

UNIVERSITÀ DEGLI STUDI DI PADOVA  
DIPARTIMENTO DI SCIENZE STATISTICHE  
CORSO DI LAUREA MAGISTRALE IN  
SCIENZE STATISTICHE



**Studio di curve di mortalità:  
un'analisi tramite scomposizioni tensoriali**

Relatore Prof. Emanuele Aliverti  
Dipartimento di Scienze Statistiche

Laureando Paolo Dallavalle  
Matricola 2021412

Anno Accademico 2022/2023



# Indice

<b>Introduzione</b>	<b>1</b>
<b>1 La distribuzione dell'età alla morte</b>	<b>3</b>
1.1 Fonte dei dati . . . . .	3
1.2 Descrizione e caratteristiche dei dati . . . . .	4
<b>2 Introduzione ai metodi tensoriali</b>	<b>5</b>
2.1 Tensori: nozioni, caratteristiche e proprietà . . . . .	5
2.1.1 Definizione e notazione . . . . .	5
2.1.2 Tensore diagonale . . . . .	7
2.1.3 <i>Matricizzazione</i> . . . . .	7
2.1.4 Prodotti matriciali di Kronecker, di Khatri–Rao e di Hadamard . . . . .	7
2.1.5 Norma di Frobenius, prodotto interno e prodotto esterno . . . . .	9
2.1.6 Rango tensoriale . . . . .	10
2.1.7 Prodotto <i>n-modale</i> . . . . .	10
2.1.8 Rango <i>n-esimo</i> . . . . .	11
2.2 Scomposizione CANDECOMP/PARAFAC . . . . .	11
2.3 Scomposizione di Tucker . . . . .	14
2.4 Algoritmo EM per clustering tensoriale . . . . .	17
2.5 Clustering multi-direzionale tramite il modello tensoriale a blocchi . . . . .	21
<b>3 Applicazione ai dati</b>	<b>25</b>
3.1 Creazione dei tensori . . . . .	25
3.2 Riduzione della complessità tramite scomposizione tensoriale CP . . . . .	25
3.3 Riduzione della complessità tramite scomposizione di Tucker . . . . .	30
3.4 Ricerca di gruppi tra i tensori tramite DEEM clustering . . . . .	35
3.5 Ricerca di gruppi all'interno dei tensori tramite TBM clustering . . . . .	39
<b>4 Conclusioni</b>	<b>45</b>
<b>Appendice</b>	<b>47</b>
<b>Bibliografia</b>	<b>51</b>



# Notazioni

$\mathcal{X} \in \mathbb{R}^{I_1 \times \dots \times I_N}$	Tensore a valori reali di ordine $N$ e dimensioni $I_1 \times \dots \times I_N$
$A \in \mathbb{R}^{I \times J}$	Matrice a valori reali di dimensioni $I \times J$
$a \in \mathbb{R}^I$	Vettore a valori reali di dimensione $I$
$\mathcal{X}_{i_n=1}$	Sotto tensore di $\mathcal{X}$ fissato il valore della moda $i_n$ pari a 1
$x_{l;i,j,k}$	Generico elemento $i, j, k$ del tensore a 3 vie $\mathcal{X}_l$
$x_{l;:,j,k}$	Fibra di moda 1 del tensore a 3 vie $\mathcal{X}_l$
$X_{l;::,k} = X_k$	<i>Slices</i> frontali del tensore a 3 vie $\mathcal{X}_l$
$a \circ b$	Prodotto esterno tra il vettore $a$ e il vettore $b$
$\text{rank}(\mathcal{X})$	Rango tensoriale di $\mathcal{X}$
$X_{(n)}$	<i>Matricizzazione</i> (o <i>unfolding</i> ) della moda $n$ del tensore $\mathcal{X}$
$A \otimes B$	Prodotto di Kronecker tra la matrice $A$ e la matrice $B$
$A \odot B$	Prodotto di Khatri-Rao tra la matrice $A$ e la matrice $B$
$A \star B$	Prodotto di Hadarmard tra la matrice $A$ e la matrice $B$
$A^\dagger$	Pseudo-inversa di <i>Moore-Penrose</i> della matrice $A$
$\ \mathcal{X}\ _F$	Norma di Frobenius del tensore $\mathcal{X}$
$\ \mathcal{X}\ _F^2$	Quadrato della norma di Frobenius del tensore $\mathcal{X}$
$\langle \mathcal{X}, \mathcal{Y} \rangle$	Prodotto interno tra i tensori $\mathcal{X}$ e $\mathcal{Y}$
$\mathcal{X} \times_n A$	Prodotto <i>n-modale</i> tra il tensore $\mathcal{X}$ e la matrice $A$
$\text{rank}_n(\mathcal{X})$	Rango <i>n-esimo</i> del tensore $\mathcal{X}$
$[[\mathcal{X}; A^{(1)}, \dots, A^{(N)}]]$	Serie di prodotti <i>modali</i> ordinati tra il tensore $\mathcal{X}$ e le $N$ matrici $A^{(n)}$ per ogni $n = 1, \dots, N$
$TN(\mu; \Sigma_1, \dots, \Sigma_N)$	Distribuzione normale tensoriale con media $\mu$ e matrice di varianza lungo la moda $n$ pari a $\Sigma_n$ per ogni $n = 1, \dots, N$



