	0	1	0	3	3

$$\hat{\rho}^2_{YT} = \frac{\sigma^2_T}{\sigma^2_T + \sigma^2_E} = \hat{\rho}^2_{YT} = \frac{1.787^2}{1.787^2 + 1.161^2} = 0.703$$

Relazione tra  $\rho_{YT}$  e  $m$

Consideriamo  $y = (y_1, \dots, y_m)$  v.a. e  $y_{TOT} = \sum_j y_j$  la quantita' aleatoria che forma il c.d. punteggio totale. Supponiamo di voler aggiungere  $m'$  misure alla collezione  $y$  (dello stipo tipo). Quanti item/misure  $m'$  occorre aggiungere a  $y$  affine' di raggiungere una certa attendibilita'? La risposta e' data dalla c.d. formula profetica di Spearman-Brown:



$$\hat{\rho}^1_{YT} = \frac{m' \cdot \hat{\rho}_{YT}}{1 + (m' - 1) \hat{\rho}_{YT}}$$

dove:  $m'$  e' il numero di item da aggiungere  
 $\hat{\rho}_{YT}$  e' l'attendibilita' iniziale del test.

**Attendibilita' di una misura composta e lunghezza:** Curve di attendibilita' di una misura composta (o test) quando questa e' allungata per un intero  $m$ . Le curve sono in funzione dell'attendibilita' iniziale  $\rho^2_{XX'}$  della misura composta (colore delle curve) e mostrano come questa, tenendo fisso tale valore, cambia in precisione  $\rho^2_{XX'm}$  (ordinata) quando  $m$  nuove misure componenti (ascissa) sono aggiunte a formare il test.



Stima dei punteggi finali ad un test

Consideriamo  $y_1, \dots, y_m$  v.z. (possibilmente parallele) secondo lo schema:

$$(y_1 = T + E_1, \dots, y_j = T + E_j, \dots, y_m = T + E_m)$$

Due possibili stimatori di  $T$  sono i seguenti:

- 1)  $\hat{T} = \sum_j y_j$  (somma)
- 2)  $\hat{T} = \sum_j y_j w_j$ ,  $w_j$  è una quantità reale positiva (somma pesata)

Usando (2), possiamo specificare  $w_1, \dots, w_m$  come segue:

$$\hat{T} = \hat{\rho}_{YT} \sum_j y_j + (1 - \hat{\rho}_{YT}) \frac{1}{m} \sum_j y_j$$

dove  $\hat{\rho}_{YT}$  è l'attendibilità della collezione  $Y$ . Si verifica quanto segue:

- (i)  $\hat{\rho}_{YT} \rightarrow 1$ :  $\hat{T}$  stimato usando  $y^{TOT}$
- (ii)  $\hat{\rho}_{YT} \rightarrow 0$ :  $\hat{T}$  stimato usando  $\frac{1}{m} y^{TOT}$  (media)

Esempio (10 partecipanti, 6 item):

	item 1	item 2	item 3	item 4	item 5	item 6	$y_{totA}$	$y_{totB}$	$y_{tot}$	$\tau$
1	0	0	0	0	0	0	0	0	0	0.861
2	0	0	0	0	1	0	1	0	1	1.564
3	1	0	1	1	1	0	3	1	4	3.673
4	1	1	1	1	1	1	3	3	6	5.079
5	1	1	1	1	1	1	3	3	6	5.079
6	0	0	1	0	0	0	1	0	1	1.564
7	0	0	1	1	1	0	2	1	3	2.970
8	0	0	0	1	0	0	0	1	1	1.564
9	1	0	1	1	1	0	3	1	4	3.673
10	0	1	0	1	0	1	0	3	3	2.973

$\hat{\rho}_{YT} = 0.703$

punteggi grezzi

punteggi stimati

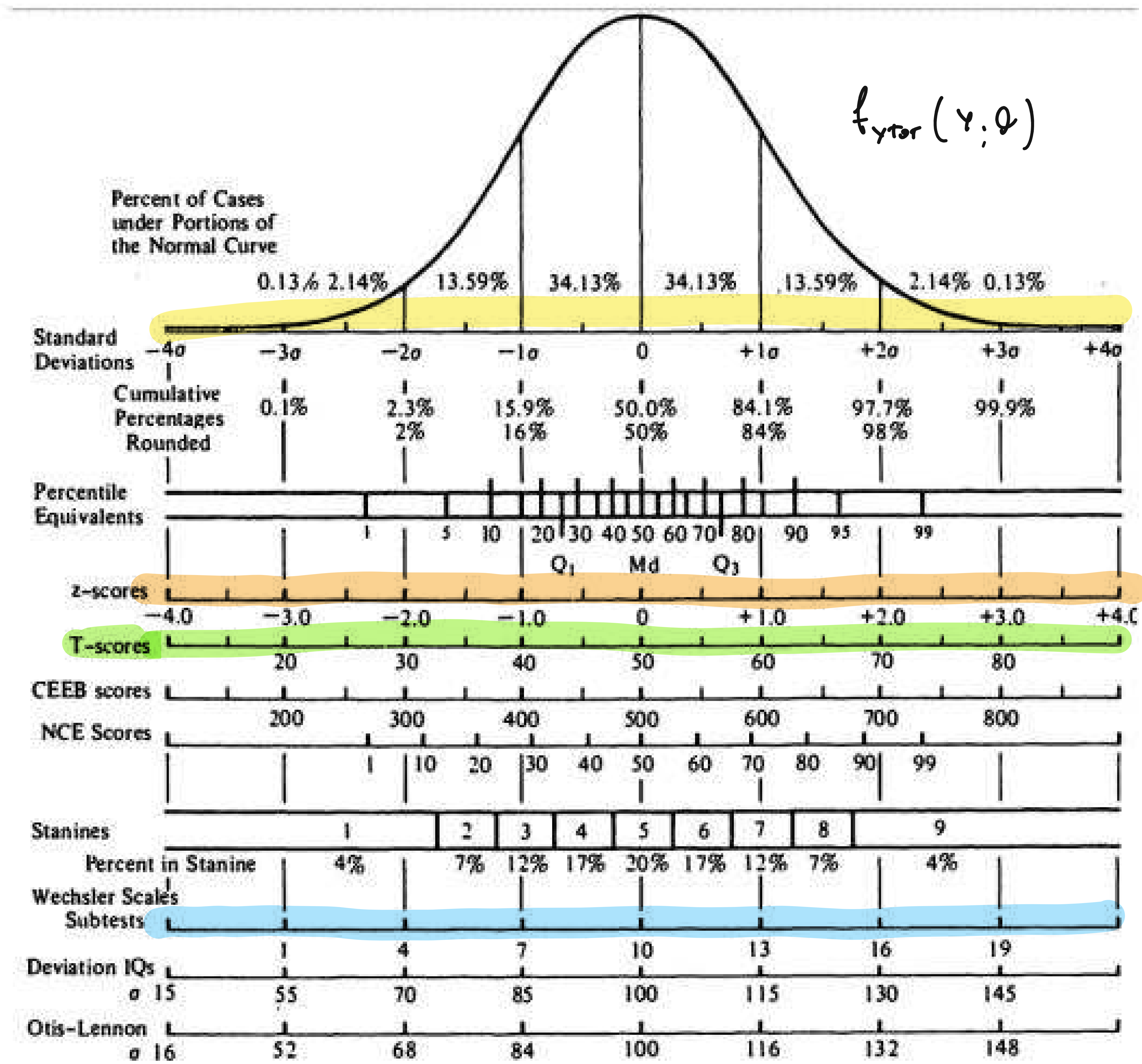
Esempio (10 partecipanti, 6 item):

	item 1	item 2	item 3	item 4	item 5	item 6	$y_{totA}$	$y_{totB}$	$y_{tot}$	$\tau$
1	0	0	0	0	0	0	0	0	0	2.871
2	0	0	0	0	1	0	1	0	1	2.881
3	1	0	1	1	1	0	3	1	4	2.911
4	1	1	1	1	1	1	3	3	6	2.931
5	1	1	1	1	1	1	3	3	6	2.931
6	0	0	1	0	0	0	1	0	1	2.881
7	0	0	1	1	1	0	2	1	3	2.901
8	0	0	0	1	0	0	0	1	1	2.881
9	1	0	1	1	1	0	3	1	4	2.911
10	0	1	0	1	0	1	0	3	3	2.901

$\hat{\rho}_{YT} = 0.01$



Spesso c'è utile trasformare  $y_{TOT}$  secondo una scala che ne favorisca una migliore interpretazione (6) o un migliore confronto (ad esempio con punteggi/soglie normative di riferimento). La figura che segue mostra un esempio di trasformazione dei punteggi osservati  $y_{TOT}$  in punteggi  $T$ ,  $Z$ , Wechsler, ecc.). Solitamente la trasformazione adottata dipende dallo specifico test in uso.



Maggiori info: Barb. 2.12, 2.12.3



La CFA è un modello psicometrico-statistico che propone di spiegare la variabilità tra le variabili osservate  $Y_1, \dots, Y_m$  introducendo una collezione di variabili "latenti"  $M_1, \dots, M_Q$ , ritenute precedenti rispetto alle variabili osservate.



CFA: variabili continue,  $(Y_1, \dots, Y_m) \in \mathbb{R}^m$   
 $(M_1, \dots, M_Q) \in \mathbb{R}^Q$

CFA nel caso di vars continue

CFA nel caso di vars categoriali

$(Y_1, \dots, Y_m) \in \mathcal{C} \subseteq \mathbb{N}^m$  variabili categoriali  
 $(M_1, \dots, M_Q) \in \mathbb{R}^Q$

CFA: 1) dopo aver eseguito un pretrattamento di  $(Y_1, \dots, Y_m) \leftarrow$  quantificazione  
 2) si calcolano statistiche idonee dai dati  $(Y_1, \dots, Y_m)$  e si esegue la CFA tradizionale

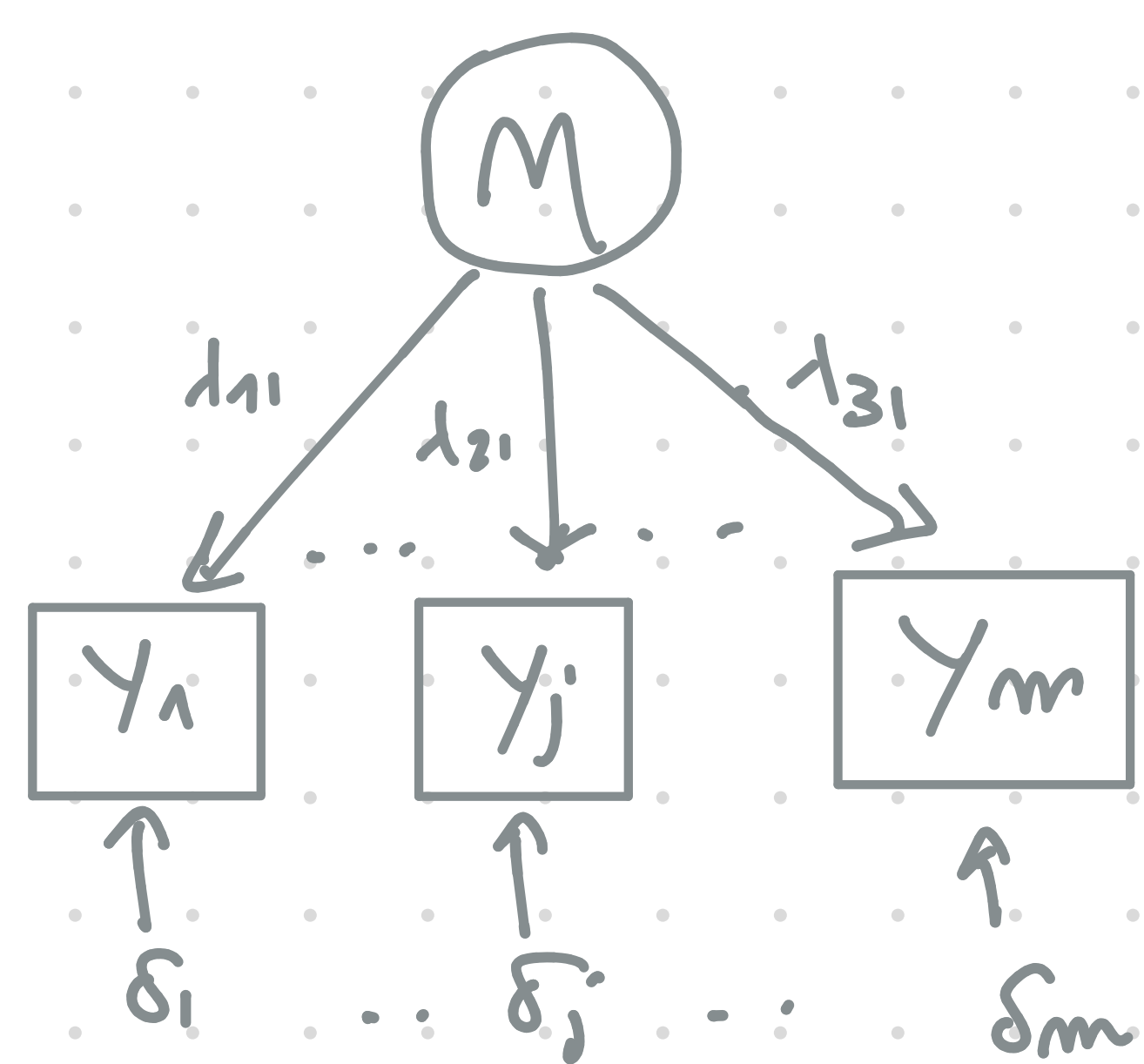
## Introduzione

Consideriamo:

$$Y_{m \times m} = \begin{pmatrix} Y_{11} & Y_{12} & \dots & Y_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{m1} & Y_{m2} & \dots & Y_{mm} \end{pmatrix}$$

Dopo esserci assicurati che ①  $m \gg m$ , ②  $\det(S) \neq 0 \Rightarrow r(S) = m$ , con  $S_{m \times m} = \text{Cov}[Y]$ , possiamo introdurre il modello della CFA come segue (modello unidimensionale):

$$Y_i = (Y_{i1}, \dots, Y_{ij}, \dots, Y_{im})$$



$$Y_{i1} = \gamma_1 + M_i \lambda_{11} + \delta_{i1}$$

$\vdots$

$$Y_{ij} = \gamma_j + M_i \lambda_{j1} + \delta_{ij}$$

$\vdots$

$$Y_{im} = \gamma_m + M_i \lambda_{m1} + \delta_{im}$$

$$\begin{matrix} Y_i & = & \gamma & + & \lambda M_i & + & \delta_i \\ (1 \times m) & & (1 \times m) & & (1 \times m)(1 \times 1) & & (1 \times m) \end{matrix}$$

dove:

$\gamma_j$  è l'intercetta del modello ("difficoltà dell'item")

$M_i$  è la v. latente detta "fattore" che sottende le misurazioni empiriche ("grado di discriminatività")

$\lambda_j$  è il coefficiente (peso) che lega la v. latente  $M$  all'osservato  $Y_j$  ("factor loading")

$\delta_{ij}$  è l'errore di misurazione (errore casuale)

Note:

(i) Il modello è lineare e additivo

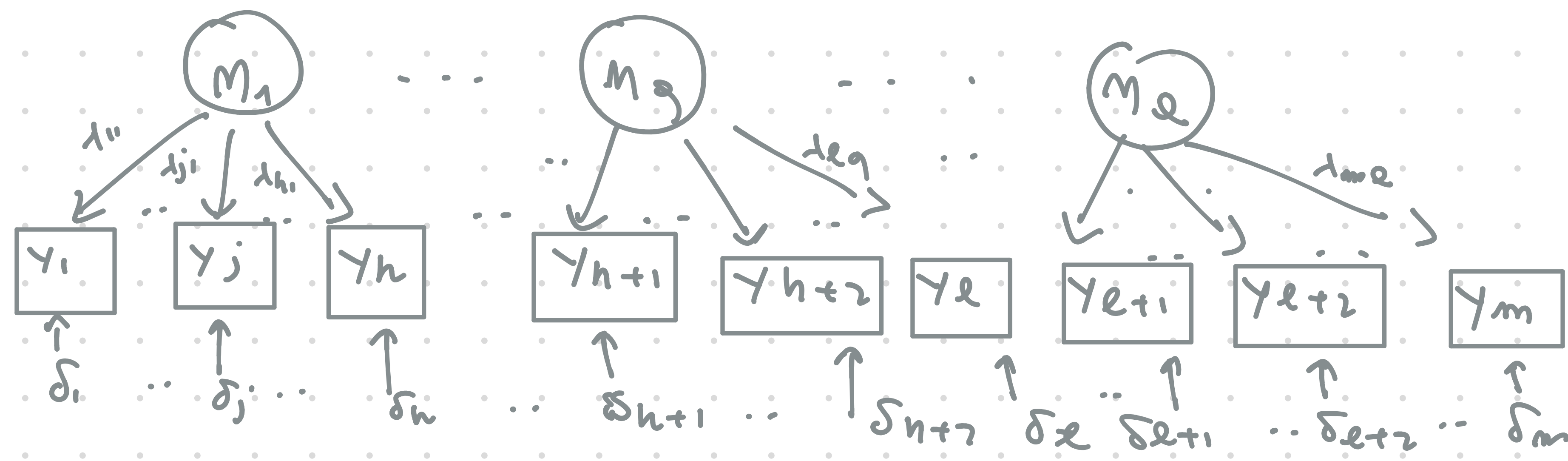
(ii) Il modello è unidimensionale poiché  $Q=1$  (una sola v. latente)

(iii) Parimenti alla TCT,  $Y_j$  è in dipendenza lineare da  $M$  ("modello riflessivo")



Il modello unidimensionale può essere generalizzato al caso  $Q > 1$ . In questo caso il modello è detto multidimensionale.