# Introduzione

A partire dall'istituzione dell'Organizzazione delle Nazioni Unite (*ONU*) e dalla conseguente fondazione dell'Organizzazione Mondiale della Sanità (*OMS*) lo sviluppo socio-sanitario ha gradualmente portato ad importanti miglioramenti del livello di mortalità in tutto il mondo. Questa diminuzione della mortalità ha inevitabilmente avuto un impatto sulla distribuzione dell'età alla morte, il cui studio rappresenta un importante strumento a supporto dei governi per misurare e garantire la sostenibilità dei sistemi di sicurezza sociale e di assistenza sanitaria.

Negli ultimi vent'anni sono stati intrapresi molti studi, che hanno coinvolto gruppi di persone differenziate per diverse caratteristiche, cercando di individuare le determinanti di questa distribuzione. La maggior parte di questi studi utilizzano classiche metodologie statistiche, parametriche o non parametriche, che si basano sulla ricostruzione dei dati multidimensionali in formato vettoriale o matriciale perdendo così la struttura multilineare dei dati.

In questo lavoro di tesi si vuole fornire un approccio differente allo studio della distribuzione dell'età alla morte. In particolare, si vuole trattare i dati relativi alla distribuzione dell'età alla morte separatamente per i maschi e per le femmine di 12 diversi paesi negli ultimi 9 decenni come 12 tensori a 3 vie/mode (uno per ogni paese) così da sfruttare la struttura multilineare dei dati. Verranno utilizzati metodi di scomposizione tensoriale per cercare di mantenere l'informazione riducendo la complessità del tensore (rappresentata dal rango tensoriale) e metodi di *clustering* tensoriale per cercare dei gruppi tra tensori e blocchi all'interno dei tensori stessi.

Nel primo capitolo si presentano i dati, si descrive come sono stati ottenuti e si evidenziano alcune caratteristiche.

Nel secondo capitolo si introducono alcune nozioni e proprietà dei tensori. Successivamente, vengono introdotte la scomposizione *CANDECOMP/PARAFAC*, la scomposizione di Tucker, l'algoritmo di *clustering Doubly Enhanced Expectation Maximization* (DEEM) e il metodo di *clustering Tensor Block Model* (TBM).

Nel terzo e ultimo capitolo viene presa in considerazione l'applicazione dei metodi illustrati ai dati di età alla morte presentati nel primo capitolo. L'analisi viene svolta utilizzando in particolare i pacchetti R specifici per dati tensoriali: *rTensor*, *tensorsparse*, *TensorClustering*.

Nell'appendice sono riportati gli pseudo codici relativi alle procedure illustrate nel secondo capitolo.



# Capitolo 1

## La distribuzione dell'età alla morte

Nel corso dell'ultimo secolo, si è assistito a importanti miglioramenti del livello di mortalità in tutto il mondo. Questa diminuzione della mortalità ha inevitabilmente portato a modifiche sostanziali nella forma della distribuzione dell'età alla morte, con importanti conseguenze a livello socio-politico che hanno obbligato i governi di tutto il mondo ad attuare decisioni tra le quali spiccano quelle delle riforme fiscali e del sistema pensionistico. Il progresso di queste curve è di particolare interesse per individuarne le determinanti così da prendere decisioni sulla base delle possibili evoluzioni future.

Lo studio presentato in questa tesi cerca di analizzare queste curve prendendo in considerazione le diverse caratteristiche in maniera simultanea tramite il formato tensoriale dei dati.

### 1.1 Fonte dei dati

Dal sito *Human Mortality Database* sono state scaricate le tavole di mortalità per 12 paesi (Belgio, Danimarca, Finlandia, Francia, Islanda, Italia, Norvegia, Spagna, Svezia, Svizzera, Regno Unito e Stati Uniti) prendendo le tavole  $1 \times 10$  (*Age interval*  $\times$  *Year interval*) e mantenendo poi solo i periodi comuni (cioè i 9 decenni a partire da quello degli anni 30 del ventesimo secolo fino al secondo del ventunesimo secolo). Di queste tavole è stato preso solo il dato relativo ai decessi (dx). In particolare, per ogni sesso di ogni paese in ogni periodo è stato diviso il valore dei decessi ad ogni età per la somma totale dei decessi stessi ottenendo così la distribuzione dell'età alla morte.

Per gli Stati Uniti non sono disponibili i dati dal 1930 al 1932 ma prendendo in considerazione delle proporzioni di morti nelle varie età sul totale di morti si è mantenuto

il dato relativo agli anni dal 1933 al 1939 come dato per il decennio relativo agli anni '30 del ventesimo secolo.

## 1.2 Descrizione e caratteristiche dei dati

I dati analizzati in questo elaborato sono curve di distribuzione dell'età alla morte divise per sesso, per i 12 diversi paesi nei 9 decenni considerati. In totale si dispone quindi di 216 curve che evidenziano degli scostamenti nella mortalità infantile (0-5 anni) e nella mortalità per le persone adulte, come si può notare dalla Figura 1.1.

Dalla prima immagine in Figura 1.1 si notano le differenze delle curve al variare del decennio considerato e si vede come le curve si siano sempre più spostate verso le età anziane con il passare degli anni. Dalla seconda immagine in Figura 1.1 risultano evidenti le differenze di genere, infatti le femmine sembrano avere curve più spostate verso le età anziane rispetto i maschi. Infine, da entrambe le Figure si notano 3/4 curve che hanno un picco intorno ai 20 anni, queste curve sono relative ai maschi nei primi decenni considerati e sono quelle caratterizzate dalle morti durante la seconda guerra mondiale.