(2)



Nota: si omette i per semplicità notazionale.

$$\begin{aligned} Y_1 &= \tau_1 + \lambda_{11}\eta_1 + \delta_1 \\ &\vdots \\ Y_j &= \tau_j + \lambda_{j1}\eta_1 + \delta_j \\ &\vdots \\ Y_l &= \tau_l + \lambda_{lq}\eta_q + \delta_l \\ &\vdots \\ Y_m &= \tau_m + \lambda_{mq}\eta_q + \delta_m \end{aligned}$$

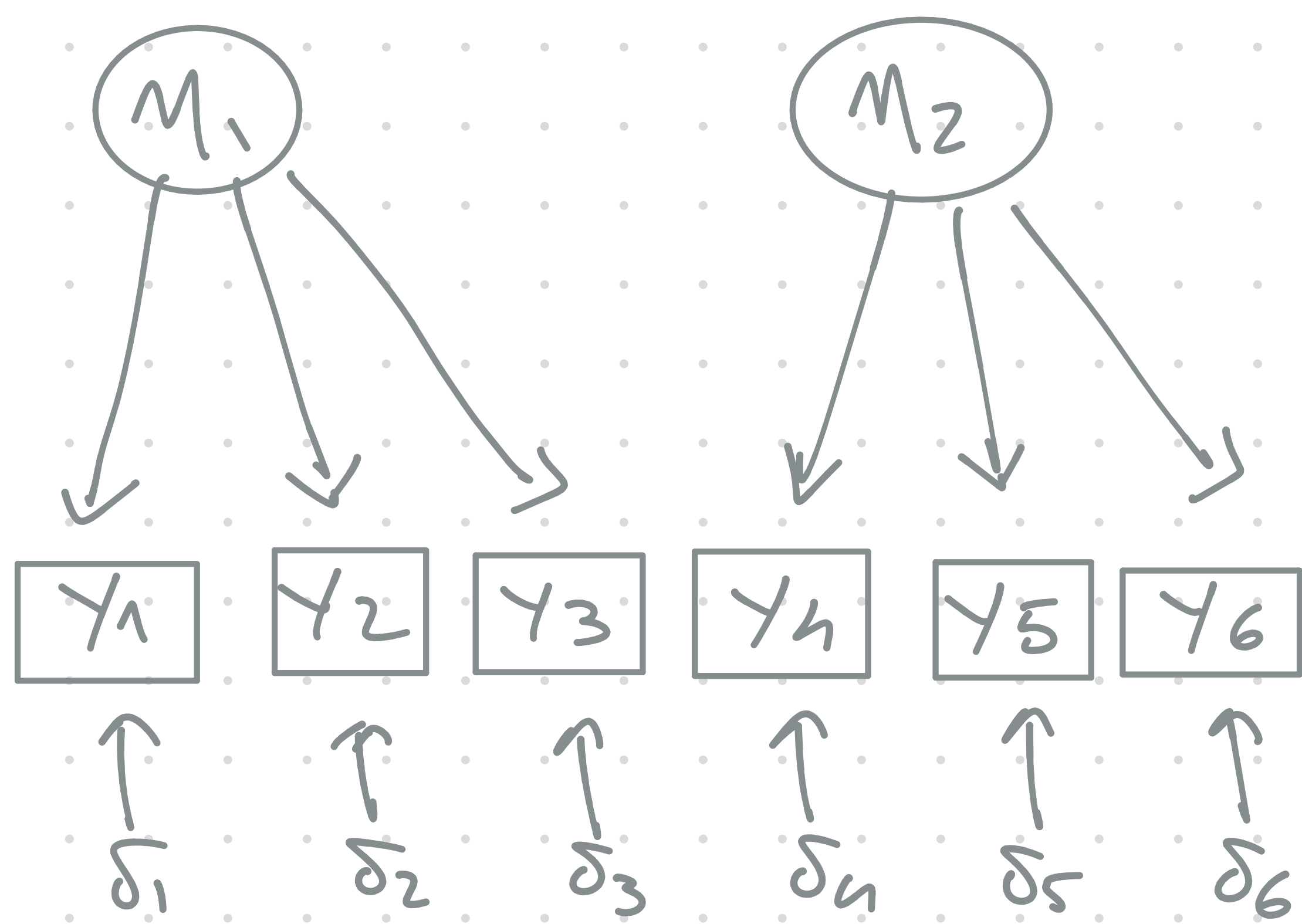
$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_m \end{pmatrix} = \begin{pmatrix} \tau_1 \\ \vdots \\ \tau_m \end{pmatrix} + \begin{pmatrix} \lambda_{11} & \dots & \lambda_{1q} & \dots & \lambda_{1Q} \\ \vdots & & \vdots & & \vdots \\ \lambda_{j1} & \dots & \lambda_{jq} & \dots & \lambda_{jQ} \\ \vdots & & \vdots & & \vdots \\ \lambda_{m1} & \dots & \lambda_{mq} & \dots & \lambda_{mQ} \end{pmatrix} \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_q \\ \vdots \\ \eta_Q \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \vdots \\ \delta_j \\ \vdots \\ \delta_m \end{pmatrix}$$

$$\underset{(m \times 1)}{Y} = \underset{(m \times 1)}{\tau} + \underset{(m \times Q)}{\Lambda} \underset{(Q \times 1)}{\eta} + \underset{(m \times 1)}{\delta}$$

Rappresentazione algebrica del modello

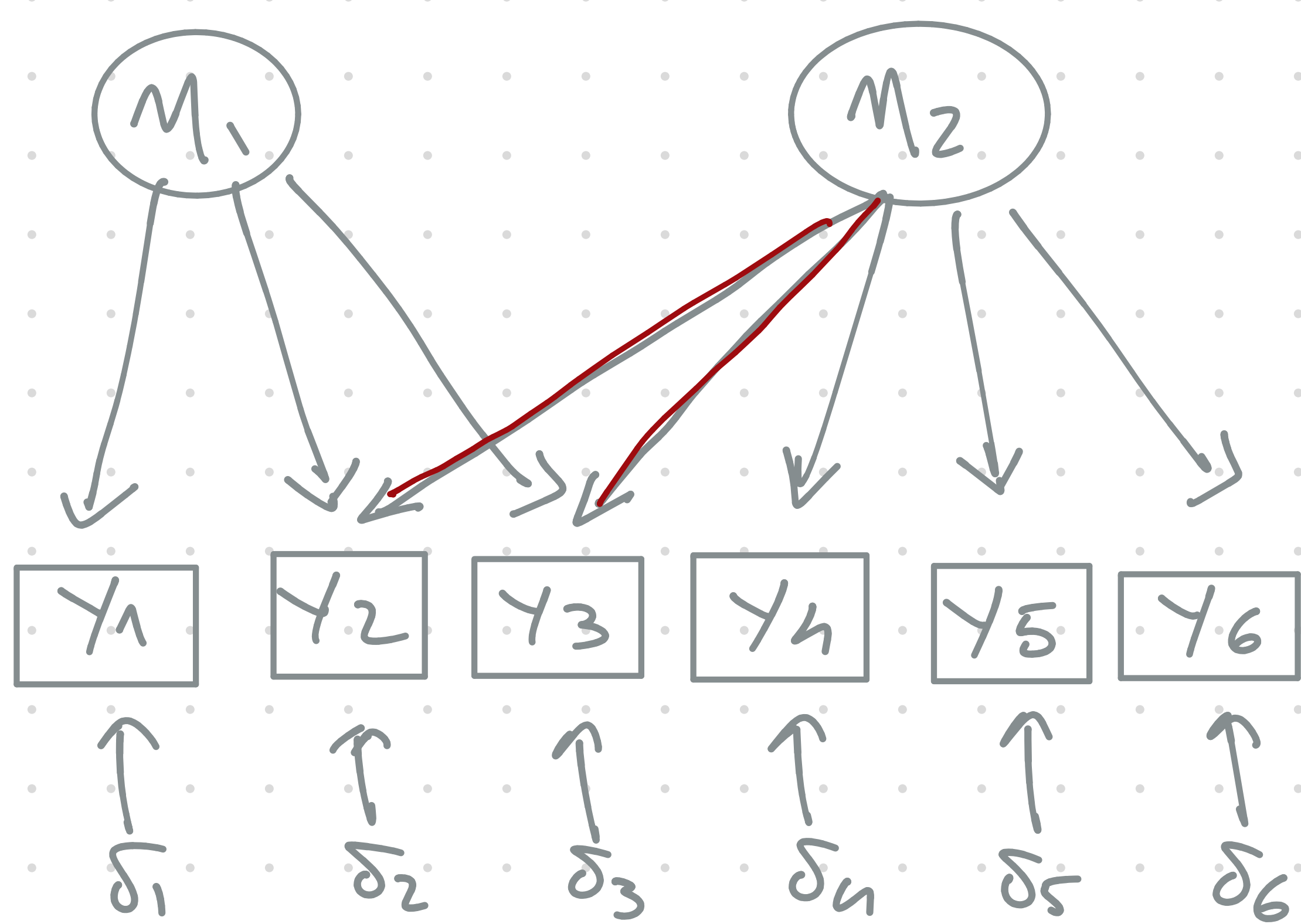
Casi notevoli di CFA ( $m=6, Q=2$ )

Es. 1 (golden standard sec. TCT)



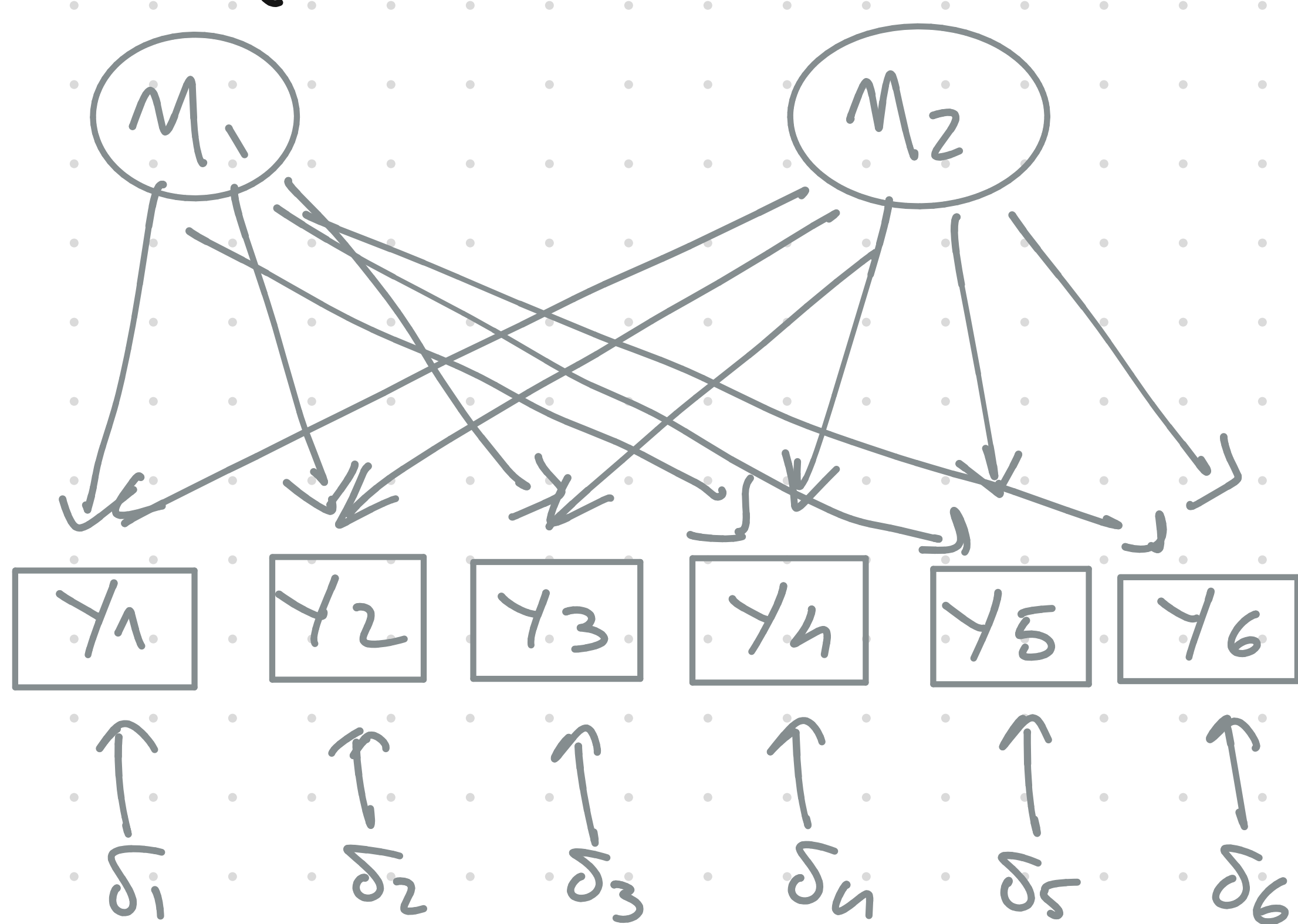
$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{pmatrix} = \begin{pmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \\ \tau_5 \\ \tau_6 \end{pmatrix} + \begin{pmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \end{pmatrix}$$

Es. 2 (cross-loadings)



$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{pmatrix} = \begin{pmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \\ \tau_5 \\ \tau_6 \end{pmatrix} + \begin{pmatrix} \lambda_{11} & 0 \\ \lambda_{21} & \lambda_{22} \\ \lambda_{31} & \lambda_{32} \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \end{pmatrix}$$

Es. 3 (EFA)



$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{pmatrix} = \begin{pmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \\ \tau_5 \\ \tau_6 \end{pmatrix} + \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \\ \lambda_{31} & \lambda_{32} \\ \lambda_{41} & \lambda_{42} \\ \lambda_{51} & \lambda_{52} \\ \lambda_{61} & \lambda_{62} \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \end{pmatrix}$$



Note:

(3)

La matrice  $\Lambda_{m \times Q}$  contiene le ipotesi del ricercatore circa la relazione tra le v. osservate e le v. latenti (fattori latenti).

In generale:

$\lambda_{jp} = 1^0 \Rightarrow$  v. osservata  $y_j$  riflette la var. latente  $\eta_p$  secondo la magnitudine  $1^0$

$1^0 \in \mathbb{R}_{+0}$ .  $\lambda_{jp}$  è un parametro fissato (può essere diverso da zero)

$$\Lambda_{3 \times 2} = \begin{pmatrix} 1^0 & 0.8 \\ 0 & 1.22 \\ 1.21 & 0 \end{pmatrix}$$

$\lambda_{jp} \neq 1^0 \Rightarrow$  v. osservata  $y_j$  riflette la v. latente  $\eta_p$

$\lambda_{jp}$  è un parametro libero

La matrice  $\Lambda_{m \times Q}$  esprime la conoscenza circa il modo con cui  $\eta_{1 \times Q}$  è misurato attraverso  $y_{1 \times m}$  (oppure: è riflesso in  $y_{1 \times m}$ )

tutto ciò premesso, misurare significa ipotizzare un modello di misura (CFA) e verificarlo rispetto a  $y_{m \times m}$  (matrice dei dati osservati).

## Struttura probabilistica del modello

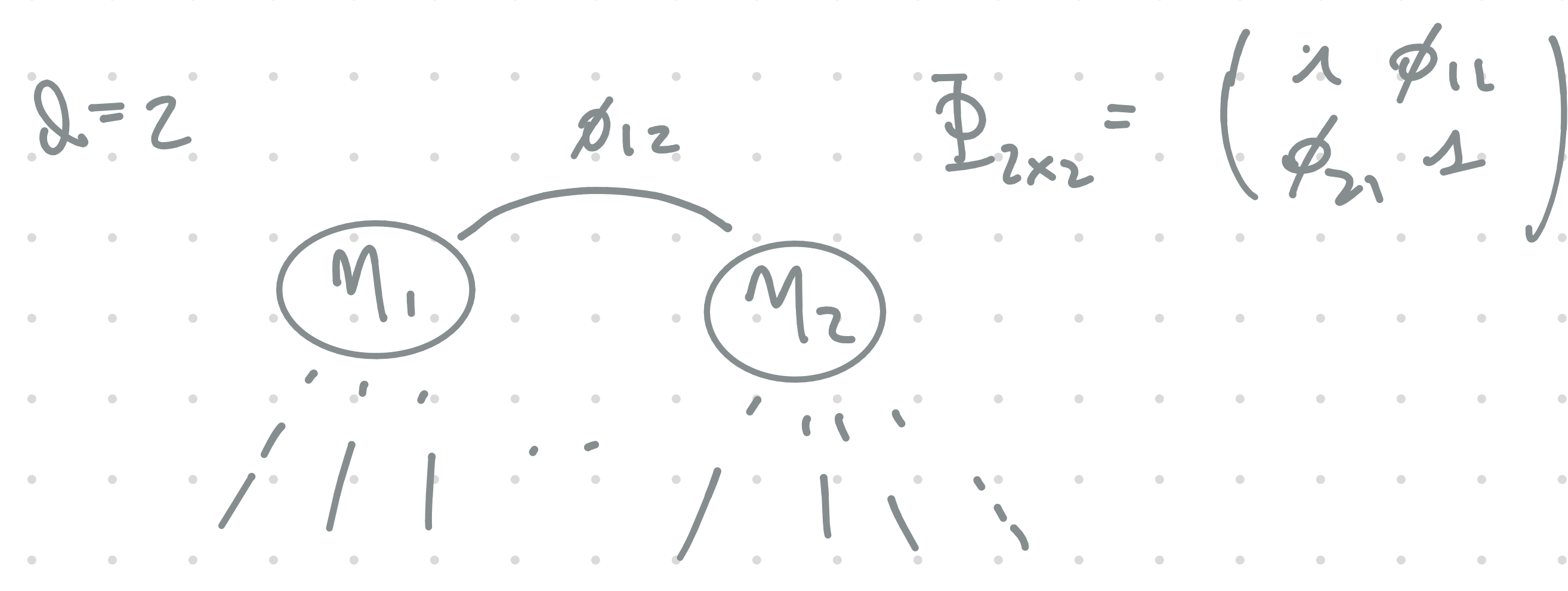
$$y_{m \times 1}^{(i)} = \gamma_{m \times 1} + \Lambda_{m \times Q} \eta_{Q \times 1}^{(i)} + \delta_{m \times 1}^{(i)} \quad i=1, \dots, n$$

$$1) \eta_{Q \times 1}^{(i)} = (\eta_1, \dots, \eta_Q)^T \sim N_Q(i; \mu_\eta, \Phi)$$

$\mu_\eta$ : vettore  $Q \times 1$  delle medie latenti

$\Phi$ : matrice  $Q \times Q$  delle correlazioni tra le v. latenti

Nota:  $\mu_\eta = 0_{Q \times 1} \Rightarrow \eta \sim N_Q(i; 0, \Phi)$



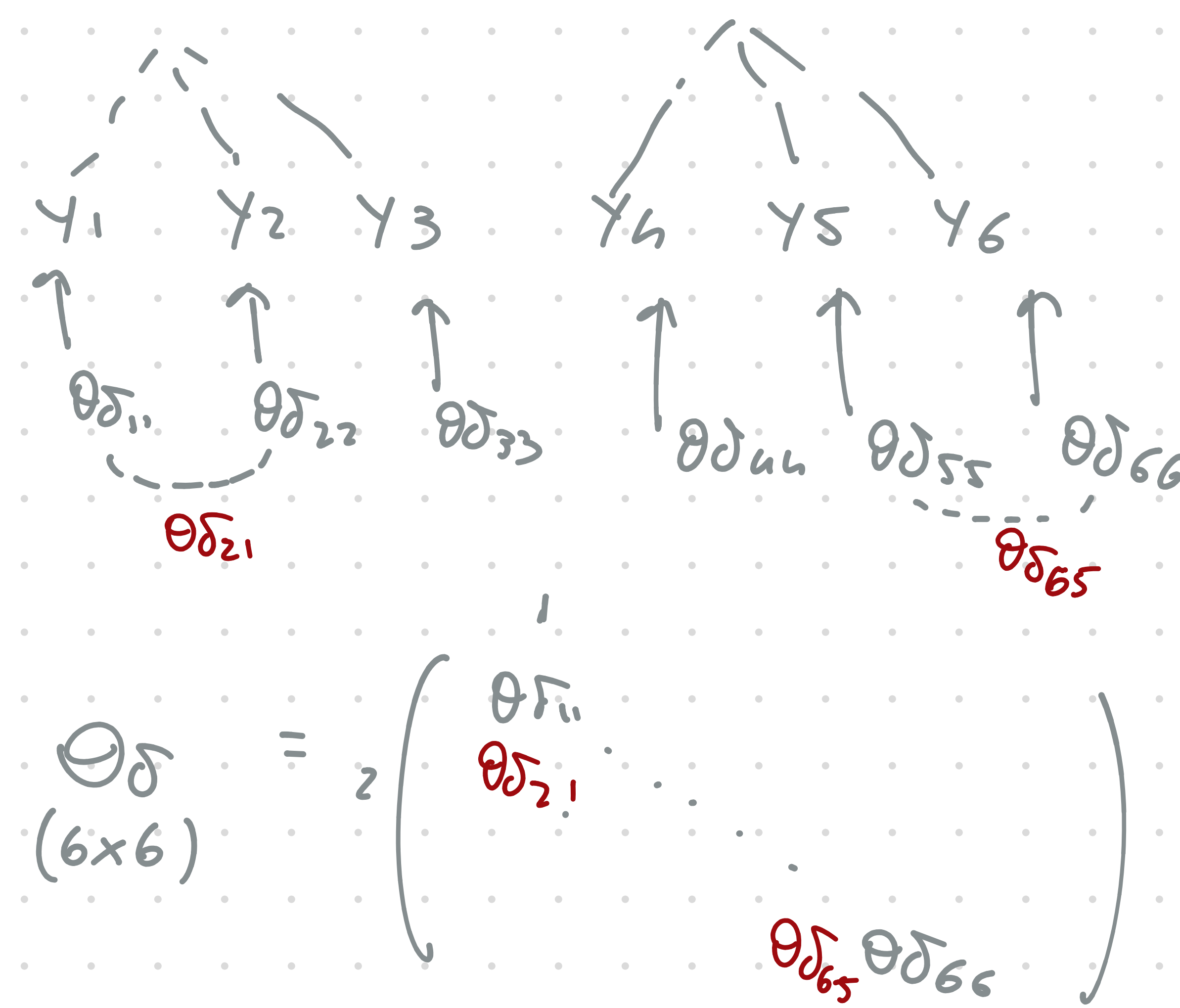
$$2) \delta_{m \times 1}^{(i)} = (\delta_1, \dots, \delta_m)^T \sim N_m(i; 0_m, \Theta_\delta)$$

$E[\delta] = 0_m$ : errore casuale e non sistematico

$\Theta_\delta$ : matrice  $m \times m$  delle correlazioni/covarianze tra gli errori delle misurazioni.

Solitamente  $\Theta_\delta$  è una matrice diagonale.

$Q=2, m=6$



$$3) y_{m \times 1}^{(i)} \sim N_m(i; \mu_y, \Sigma_y)$$

$\mu_y$ : vettore  $m \times 1$  delle medie 'riprodotte' dal modello

$\Sigma_y$ : matrice di cov/corr.  $m \times m$  'riprodotta' dal modello

Nota: (i)  $\mu_y$  e  $\Sigma_y$  sono quantità teoriche, indotte dal modello, non sono quantità osservate. (non sono q.tà osservate)

(ii)  $\gamma_y = 0_m$

$$\mu_y = E[y] = E[\gamma + \Lambda \eta + \delta] = \gamma + \Lambda E[\eta] + E[\delta] = 0$$

$$\Rightarrow y_{m \times 1}^{(i)} \sim N_m(i; 0_m, \Sigma_y)$$

"matrice attesa" (indotta dal modello)

Semplici fissature del modello che conduce ad avere  $\Sigma_y$  quale unico parametro da stimare



$$\begin{aligned}
\Sigma_Y &= \text{Cov}[Y] \\
&= E[Y Y^T] \\
&= E[(X + \Lambda \eta + \delta)(X + \Lambda \eta + \delta)^T] \\
&= E[(\Lambda \eta + \delta)(\eta^T \Lambda^T + \delta^T)] \\
&= E[\Lambda \eta \eta^T \Lambda^T + \Lambda \eta \delta^T + \delta \eta^T \Lambda^T + \delta \delta^T] \\
&= \Lambda E[\eta \eta^T] \Lambda^T + 2\Lambda E[\eta] E[\delta^T] + E[\delta \delta^T] \\
&= \Lambda \Phi \Lambda^T + \Theta \delta
\end{aligned}$$

Nota: In generale,  
 $\text{Cov}[X, Y] = E[XY] - E[X]E[Y]$

(4)

Notiamo che le relazioni tra le v. osservate  $y_1, \dots, y_m$  sono modellate in termini di  $\Sigma_Y$  che a sua volta è scritta in funzione dei seguenti parametri:

$\Lambda_{m \times Q}$ : matrice che contiene informazioni sulla struttura di misurazione

$\Phi_{Q \times Q}$ : matrice che contiene informazioni sulle relazioni tra le v. latenti

$\Theta_{m \times m}$ : matrice che contiene informazioni sulle relazioni tra gli errori di misura (solitamente diagonale)

Sulla base di  $\Phi_{Q \times Q}$  avremo quanto segue:

- modelli unidimensionali ( $Q=1$ )  $\{\phi_{11}\}$
- modelli bidimensionali ( $Q=2$ )  $\{\phi_{11}, \phi_{21}, \phi_{12}\}$
- modelli multidimensionali ( $Q>2$ )  $\{\phi_{11}, \dots, \phi_{Q2}, \frac{1}{2}Q(Q-1) \text{ cov/corr}\}$

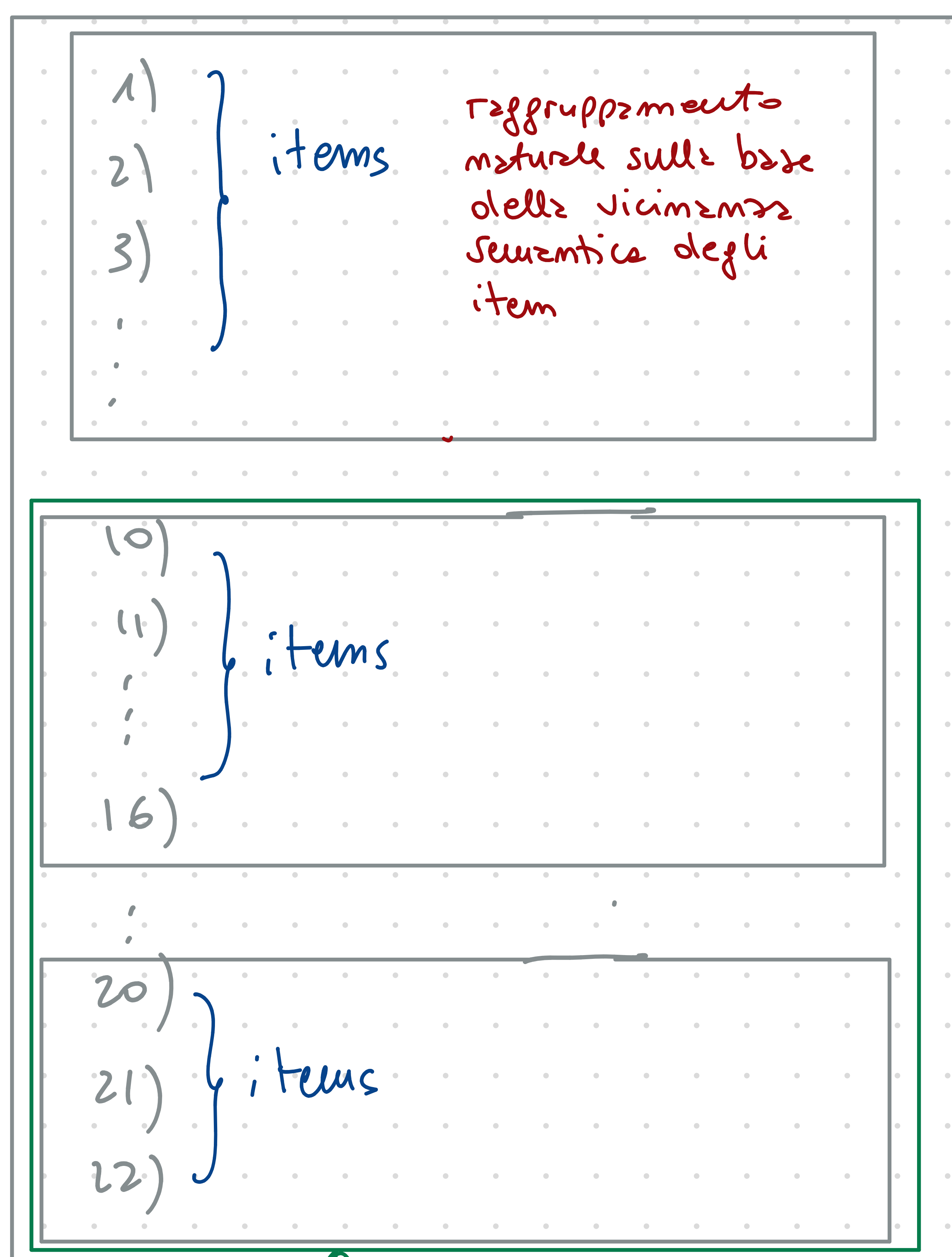


Note:

- La dimensionalità di un test/questionario è studiata sulle v. latenti  $\eta_1, \dots, \eta_Q$
- Se  $\Phi_{Q \times Q} = I_{Q \times Q}$  il modello CFA è detto ortogonale (v. latenti non correlate)
- Se  $\Phi = I$  e  $\Lambda$  non presenta cross-loadings, il modello CFA è detto modello di misura a struttura semplice

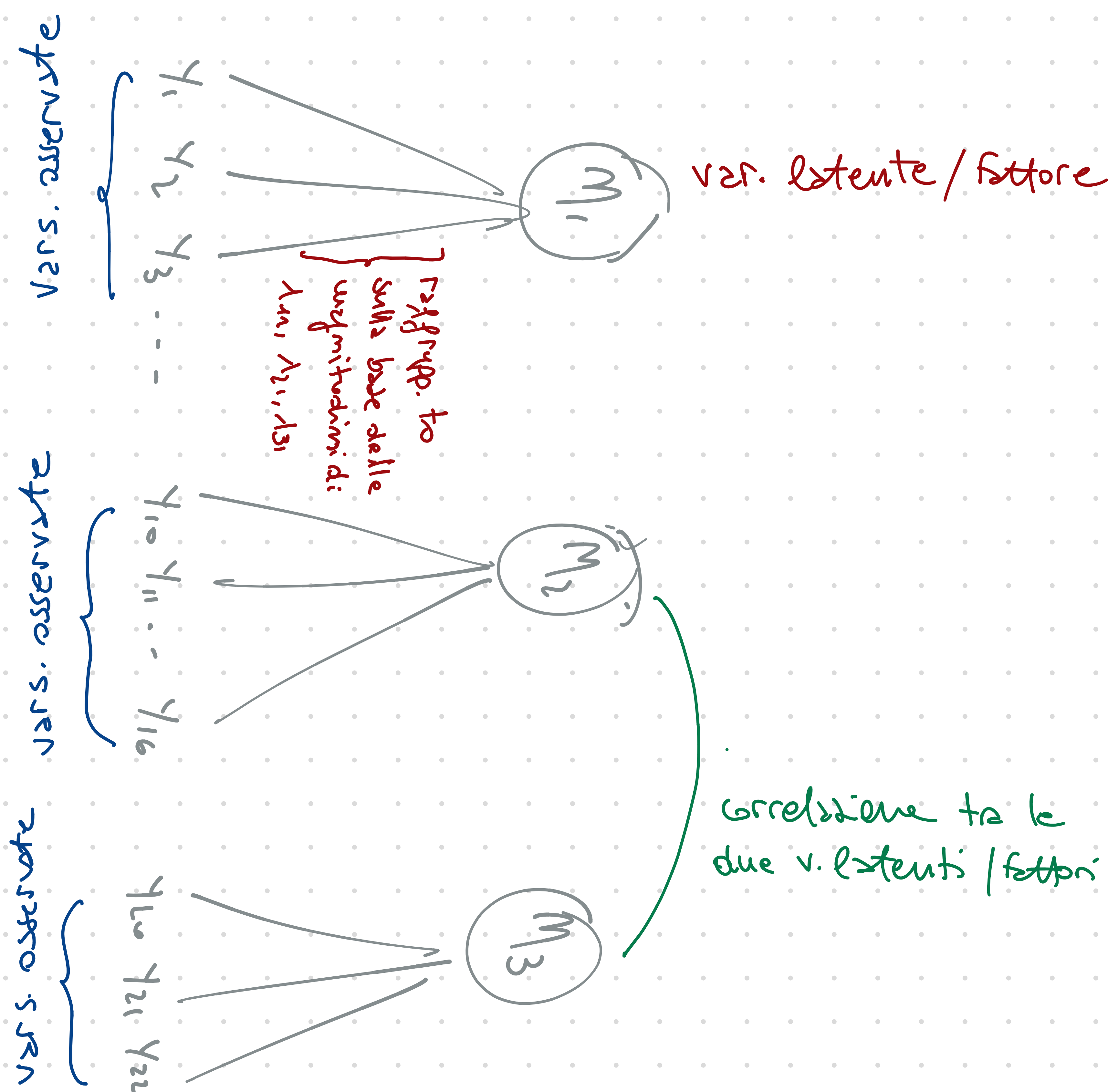
Di seguito si propone una schematizzazione semplificata delle relazioni (esprrese con diversi colori) tra un questionario e un modello CFA.