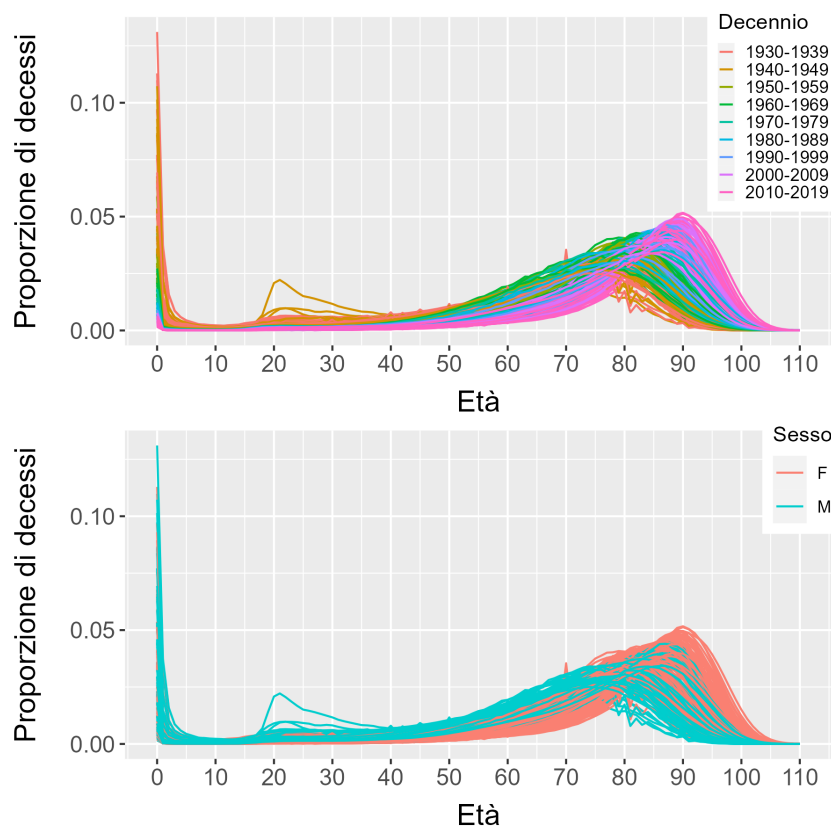


FIGURA 1.1: Curve di distribuzione dell'età alla morte prima colorate per decennio poi colorate per sesso.

# Capitolo 2

## Introduzione ai metodi tensoriali

### 2.1 Tensori: nozioni, caratteristiche e proprietà

I metodi classici di analisi dei dati ad alta dimensionalità si basano sulla ricostruzione dei dati multidimensionali in formato vettoriale o matriciale. Tuttavia, questa *vettorizzazione* o *matricizzazione* fa perdere la struttura multilineare dei dati portando così a prestazioni non ottimali dell’elaborazione. Un tensore, invece, è una generalizzazione di vettori e matrici, ed è la miglior rappresentazione se si vuole “rispettare” questa struttura dei dati.

#### 2.1.1 Definizione e notazione

Un tensore è un *array* multidimensionale il cui ordine è definito come il numero di dimensioni, anche note come vie o mode. Più formalmente un tensore a  $N$  vie (o di ordine  $N$ -esimo) è un elemento del prodotto tensoriale di  $N$  spazi vettoriali, ognuno dei quali ha un proprio sistema di coordinate. Il tensore è una generalizzazione di vettore e di matrice infatti un tensore ad una via (tensore di primo ordine) è un vettore mentre un tensore a due vie (tensore di secondo ordine) è una matrice. In Figura 2.1 si mostra un tensore di terzo ordine.

Per cercare di rimanere coerenti con la letteratura sull’analisi tensoriale, la notazione proposta in questo lavoro è la stessa usata in Kolda & Bader (2009), quindi l’elemento  $(i, j, k)$  di un tensore di terzo ordine  $\mathcal{X}$  viene definito come  $x_{ijk}$ .

Le fibre di un tensore sono l’equivalente multidimensionale delle righe e delle colonne di una matrice, infatti una fibra è definita fissando tutti gli indici tranne uno. In una matrice una colonna è una fibra di moda 1 mentre una riga è una fibra di moda 2. In un tensore di terzo ordine le fibre di moda 3 sono chiamate tubi.

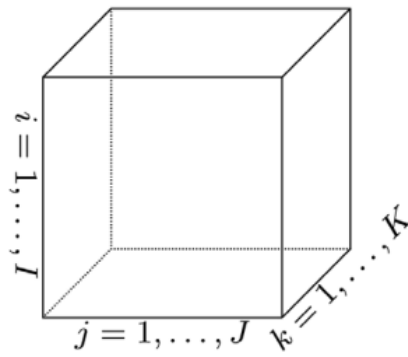
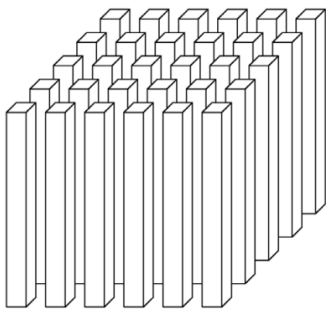
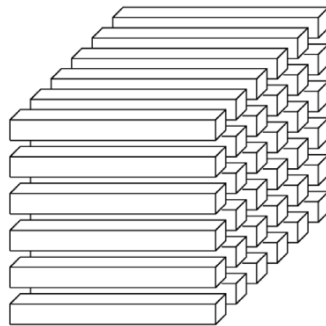


FIGURA 2.1: Tensore di terzo ordine. Figura presa da Kolda & Bader (2009).

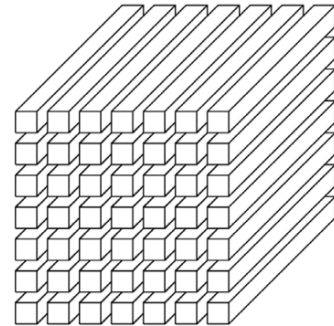
Gli *slices* di un tensore sono le sue sezioni bidimensionali, ottenute quindi fissando tutti gli indici tranne due. Nella Figura 2.2 si mostrano fibre e *slices* di un tensore di terzo ordine.



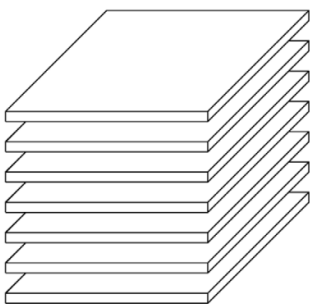
(a) Fibre di moda 1 (colonne):  $x_{:jk}$



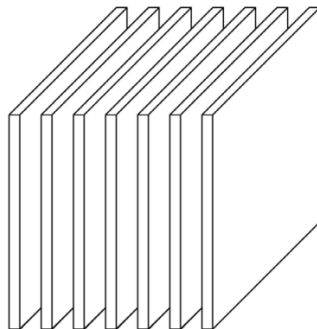
(b) Fibre di moda 2 (righe):  $x_{i:k}$



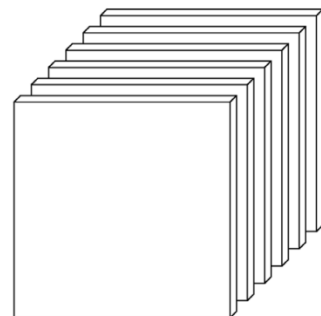
(c) Fibre di moda 3 (tubi):  $x_{ij:}$



(d) Slices orizzontali:  $\mathcal{X}_{i:}$



(e) Slices laterali:  $\mathcal{X}_{:j}$



(f) Slices frontali:  $\mathcal{X}_{::k}$

FIGURA 2.2: Fibre e *Slices* di un tensore di terzo ordine. Immagini prese da Kolda & Bader (2009).

### 2.1.2 Tensore diagonale

Un tensore  $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$  è diagonale se

$$x_{i_1, i_2, \dots, i_N} \neq 0 \iff i_1 = i_2 = \dots = i_N.$$

In Figura 2.3 si mostra un tensore diagonale di terzo ordine con la *superdiagonale* di soli uni.

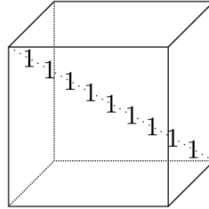


FIGURA 2.3: Tensore diagonale di terzo ordine con uni sulla *superdiagonale*. Figura presa da Kolda & Bader (2009)