Questionario / test



raggruppamento di blocchi di items sulle basi delle vicinanza semantiche

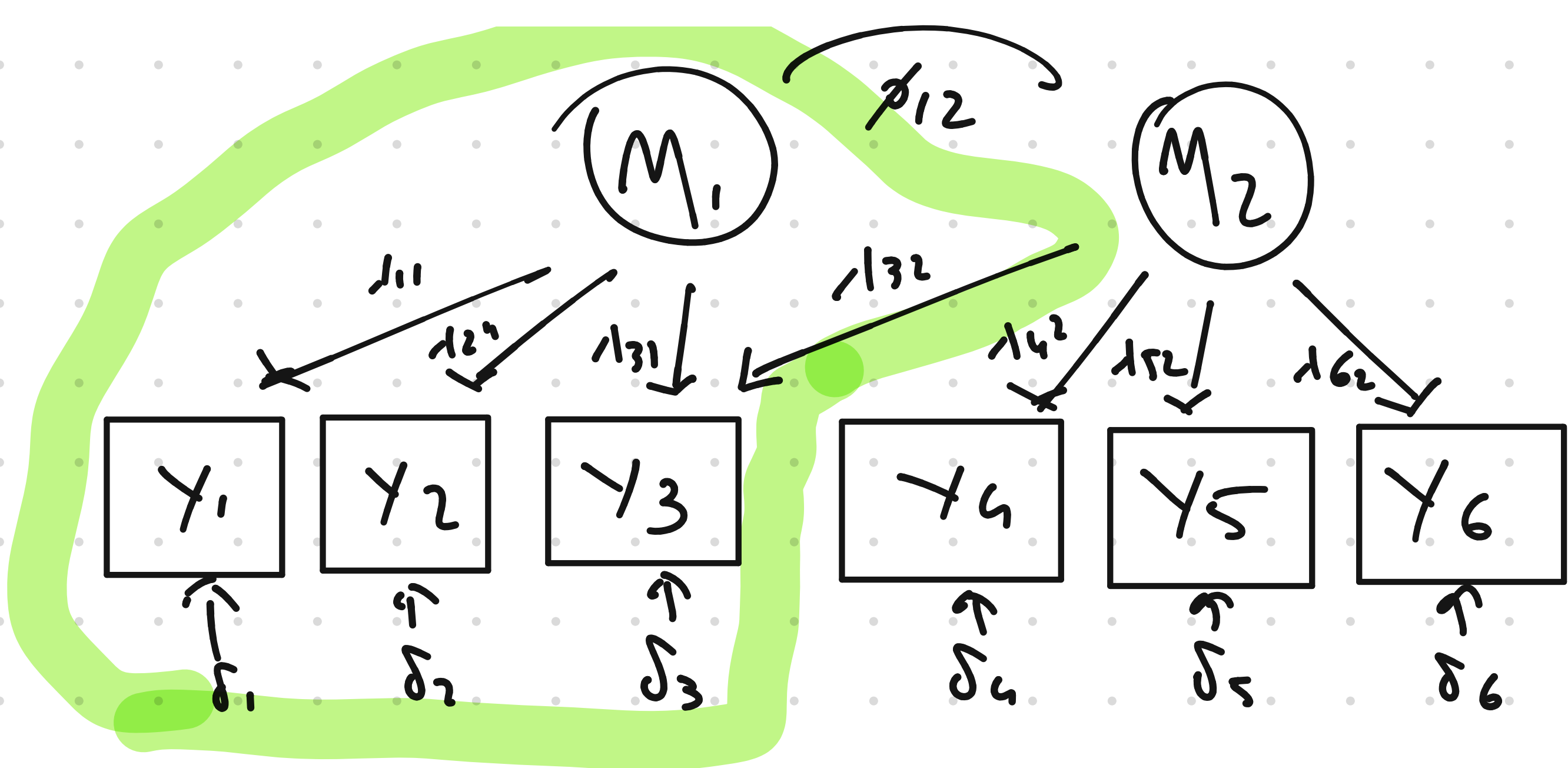
Modello fattoriale





# Decomposizione delle cov. osservate

Il modello CFA riproduce la c.d. matrice di cov. attesa  $\Sigma_Y$  (matrice dei legami teorici che il modello ipotizzato riproduce sempre senza considerare i dati) che successivamente viene studiata in relazione a  $S_Y^{oss}$  (matrice di cov. osservate) mediante opportuna metrica  $d(S_Y^{oss}, \Sigma_Y)$  minimizzando la quale si ottiene la matrice di cov. stimata  $\hat{\Sigma}_Y$ . Per comprendere al meglio la relazione tra  $S_Y^{oss}$  e  $\Sigma_Y$ , consideriamo l'esempio seguente:



$$\begin{aligned} Y_1 &= \eta_1 \lambda_{11} + \delta_1 \\ Y_2 &= \eta_1 \lambda_{12} + \delta_2 \\ Y_3 &= \eta_1 \lambda_{13} + \eta_2 \lambda_{23} + \delta_3 \\ Y_4 &= \eta_2 \lambda_{21} + \delta_4 \\ Y_5 &= \eta_2 \lambda_{22} + \delta_5 \\ Y_6 &= \eta_2 \lambda_{23} + \delta_6 \end{aligned}$$

$$\Sigma_{Y_{6 \times 6}} = \begin{pmatrix} \sigma_{11}^2 & & & & & \\ & \sigma_{22}^2 & & & & \\ & & \sigma_{33}^2 & & & \\ & & & \sigma_{44}^2 & & \\ & & & & \sigma_{55}^2 & \\ & & & & & \sigma_{66}^2 \end{pmatrix}$$

Ci concentriamo solo sulla scala  $M_1$  (v. latenti sono nulli di  $Y_1, Y_2, Y_3$ ).

$$\Sigma_{Y_{3 \times 3}} = \begin{pmatrix} \sigma_{11}^2 & - & - \\ \sigma_{21} & \sigma_{22}^2 & - \\ \sigma_{31} & \sigma_{32} & \sigma_{33}^2 \end{pmatrix}$$

$$\Sigma_{Y_{3 \times 3}} = \lambda_{3 \times 1} \phi \lambda_{3 \times 1}^T + \theta \delta_{3 \times 3}$$

$$\begin{aligned} \sigma_{11}^2 = \text{Var}[Y_1] &= E[Y_1 Y_1^T] \\ &= E[(\eta_1 \lambda_{11} + \delta_1)(\eta_1 \lambda_{11} + \delta_1)^T] \\ &= \dots \\ &= E[\eta_1 \eta_1^T] \lambda_{11}^2 + E[\delta_1 \delta_1^T] + 2 \lambda_{11} E[\eta_1 \delta_1] = 0 \\ &= \lambda_{11}^2 \phi_{11}^2 + \theta \delta_{11} \end{aligned}$$

varianza spiegata di  $\eta_1$  (var. comune) C
varianza non spiegata di  $\eta_1$  (var. unica) u

$$\begin{aligned} \sigma_{33}^2 = \text{Var}[Y_3] &= E[Y_3 Y_3^T] \\ &= E[(\eta_1 \lambda_{31} + \eta_2 \lambda_{32} + \delta_3)^2] \\ &= \dots \\ &= E[\eta_1^2] \lambda_{31}^2 + E[\eta_2^2] \lambda_{32}^2 + E[\delta_3^2] + 2 E[\eta_1 \eta_2] \lambda_{31} \lambda_{32} \\ &= \lambda_{31}^2 \phi_{11}^2 + \lambda_{32}^2 \phi_{22}^2 + 2 \lambda_{31} \lambda_{32} \phi_{21} + \theta \delta_{33} \end{aligned}$$

var. comune C
var. unica u

Nota: l'introduzione del legame diretto  $\eta_2 \rightarrow Y_3$  e del legame indiretto  $Y_3 \leftarrow \eta_1 \leftrightarrow \eta_2$  consente di assorbire maggiore varianza di  $Y_3$  qualora il legame  $\eta_1 \rightarrow Y_3$  non fosse in grado di farlo. Cio' sottrae varianza alla componente unica (di errore).

$$\begin{aligned} \sigma_{21} = \text{Cov}[Y_2, Y_1] &= E[Y_2 Y_1^T] \\ &= E[(\eta_1 \lambda_{12} + \delta_2)(\eta_1 \lambda_{11} + \delta_1)^T] \\ &= \dots \\ &= \lambda_{11} \lambda_{21} E[\eta_1 \eta_1^T] + E[\delta_1 \delta_2^T] \\ &= \phi_{11}^2 \lambda_{11} \lambda_{21} + \theta \delta_{21} = 0 \text{ (poiche' non ipotizzato)} \\ &= \phi_{11}^2 \lambda_{11} \lambda_{21} \end{aligned}$$

Se  $\phi_{11}^2 = 1$  allora  $\sigma_{21} = \lambda_{11} \lambda_{21}$  (la cov. tra  $Y_1$  e  $Y_2$  e' spiegata/dovuta alla v. latente  $\eta_1$ )

$$\Sigma_Y = \begin{pmatrix} Y_1 & Y_2 & Y_3 \\ Y_1 & \lambda_{11}^2 \phi_{11}^2 + \theta \delta_{11} & ? \\ Y_2 & \phi_{11}^2 \lambda_{11} \lambda_{21} & ? \\ Y_3 & ? & ? \end{pmatrix}$$

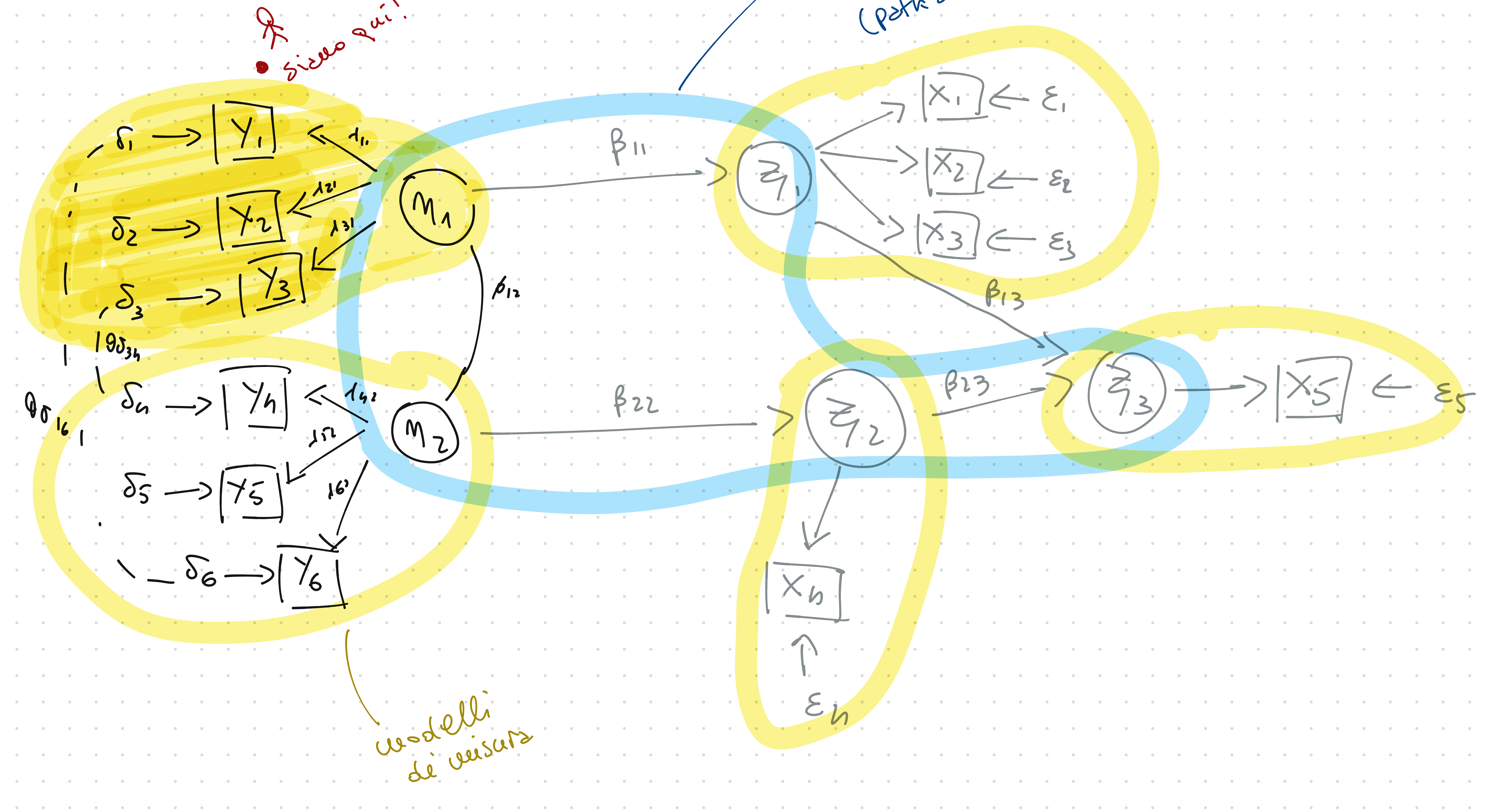
$\lambda_{31}^2 \phi_{11}^2 + \lambda_{32}^2 \phi_{22}^2 + 2 \lambda_{31} \lambda_{32} \phi_{21} + \theta \delta_{33}$



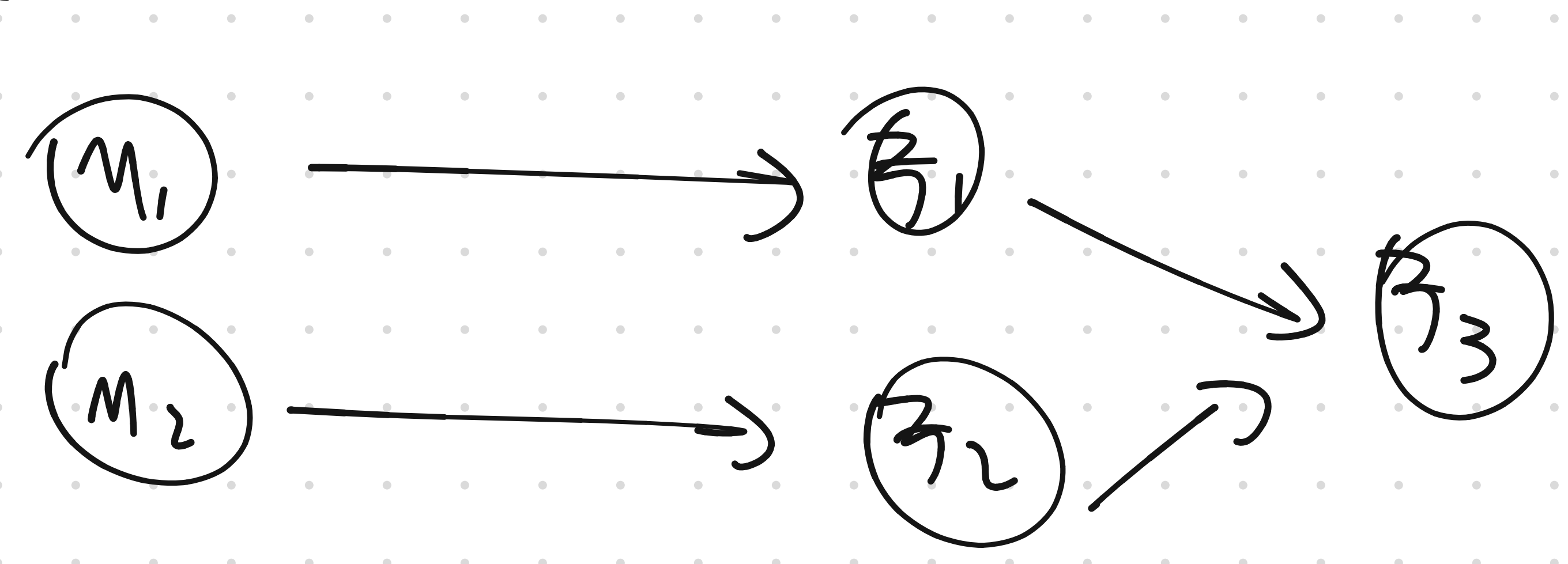
# SEM

modello strutturale (path analysis)

• siamo qui!



modelli strutturali:





$$\Sigma_Y = \Lambda \Phi \Lambda^T + \Theta\sigma$$

Il modello CFA per ricostruire i legami tra le v. osservate utilizza l'array di parametri  $\{\Lambda, \Phi, \Theta\sigma\}$ .  
Per stimare  $\hat{\Sigma}_Y$  si lavora sui parametri liberi contenuti in  $\{\Lambda, \Phi, \Theta\sigma\}$ .

Di seguito si descrivono alcuni elementi essenziali allo studio della parametrizzazione del modello CFA:

• Gradi di libertà: misura che quantifica quanto un modello è parsimonioso secondo la definizione seguente:

$$df = \left( \underbrace{\frac{1}{2}(m^2 + m)}_M - p \right) = M - p$$

M: numero di elementi non ridondanti di  $\Sigma_Y$

dove  $M$  è il numero massimo di termini che la matrice  $\Sigma_Y^{oss}$  ammette mentre  $p$  è il numero di parametri liberi del modello.

1) Se  $df = 0$  oppure  $\frac{p}{M} = 1$  allora il modello CFA è detto saturo, con l'implicazione che  $\Sigma_Y^{oss}$  è esattamente riprodotta da  $\Sigma_Y$  (il modello  $\Sigma_Y$  "adatta" perfettamente a  $\Sigma_Y$ ).

2) Se  $df > 0$  oppure  $\frac{p}{M} < 1$  allora il modello CFA è sottoidentificato (ie: più eqs che incognite / parametri), con l'implicazione che si andrà alla ricerca della soluzione ottimale tra le diverse soluzioni ammissibili.

3) Se  $df < 0$  allora il modello CFA non è identificabile\*, con l'implicazione che il modello ipotizzato necessita di essere ulteriormente revisionato (es: con l'aggiunta di vincoli).

Condizione per l'identificabilità:

$$df \geq 0$$

Nota: nella pratica,  $df > 0$

\* Un modello non identificabile è un modello non utilizzabile nella pratica di analisi dei dati.

Tecnicamente un modello CFA è identificato se:

①  $df > 0$

② se è possibile esprimere le varianze-covarianze osservate ( $\Sigma_Y$ ) in termini del modello ( $\Sigma_Y$ )

③ varianze stimate dal modello - ie:  $\text{diag}(\Theta\sigma)$ ,  $\text{diag}(\Phi)$  - sono positive

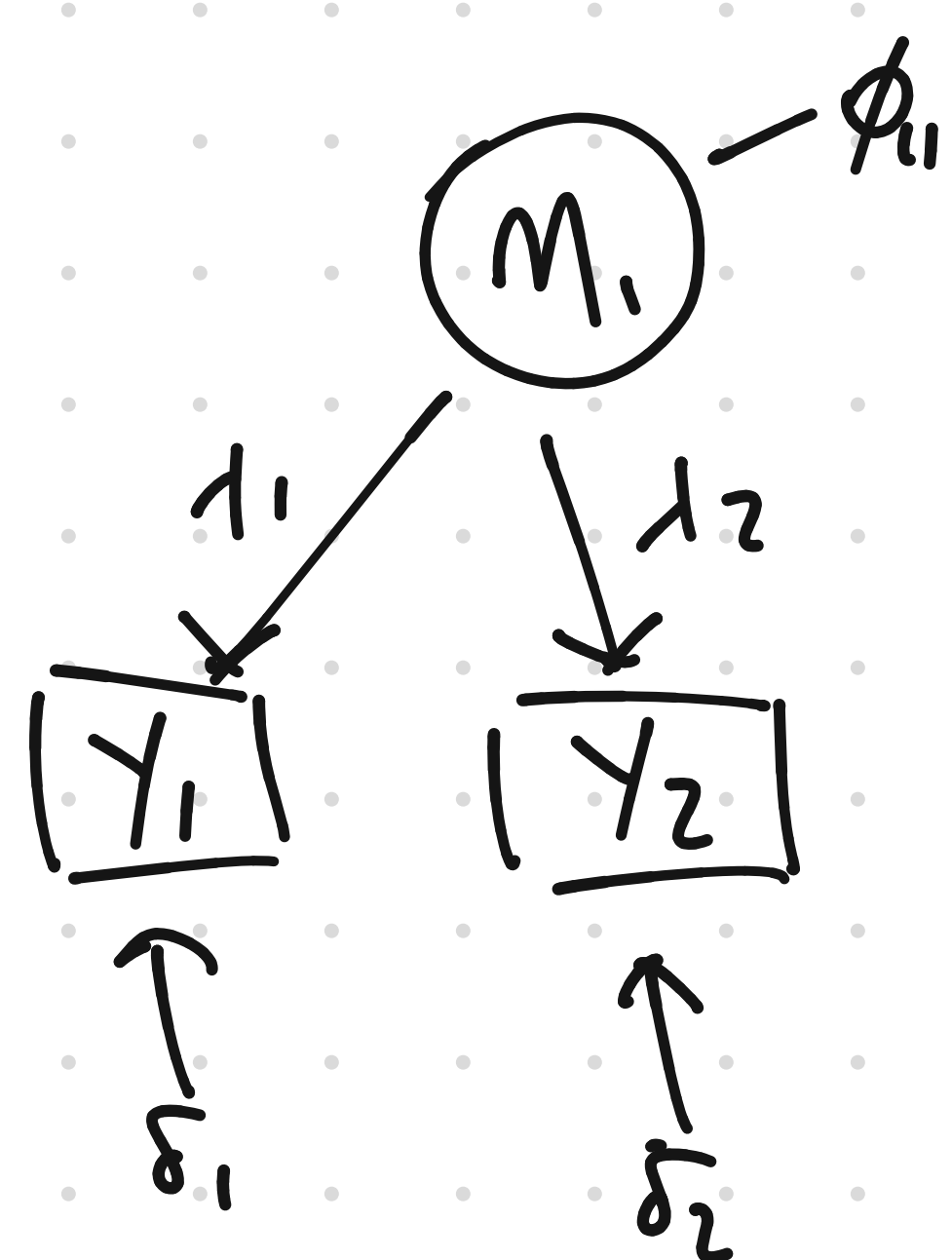
④ standard errors - ie: varianze delle stime  $\hat{\Lambda}$ ,  $\hat{\Phi}$ ,  $\hat{\Theta}\sigma$  - devono essere "contenuti" (magnitudine non eccessivamente elevata)

⑤ matrice dell'informazione attesa  $I(\theta)$  (information matrix) è semi-positiva definita, ie:  $\det(I(\theta)) \neq 0$

Se un modello CFA non è identificabile  $\Rightarrow$  aggiungere vincoli al modello (aumentare i parametri fissi / ridurre i parametri liberi)

\* Esempio completo su identificabilità: tab 6.2, pp. 138-141 (Gibetta)  $\leftarrow$  punto ②

Esempio (sull'espressione di  $\Sigma_Y$  in termini di  $\Sigma_Y$ )



$$\Sigma_Y = \begin{pmatrix} \sigma_{11}^2 & - \\ \sigma_{21} & \sigma_{22}^2 \end{pmatrix}$$

(2x2)

$$\Sigma_Y = \begin{pmatrix} s_{11}^2 & - \\ s_{21} & s_{22}^2 \end{pmatrix}$$

(2x2)

$$\begin{aligned} \sigma_{11}^2 &= \lambda_1^2 \phi_1^2 + \theta\sigma_1 & (1) \\ \sigma_{22}^2 &= \lambda_2^2 \phi_1^2 + \theta\sigma_2 & (2) \\ \sigma_{21} &= \lambda_1 \lambda_2 \phi_1^2 & (3) \end{aligned}$$

$$p = 5$$

$$M = 3$$

$$df = (3 - 5) < 0$$

Considerato da  $df < 0$  il modello necessita di ulteriori revisioni. Si vedano le soluzioni proposte di seguito.



Prima soluzione (21 problemi df < 0)

7

$$\begin{aligned}\sigma_{11}^2 &= \lambda^2 + \theta\delta_1 \\ \sigma_{22}^2 &= \lambda^2 + \theta\delta_2 \\ \sigma_{21} &= \lambda\lambda\end{aligned}$$

(vincolo di ug)

$$\lambda = \lambda_1 = \lambda_2$$

$$\phi_1^2 = 1 \quad (\text{param. fisso})$$

$$p = 3$$

$$df = 0$$

$$\{\lambda, \theta\delta_1, \theta\delta_2\}$$

$$\hat{\lambda} = \sqrt{\sigma_{21}}$$

$$\hat{\theta\delta_2} = \sigma_{22}^2 - \sqrt{\sigma_{21}}$$

$$\hat{\theta\delta_1} = \sigma_{11}^2 - \sqrt{\sigma_{21}}$$

Le soluzioni trovate sono tutte espresse in termini di variante-covarianza  $\Sigma_Y$

Seconda soluzione (21 problemi df < 0)

$$\begin{aligned}\sigma_{11}^2 &= \lambda_1^2 + 0.2 \\ \sigma_{22}^2 &= \lambda_2^2 + \theta\delta_{22} \\ \sigma_{21} &= \lambda_1\lambda_2\end{aligned}$$

$$\theta\delta_1 = 0.2 \quad (\text{param. fisso})$$

$$\phi_1 = 1 \quad (\text{param. fisso})$$

$$p = 3$$

$$df = 0$$

$$\hat{\lambda}_1 = \sqrt{\sigma_{11}^2 - 0.2}$$

$$\hat{\lambda}_2 = \frac{\sigma_{21}}{\hat{\lambda}_1}$$

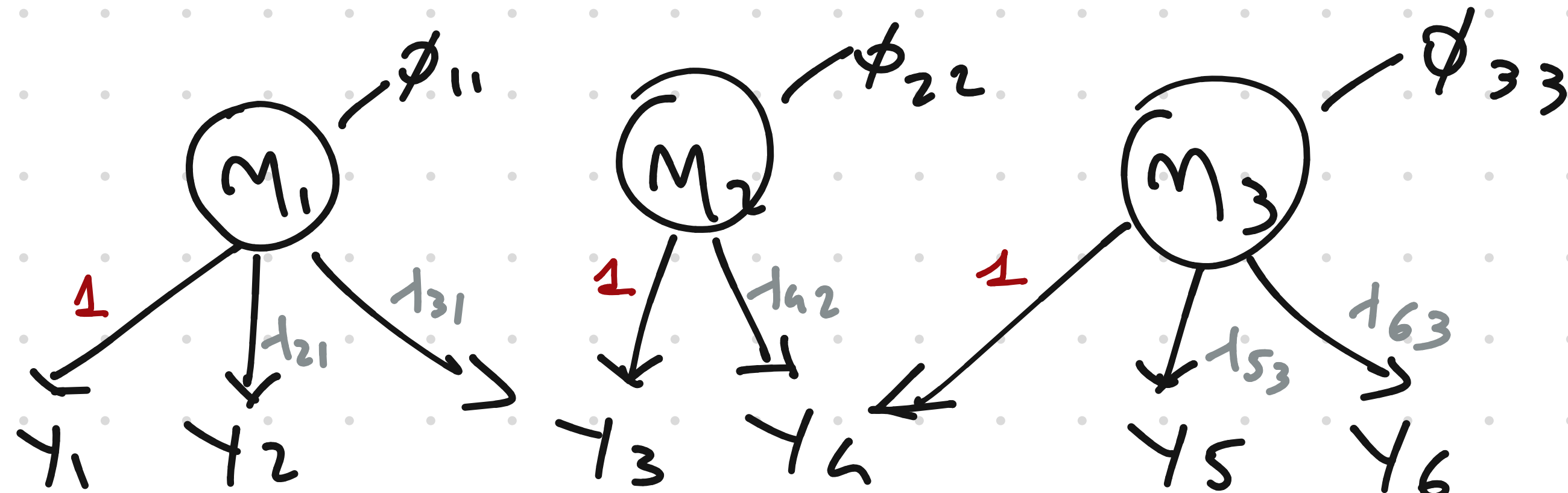
$$\hat{\theta\delta_2} = \sigma_{21}^2 - \hat{\lambda}_2$$

Le soluzioni trovate sono tutte espresse in termini di variante-covarianza  $\Sigma_Y$

• Metrics delle v. latenti: un modo 2 disposizione per ridurre il numero dei parametri liberi sfruttando il fatto che le v. latenti  $\eta_1 - \eta_k$  non hanno unità di misura.

1) ULI (unit loading identification): per ciascuna scala del modello CFA, si fissa a 1 il primo factor loading.

Es:

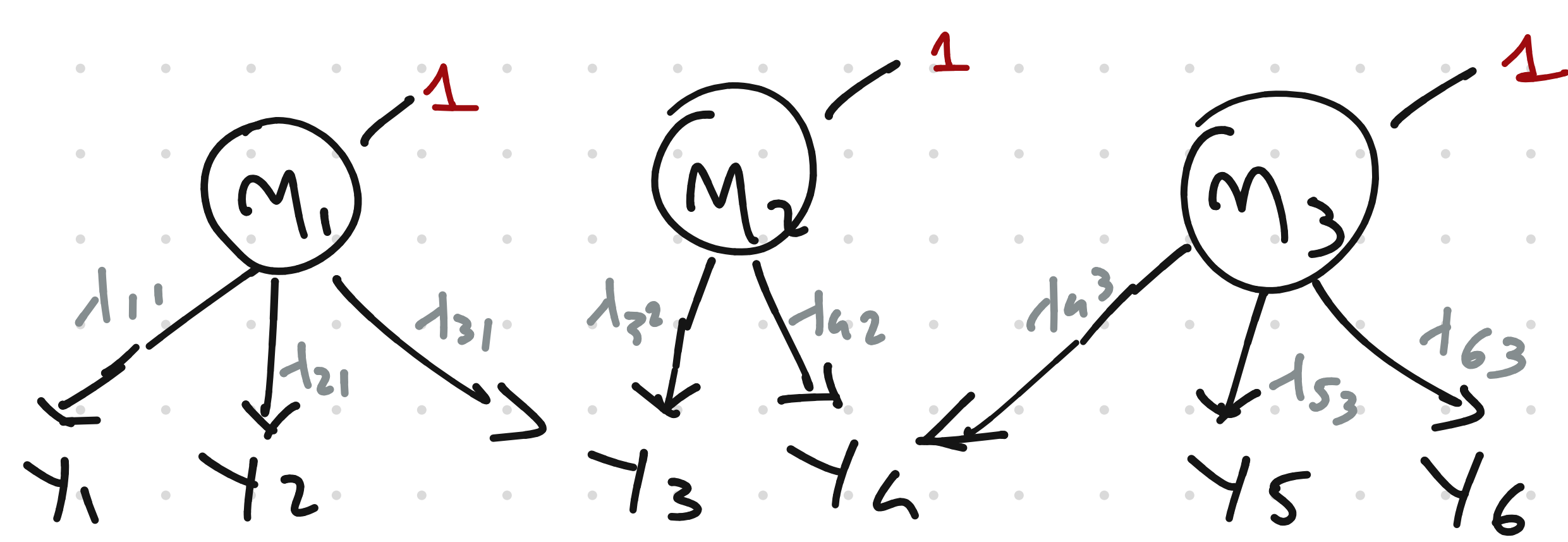


$$\Lambda_{(6 \times 3)} = \begin{pmatrix} 1 & 0 & 0 \\ \lambda_{21} & 0 & 0 \\ \lambda_{31} & 1 & 0 \\ 0 & \lambda_{42} & 1 \\ 0 & 0 & \lambda_{53} \\ 0 & 0 & \lambda_{63} \end{pmatrix}$$

$$\Phi_{(3 \times 3)} = \begin{pmatrix} \phi_{11} & & \\ & \phi_{22} & \\ & & \phi_{33} \end{pmatrix}$$

2) UVI (unit variance identification): la diagonale della matrice  $\Phi$  è uguale a 1, ossia  $\text{diag}(\Phi) = \mathbf{1}_{Q \times 1}$

Es:



$$\Lambda_{(6 \times 3)} = \begin{pmatrix} \lambda_{31} & 0 & 0 \\ \lambda_{21} & 0 & 0 \\ \lambda_{31} & \lambda_{42} & 0 \\ 0 & \lambda_{42} & \lambda_{53} \\ 0 & 0 & \lambda_{53} \\ 0 & 0 & \lambda_{63} \end{pmatrix}$$

$$\Phi_{(3 \times 3)} = \begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix}$$

Stima dei parametri

$$\Sigma_Y = \Lambda \Phi \Lambda^T + \theta\delta$$

$$(n \times n) = (n \times q)(q \times q)(n \times q) + (n \times n)$$

$$\Omega = \{\Lambda, \Phi, \theta\delta\}$$

Sotto la condizione  $df > 0$ :

• massima verosimiglianza (maximum likelihood)

$$\arg \max_{\Omega \in \Omega} \ell(\Sigma_Y) \propto \frac{n}{2} \left( \ln |\Sigma_Y^{-1}| - \text{tr} \{ \Sigma_Y^{-1} S_Y \} \right)$$

$$\propto \frac{n}{2} \left( \ln |(\Lambda \Phi \Lambda^T + \theta\delta)^{-1}| - \text{tr} \{ (\Lambda \Phi \Lambda^T + \theta\delta)^{-1} S_Y \} \right)$$



- minimi quadrati non pesati (OLS)

$$\begin{aligned} \arg \min_{\Omega \in \mathbb{R}^n} d_{OLS}(\Sigma_Y) &= \frac{1}{2} \text{tr} \{ (\Sigma_Y - \Sigma_Y)^2 \} \\ &= \frac{1}{2} \text{tr} \{ (\Sigma_Y - (\Lambda \Phi \Lambda^T + \Theta \Sigma))^2 \} \end{aligned}$$

- minimi quadrati generalizzati (GLS)

$$\arg \min_{\Omega \in \mathbb{R}^n} d_{GLS}(\Sigma_Y) = \frac{1}{2} \text{tr} \{ (\Sigma_Y^{-1/2} (\Sigma_Y - \Sigma_Y) \Sigma_Y^{-1/2})^2 \}$$

- DWLS: metodo di stima per  $(Y_1 - Y_m)$  di tipo categoriale. In breve il metodo procede in due step, dapprima stimando  $\hat{\Sigma}_Y$  (matrice di cov./corr. polinomiche) per poi stimare  $\hat{\Omega}$  condizionatamente a  $\hat{\Sigma}_Y$ .

## Valutazione del modello (adattamento ai dati)

Una volta ottenuta  $\hat{\Sigma}_Y$  e' doveroso chiedersi quanto il modello stimato si adatti ai dati rispetto a criteri globali.

- test del  $\chi^2$ : dato  $\Sigma_Y$ , si definiscono e si adattano ai dati due modelli CFA,

1)  $\mathcal{M}_1$  con  $df_1$  (modello corrente)

2)  $\mathcal{M}_2$  con  $df_2 = 0$  (modello saturo)

$$3) \chi^2_1 = -2 \ell(\hat{\Sigma}_Y) / \nu_{\mathcal{M}_1} - \ell(\hat{\Sigma}_Y) / \nu_{\mathcal{M}_2}$$

$$\chi^2_1 \stackrel{H_0}{\sim} \chi^2(df_1) \quad H_0: \mathcal{M}_1 = \mathcal{M}_2$$

Obiettivo: rigettare  $H_0 \Rightarrow$  il modello  $\mathcal{M}_1$  e' distinto da  $\mathcal{M}_2$

Nota: per  $n$  elevato il test rigetta sempre  $H_0$

- CFI: l'indice varia tra 0 e 1 e valori prossimi a 1 sono da considerarsi ottimali

$$CFI = \frac{(\chi^2_1 - df_1)}{\max \{ \chi^2_1 - df_1, \underbrace{\chi^2_0 - df_0}_{\text{stat. } \chi^2_0 \text{ che si ottiene sotto l'ipotesi di modello nullo:}} \}}$$

stat.  $\chi^2_0$  che si ottiene sotto l'ipotesi di modello nullo:

$$\Sigma_Y = \underbrace{I}_{(n \times n)} + \Theta \Sigma$$

- Altri indici: RMSEA (piu' vicino a zero e' ottimale)  
TLI/NNFI  
...

- AIC: indice che si usa nel confronto tra piu' modelli ("the lower, the better")

$$AIC = -2 \ell(\Sigma_Y) + 2m$$

### Indici assoluti

CFI  
 $\chi^2$   
RMSEA  
SRMR  
Altri (TLI, NNFI)

Ci sono dei valori soglia da rispettare

### Valori soglia:

CFI  $\geq 0.95$   
RMSEA  $\leq 0.05$  (accettabile  $\leq 0.10$ )  
SRMR  $\leq 0.08$  (accettabile  $\leq 0.10$ )

CFI  $> 0.95$   $\wedge$  SRMR  $\leq 0.08$

### Indici relativi

AIC  
parsimonieta' ( $\frac{p}{n}$ )  
df

Nessun valore soglia: dati una serie di AICs, si sceglie il minore



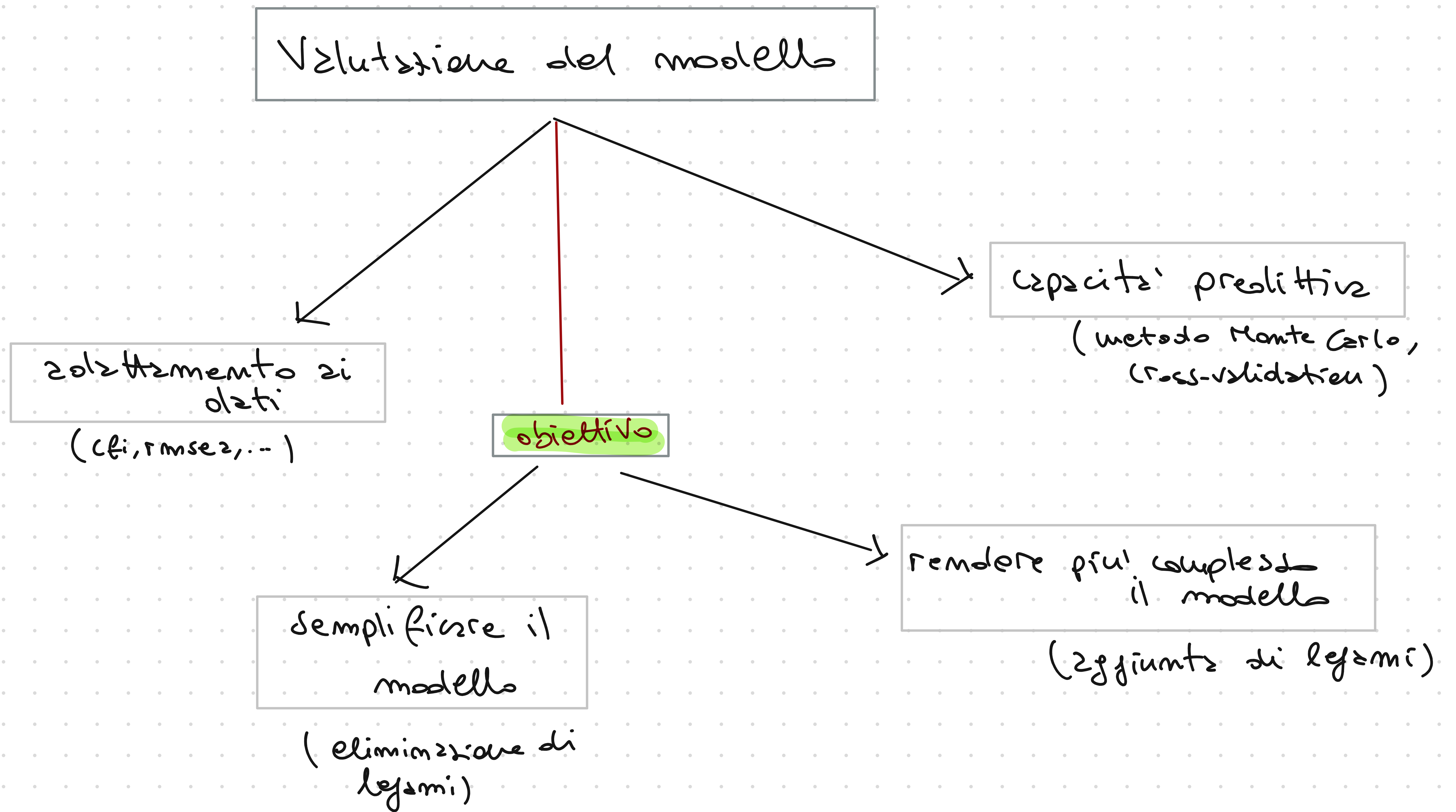
In generale la valutazione del modello CFA può essere fatta:

(9)

- in termini di adattamento ai dati  $S_y^{obs}$  (usando gli indici visti in precedenza)
- in termini di capacità predittiva (usando metodi Monte Carlo) \* Vedremo ciò nei laboratori.
- con l'obiettivo di:

⊕ Semplificare il modello, eliminando quei legami tra v. osservate e v. latenti che risultano al di sotto di una certa soglia (es:  $\lambda_{19} \leq 0.35$ )

⊕ complessificare il modello, aggiungendo parametri (ie, legami) non ipotizzati nel modello originario. La procedura utilizza i c.d. indici di modifica e sono basati su un approccio/metodo statistico (non teorico). \* Vedremo ciò nei laboratori





# Calcolo di $\hat{\eta}$

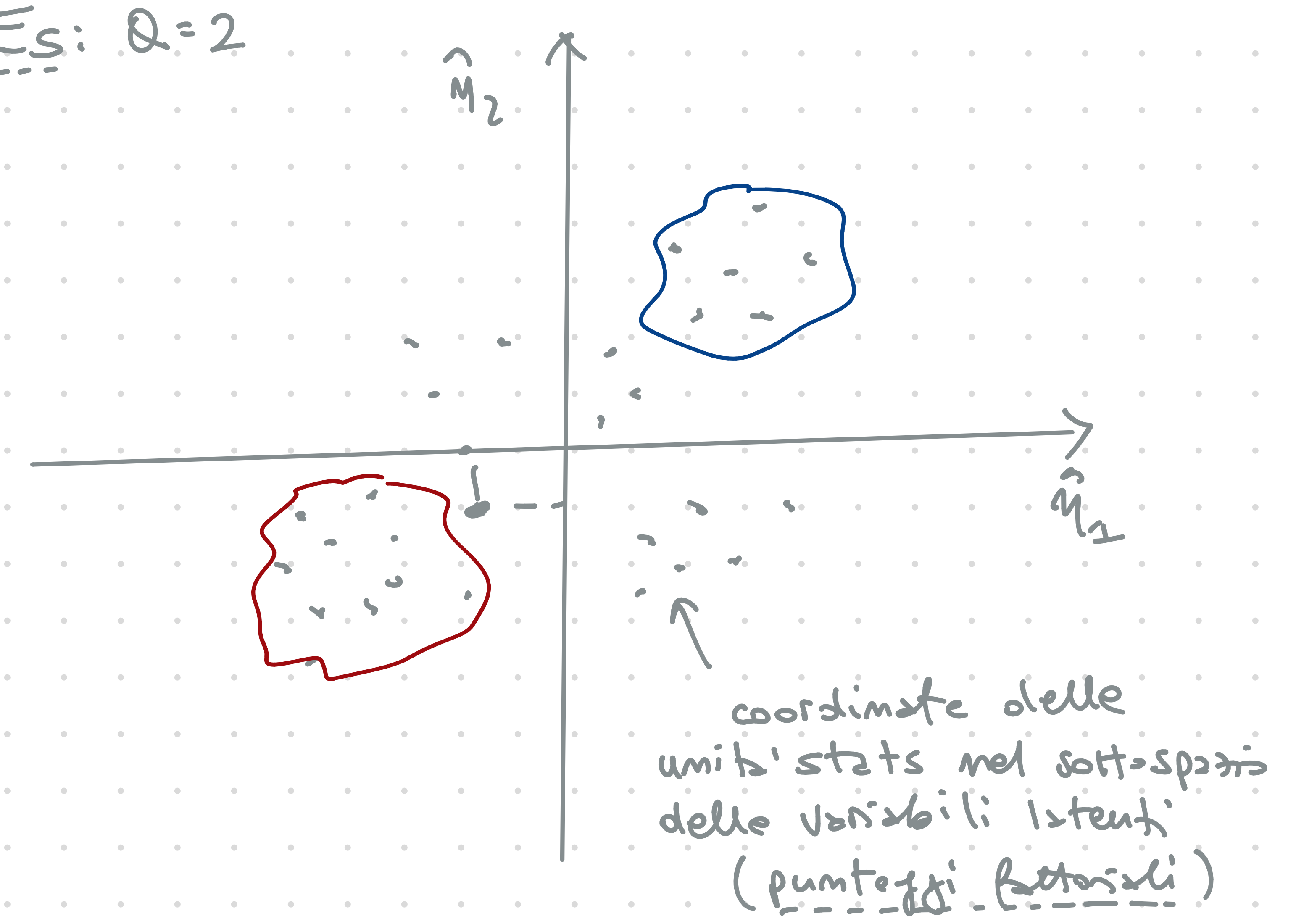
$$\eta_i \sim N_Q ( ; \mu_\eta, \Phi ) , \quad \mu_\eta = 0_{Q \times 1}$$

$$\delta_i \sim N_m ( ; 0_{m \times 1}, \Theta \delta ) , \quad \eta_i \perp \delta_i$$

$$y_i | \eta_i \sim N_m ( ; 0_{m \times 1}, \underbrace{\Sigma_y}_{= \Lambda \Phi \Lambda^T + \Theta \delta} )$$

$$\hat{\eta}_i = E[\eta_i | y_i] = \hat{\Phi} \hat{\Lambda}^T \hat{\Sigma}_y^{-1} (y_i - \hat{\mu}_y)$$

Es: Q=2



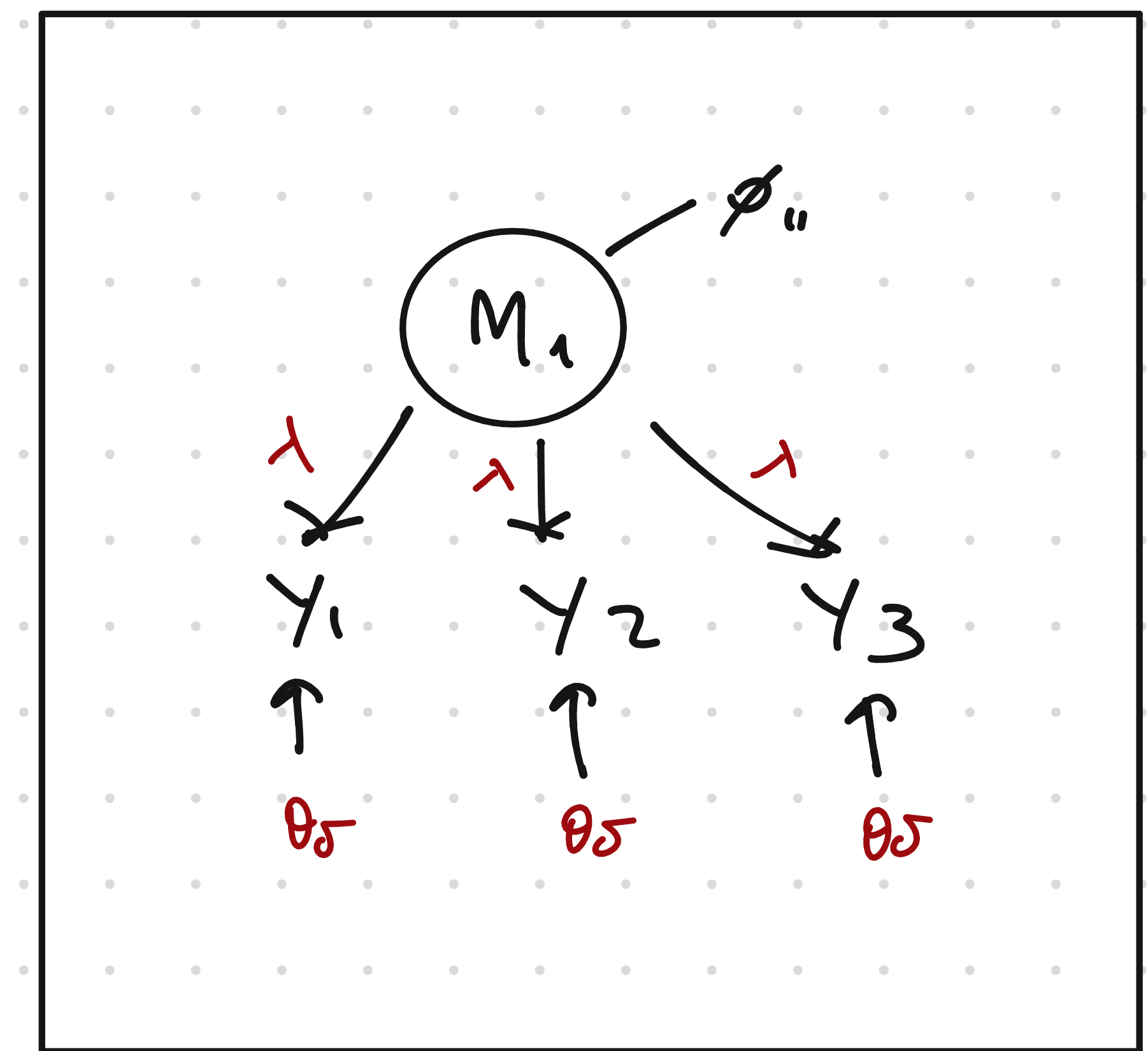
Esempio di rappresentazione grafica dei punteggi fattoriali  $\hat{\eta}_i$ . Una volta calcolati, questi possono essere utilizzati per ulteriori analisi statistiche, ad esempio analisi di raggruppamento (clustering) alla ricerca di gruppi (nel grafico sono rappresentati due ipotetici raggruppamenti).

## Modelli CFA notevoli

• Modello per test paralleli  
(caso unidimensionale)

$$\Sigma_y = \lambda \lambda^T \phi + \mathbb{I} \Theta \delta$$

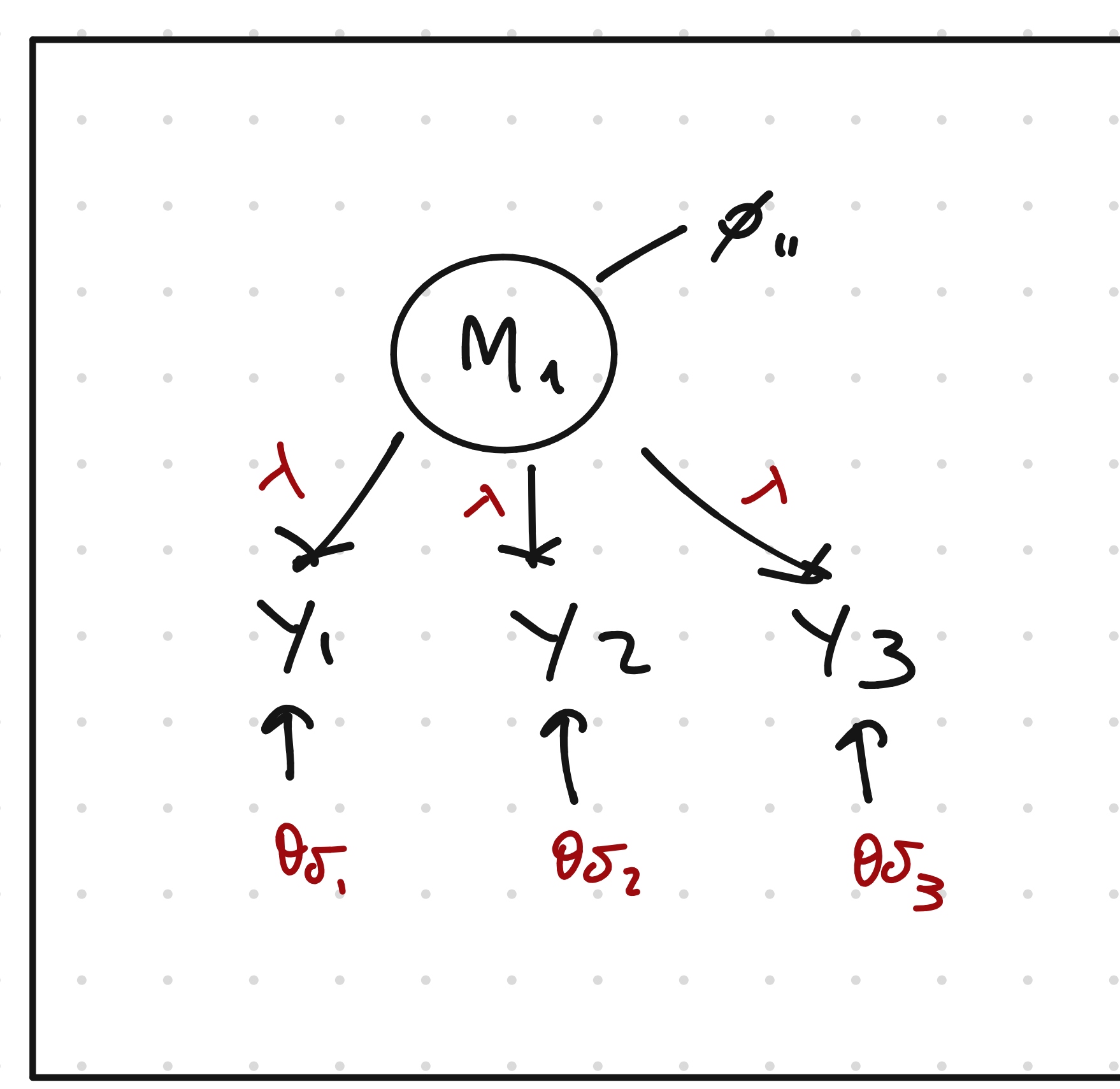
- 1)  $\lambda, \phi, \Theta \delta$  scalari
- 2)  $\lambda$  vettore  $m \times 1$  di 1
- 3)  $\mathbb{I}$  matrice  $m \times m$  (identità)
- 4)  $\lambda$  può essere stimato ( $\phi=1$ ) oppure fissato a priori ( $\phi$  viene stimato)



• Modello per test r-equivalenti  
(caso unidimensionale)

$$\Sigma_y = \lambda \lambda^T \phi + \mathbb{I} \Theta \delta$$

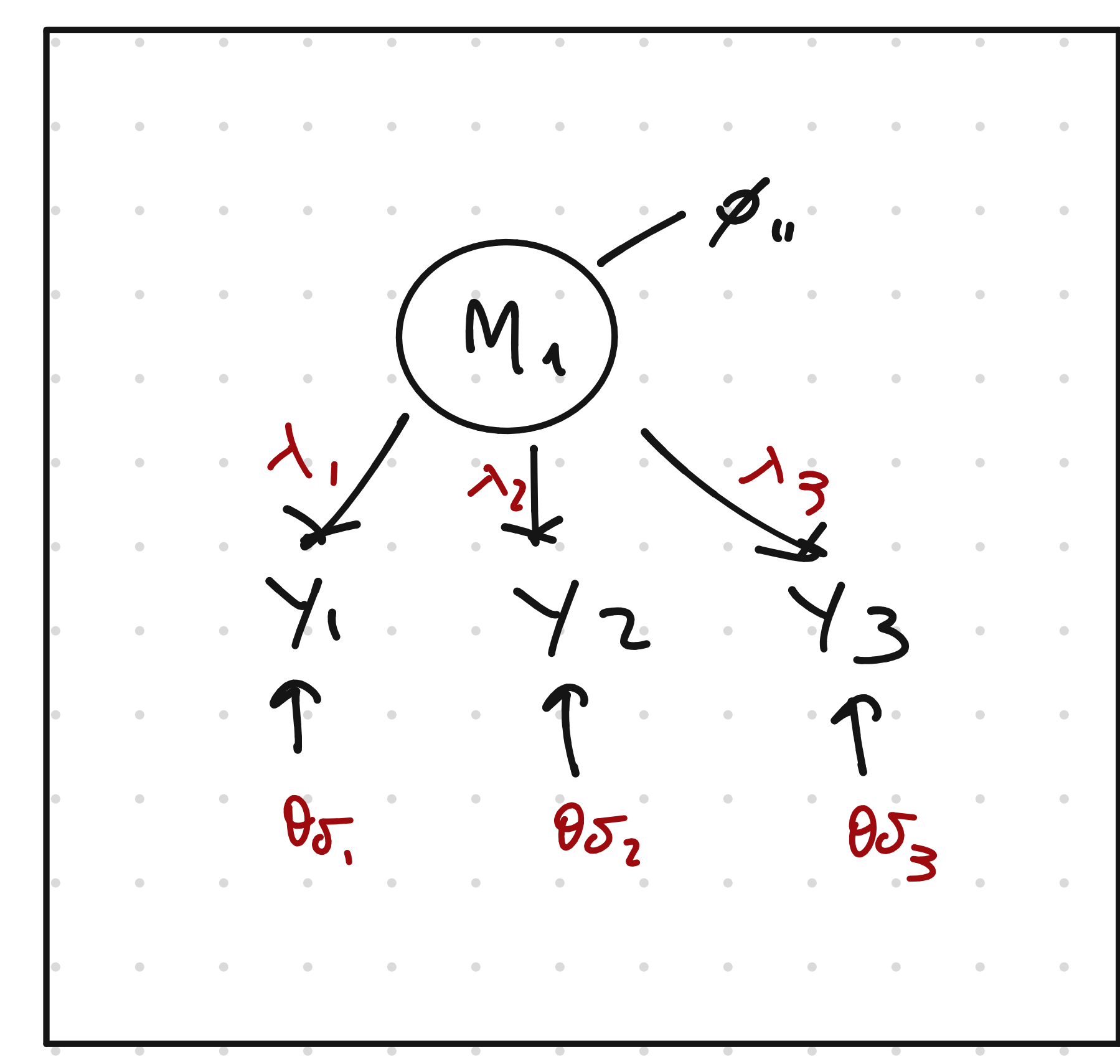
- 1)  $\lambda, \phi$  scalari
- 2)  $\Theta \delta$  è diagonale (oppure essere piena)



• Modello per test congruenti  
(caso unidimensionale)

$$\Sigma_y = \lambda \phi \lambda^T + \Theta \delta$$

Nota: Questo modello comincia con un modello CFA unidimensionale

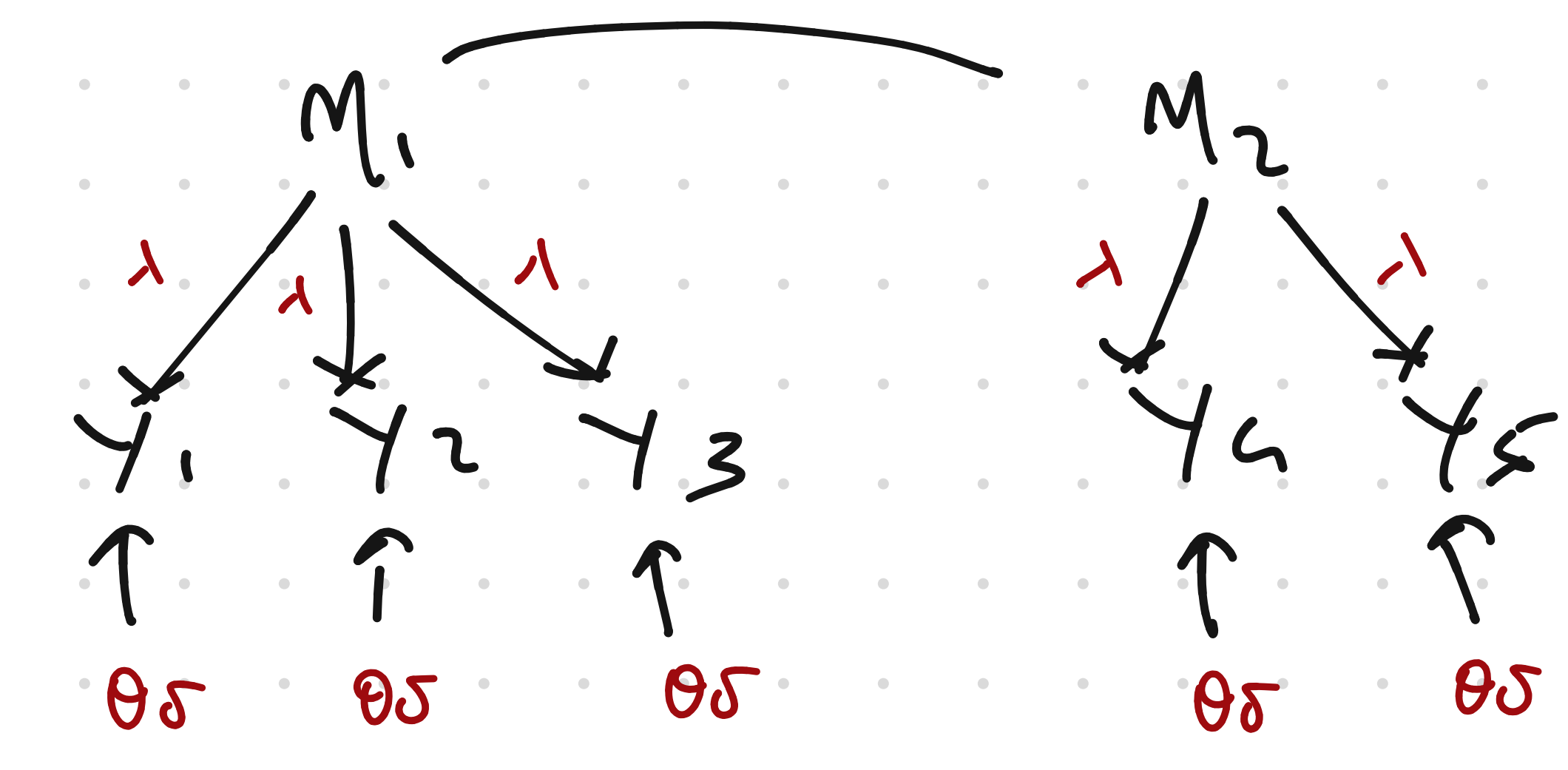


Note:

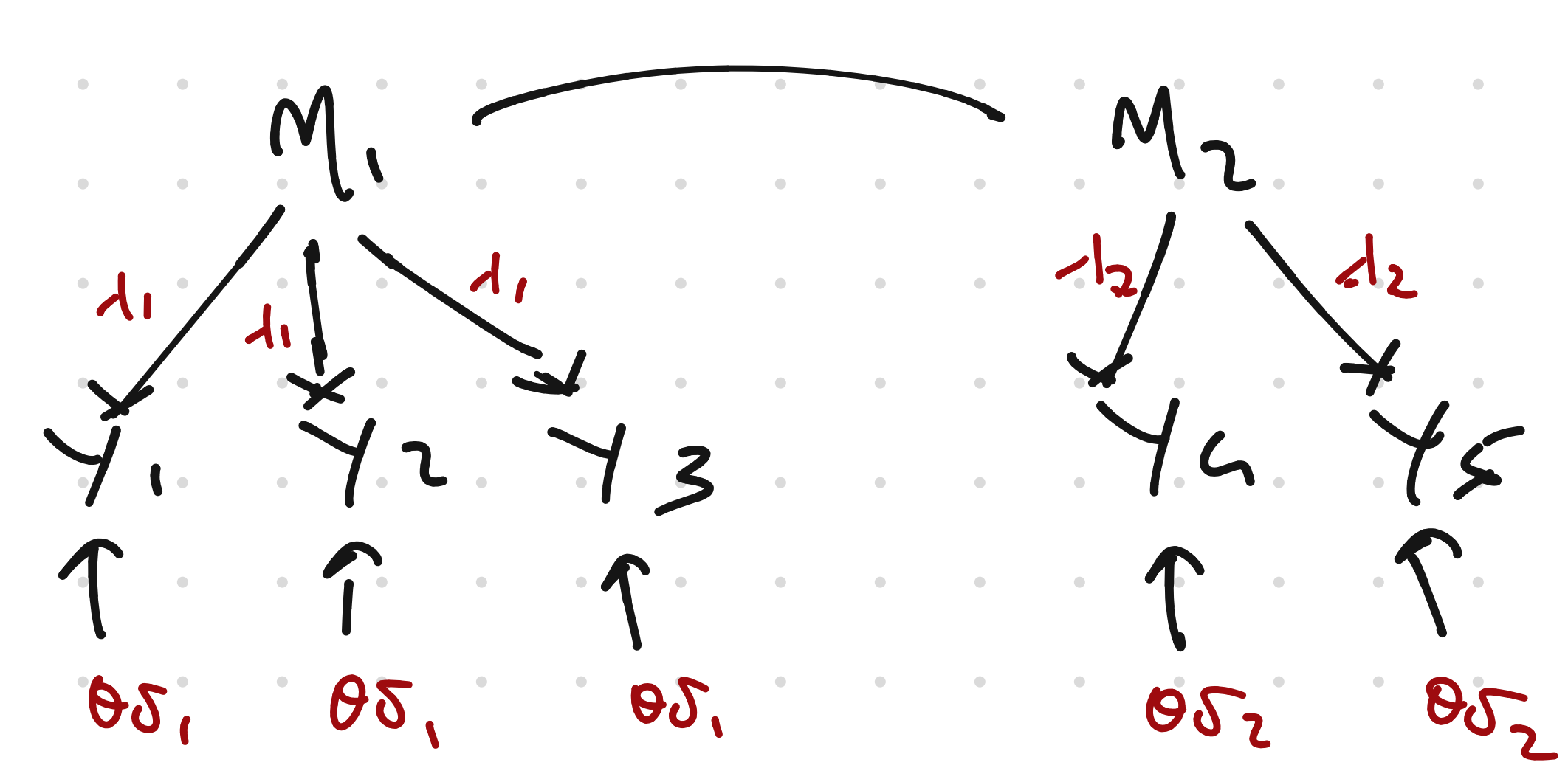
- 1) I tre modelli notevoli sono annidati (nested: parallelo > r-equiv. > congruente)
- 2) Caso multidimensionale: i tre modelli possono essere definiti: (A) a livello del test complessivo (B) a livello delle scale del test



Esempio: (caso multidimensionale di test paralleli)



caso A: test complessivo



caso B: singole scale

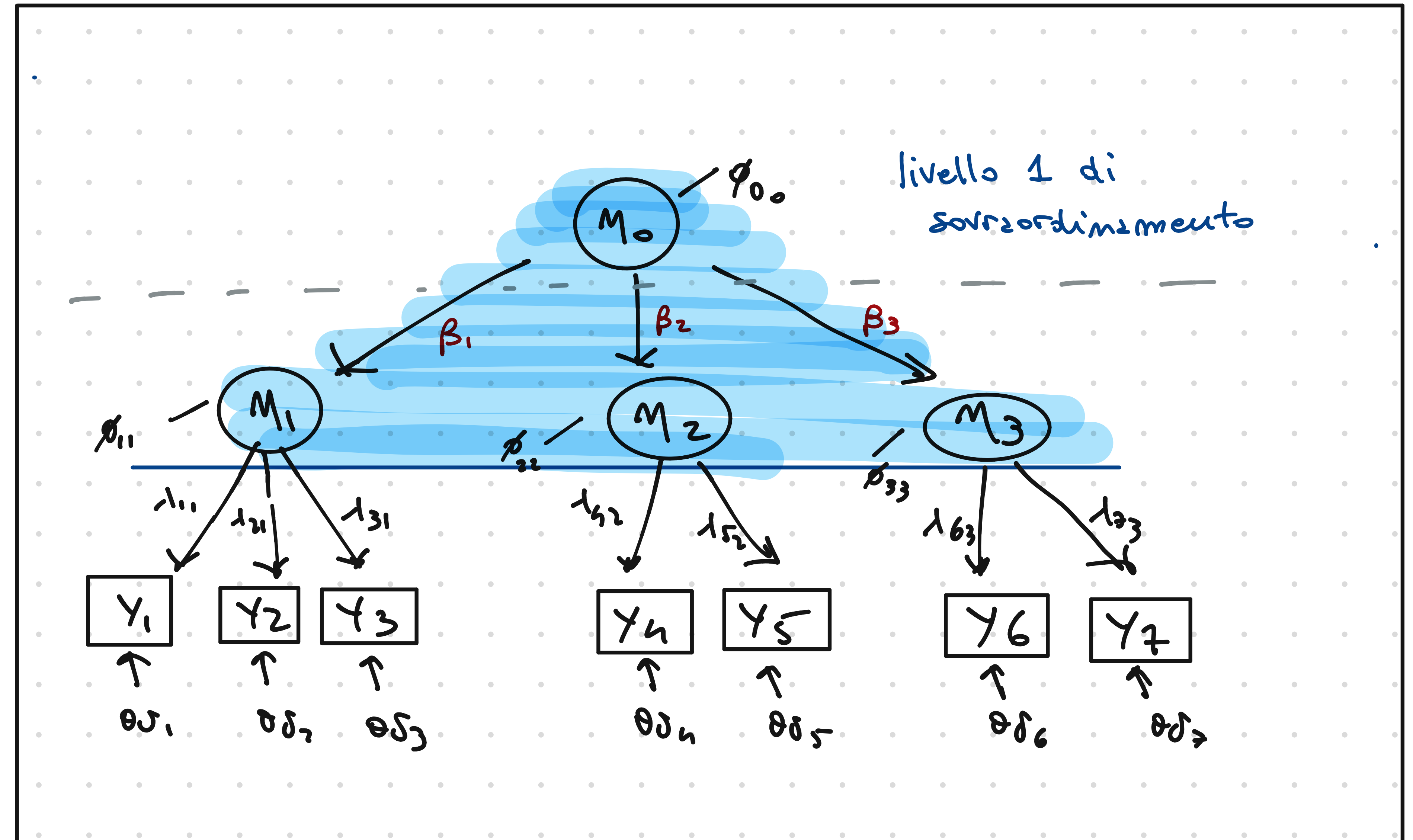
### Modello sovra-ordinato

(caso ad 1 livello di sovra-ordinamento)

$$\Sigma_Y = \Lambda (\beta \beta^T + \Phi) \Lambda^T + \theta \theta$$

- 1)  $\Lambda$  ha dimensione  $m \times (Q+1)$
- 2)  $\beta$  è un vettore  $Q \times 1$
- 3)  $\Phi$  ha dimensione  $(Q+1) \times (Q+1)$  ed è diagonale (l'introduzione di  $\eta_0$  annulla le correlazioni tra le altre v. latenti)

Es:  $\Phi = \begin{pmatrix} \phi_{11} & \phi_{21} & \phi_{31} & 0 \\ 0 & \phi_{22} & \phi_{32} & \phi_{00} \\ 0 & 0 & \phi_{33} & 0 \\ 0 & 0 & 0 & \phi_{00} \end{pmatrix}$   
 $(3+1) \times (3+1)$

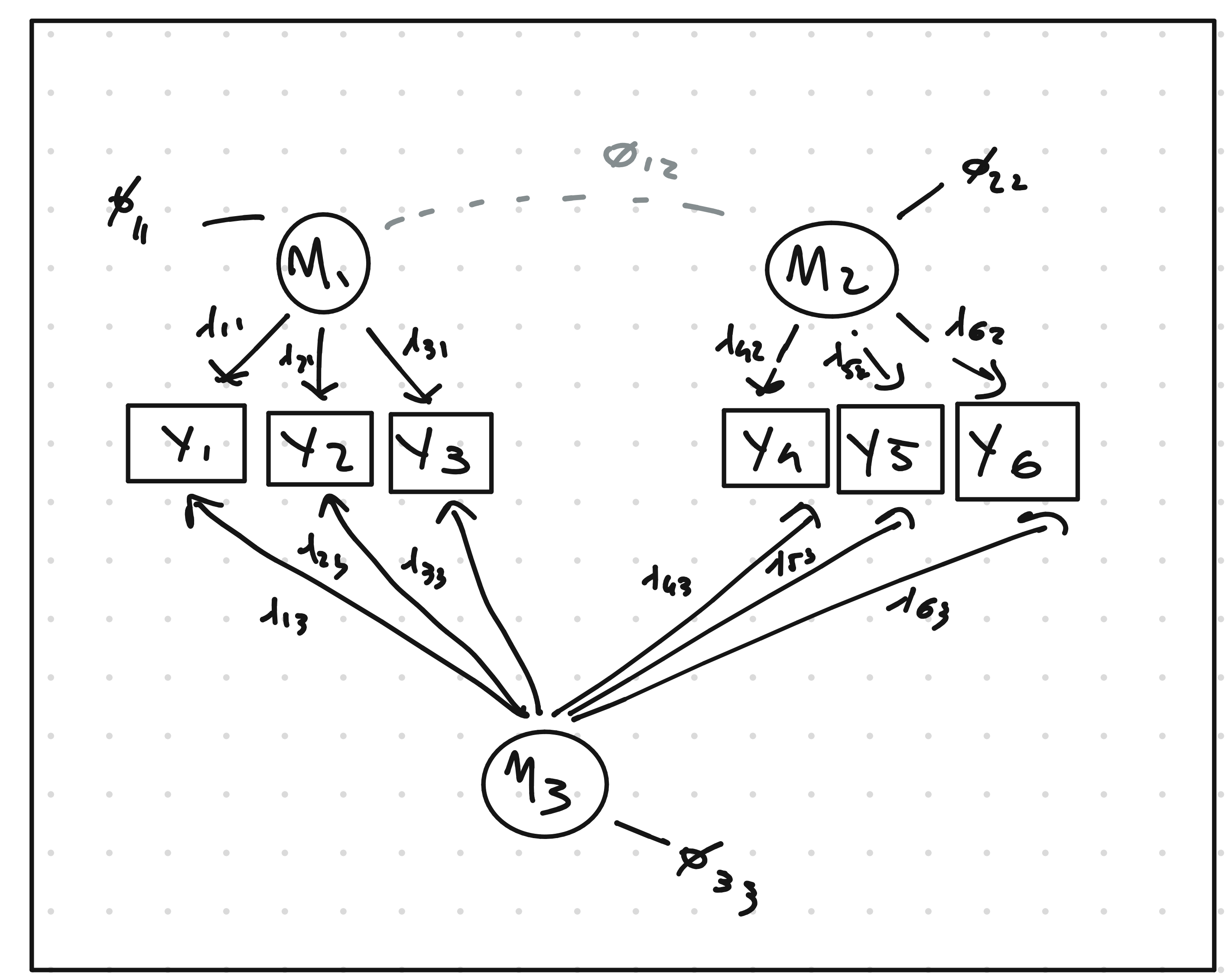


Note: l'introduzione di  $\eta_0$  non è mai suggerita dagli indici di modifica e ha senso solo se il modello di base (ie, confederico) si adatta bene ai dati. Lo scopo di questi modelli è quello di provare a "spiegare" le correlazioni tra le v. latenti del modello di base mediante l'introduzione di una v. latente "più generale" pre-esistente a quelle del modello di base.

### Modello bi-factor

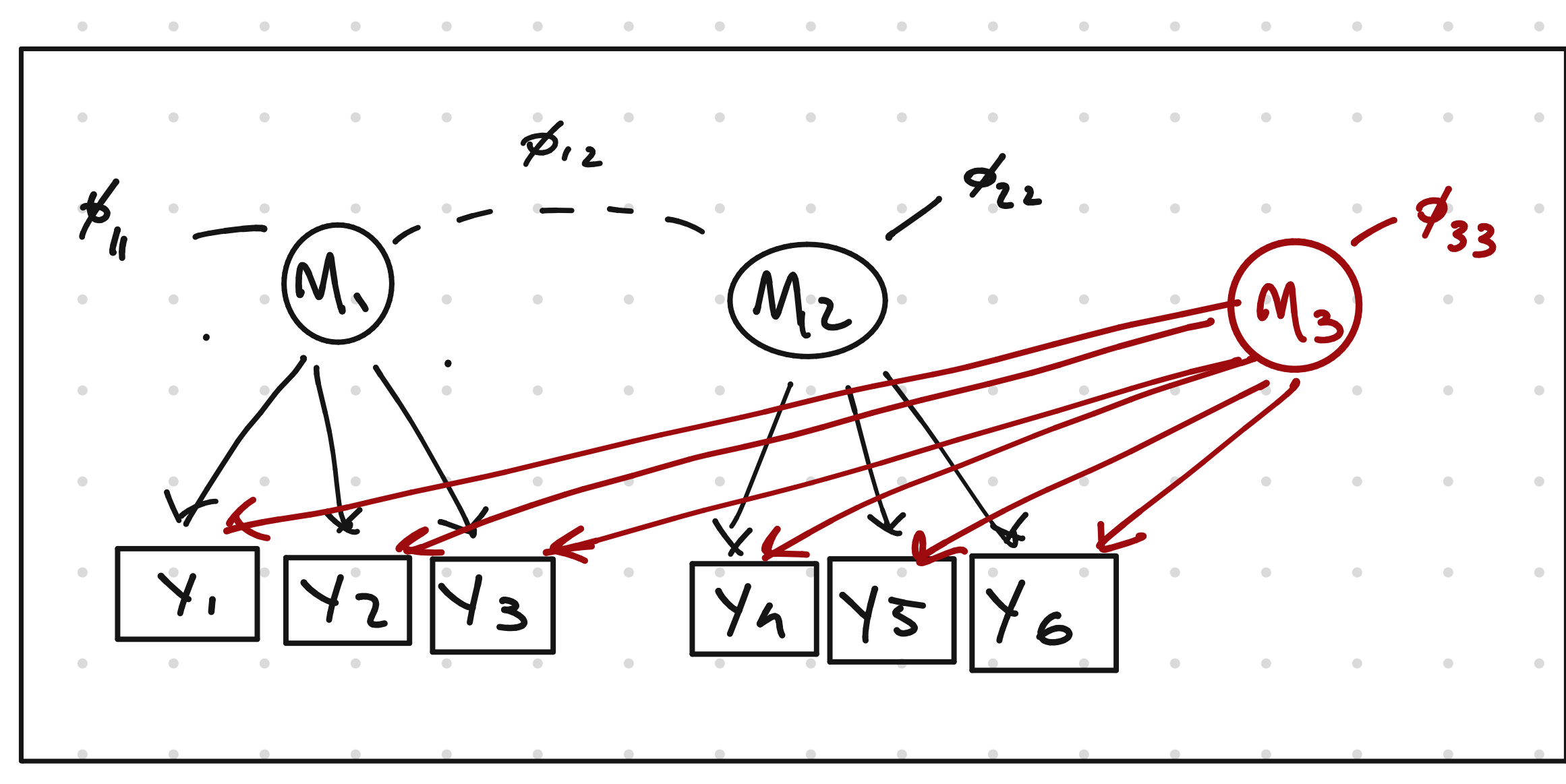
$$\Sigma_Y = \Lambda \Phi \Lambda^T + \theta \theta$$

- 1)  $\eta_3$  è una v. latente che si aggiunge alle precedenti
- 2)  $\eta_3$  non è collegato alle v. d'errore di  $Y_1 - Y_m$
- 3)  $\Phi$  può essere costruita in due modi:
  - 3.a) diagonale (ortogonalità completa)
  - 3.b)  $\eta_1 \perp \eta_3$  e  $\eta_2 \perp \eta_3$  (ortogon. parziale)



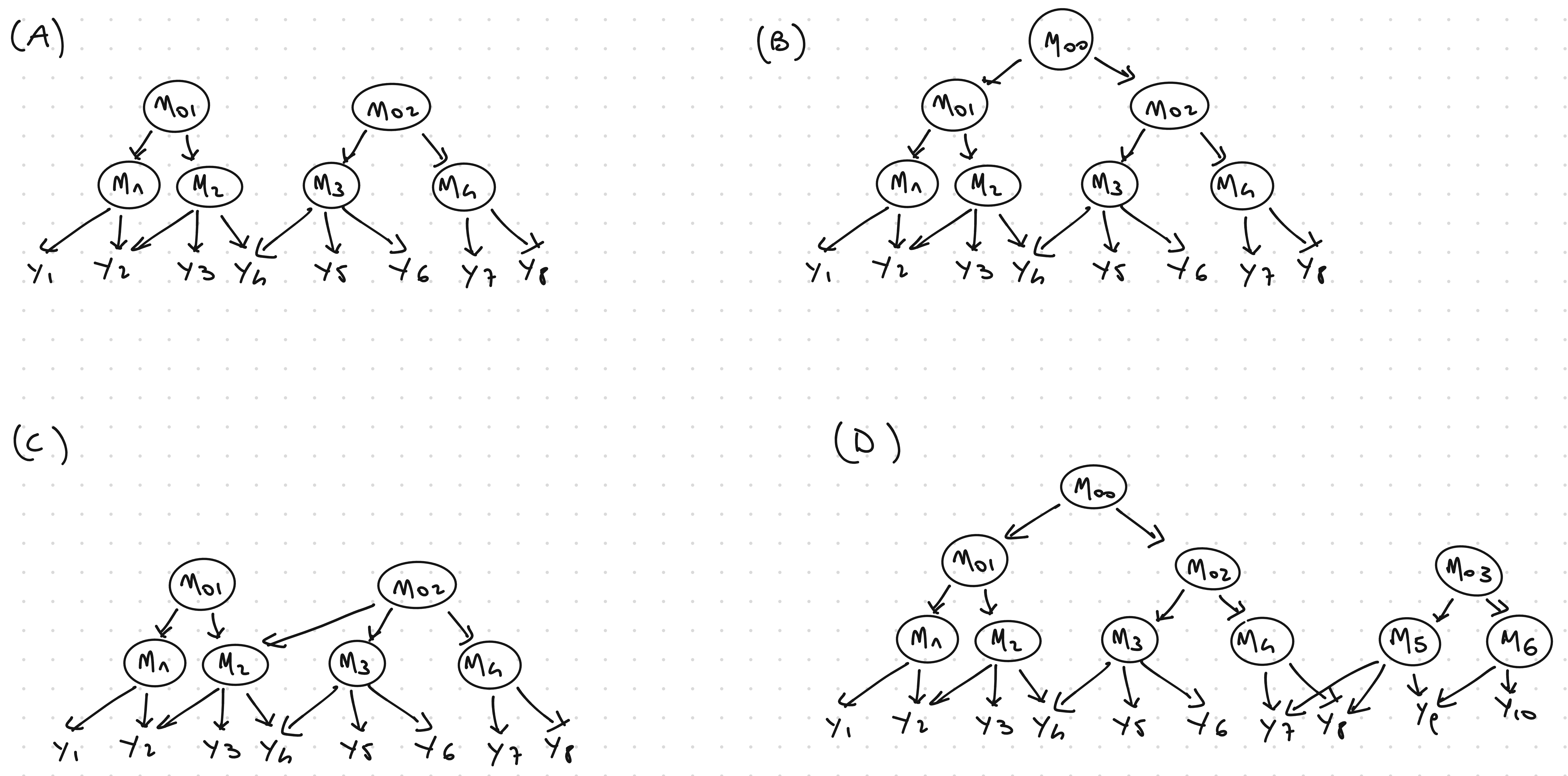
Note: in figura è rappresentato il caso 3b. Il caso 3a si ottiene, in figura, ponendo  $\phi_{12} = 0$ .

Modello 2

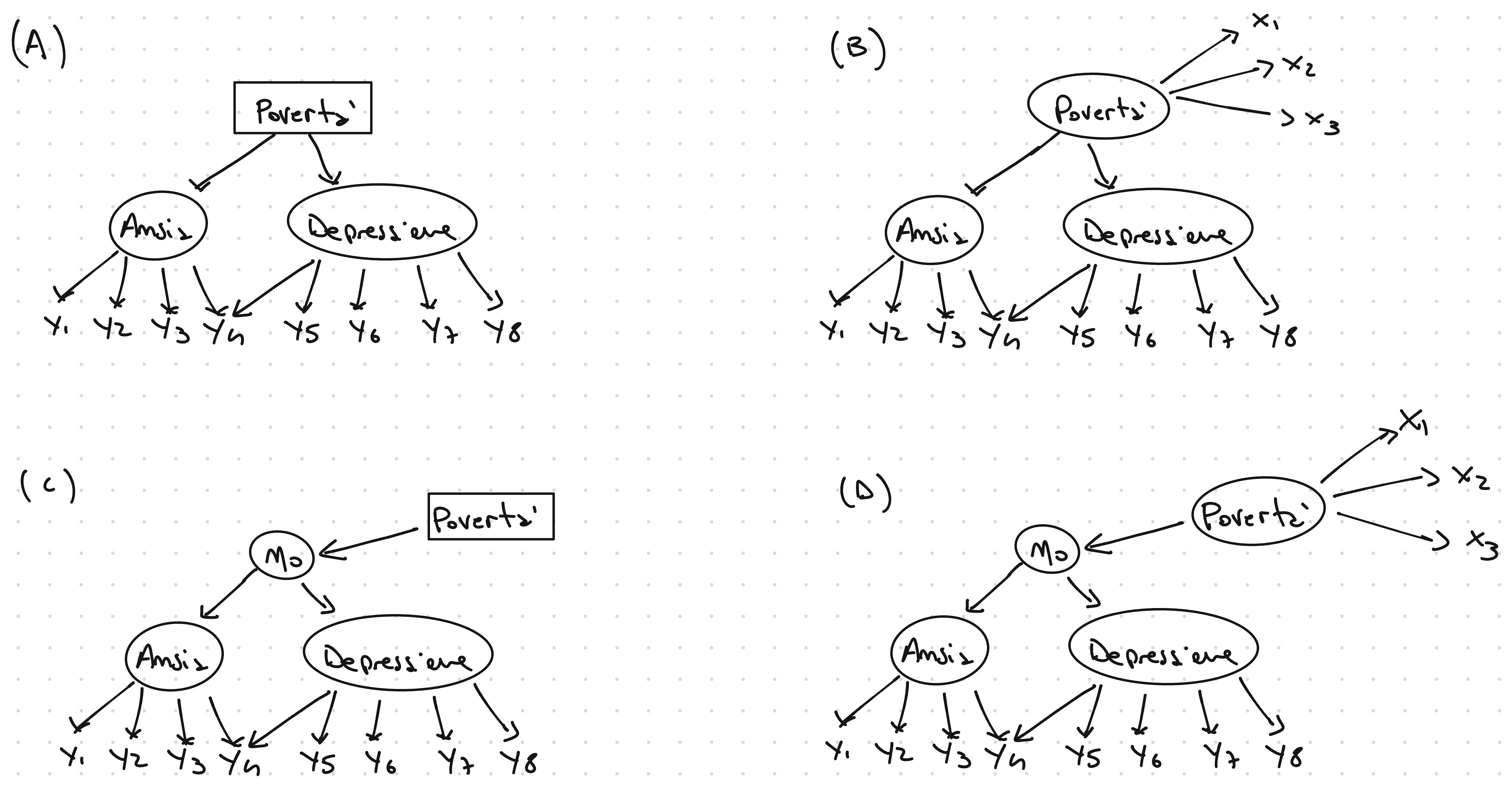




Di seguito si presentano alcuni casi esemplificativi di modelli sovrordinati.  
Per semplicità si propongono grafici semplificati rispetto ai parametri del modello.



Si riportano inoltre alcuni diagrammi esemplificativi di modelli apparentemente sovrordinati nel contesto più generale dei SEMs. Per ragioni di esemplificazione si consideri il caso di due v. latenti (ansia, depressione) e una variabile concomitante (povertà) in entrambi i casi di misurazione della variabile direttamente e indirettamente (mediante ulteriore modello CFA).



Caso A: la v. concomitante "povertà" è misurata direttamente e agisce sulle v. latenti del modello. In questo caso, queste v. latenti variano in funzione della v. concomitante. Se questa è di tipo categoriale (es, v. a 3 livelli) avremo tanti modelli CFA quanti sono i livelli della v. concomitante (es, 3 sottomodelli).



Caso B: come il caso A con la differenza che la v. concomitante "povertà" non è misurata direttamente ma usando un modello CFA.

Caso C-D: come per i casi precedenti con la differenza che la v. concomitante "povertà" (sia nel caso di misurazione diretta o indiretta) agisce sulla v. latente subordinata  $\eta_0$ .