### 2.1.3 Matricizzazione

La *matricizzazione* anche nota come *unfolding* è il processo di riordinamento degli elementi di un tensore di ordine  $N$  in una matrice. La *matricizzazione* della  $n$ -esima moda di un tensore  $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$  denotata con  $X_{(n)}$  dispone le fibre di moda  $n$  come colonne della matrice risultante. Come esempio si riporta l'*unfolding* della prima moda di un tensore di terzo ordine  $\mathcal{X}$  di dimensioni  $I_1 \times I_2 \times I_3$

$$\begin{aligned} X_{(1)} &= \begin{bmatrix} x_{:11} & x_{:21} & \dots & x_{:I_2 1} & \dots & \dots & x_{:1I_3} & x_{:2I_3} & \dots & x_{:I_2 I_3} \end{bmatrix} = \\ &= \begin{bmatrix} x_{111} & x_{121} & \dots & x_{1I_2 1} & \dots & \dots & x_{11I_3} & x_{12I_3} & \dots & x_{1I_2 I_3} \\ x_{211} & x_{221} & \dots & x_{2I_2 1} & \dots & \dots & x_{21I_3} & x_{22I_3} & \dots & x_{2I_2 I_3} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_{I_1 11} & x_{I_1 21} & \dots & x_{I_1 I_2 1} & \dots & \dots & x_{I_1 1 I_3} & x_{I_1 2 I_3} & \dots & x_{I_1 I_2 I_3} \end{bmatrix}. \end{aligned}$$

Per ulteriori dettagli sul concetto di *unfolding* si rimanda a Kolda (2006).

### 2.1.4 Prodotti matriciali di Kronecker, di Khatri–Rao e di Hadamard

Il prodotto di Kronecker tra due matrici  $A \in \mathbb{R}^{I \times J}$  e  $B \in \mathbb{R}^{K \times L}$ , indicato con  $A \otimes B$ , risulta essere una matrice di dimensioni  $(IK) \times (JL)$  definita come

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1J}B \\ a_{21}B & a_{22}B & \cdots & a_{2J}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{I1}B & a_{I2}B & \cdots & a_{IJ}B \end{bmatrix},$$

$$A \otimes B = \begin{bmatrix} a_1 \otimes b_1 & a_1 \otimes b_2 & \cdots & a_J \otimes b_{L-1} & a_J \otimes b_L \end{bmatrix}.$$

Il prodotto di Khatri-Rao è il prodotto di Kronecker tra colonne corrispondenti di due matrici con lo stesso numero di colonne. Siano  $A \in \mathbb{R}^{I \times K}$  e  $B \in \mathbb{R}^{J \times K}$ , il loro prodotto di Khatri-Rao, indicato con  $A \odot B$ , risulta essere una matrice di dimensioni  $(IJ) \times K$  definita come

$$A \odot B = \begin{bmatrix} a_{:1} \otimes b_{:1} & a_{:2} \otimes b_{:2} & \cdots & a_{:K} \otimes b_{:K} \end{bmatrix}.$$

Il prodotto di Hadamard è il prodotto tra elementi corrispondenti di due matrici delle stesse dimensioni. Date due matrici  $A$  e  $B$  entrambe di dimensioni  $I \times J$  il loro prodotto di Hadamard (qui indicato con  $A \star B$ ) è definito come

$$A \star B = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & \cdots & a_{1J}b_{1J} \\ a_{21}b_{21} & a_{22}b_{22} & \cdots & a_{2J}b_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ a_{I1}b_{I1} & a_{I2}b_{I2} & \cdots & a_{IJ}b_{IJ} \end{bmatrix}.$$

Questi prodotti hanno diverse proprietà, ma per questo lavoro è utile richiamare in particolare le seguenti tre

$$A \odot B \odot C = (A \odot B) \odot C = A \odot (B \odot C), \quad (2.1)$$

$$(A \odot B)^T (A \odot B) = (A^T A) \star (B^T B), \quad (2.2)$$

$$(A \odot B)^\dagger = ((A^T A) \star (B^T B))^\dagger (A \odot B)^T, \quad (2.3)$$

dove  $A^\dagger$  indica la pseudo-inversa di *Moore-Penrose* della matrice  $A$  (la cui definizione è illustrata in Gloub & Van Loan (1996)). La proprietà (2.3) permette perciò di passare dal dover svolgere la pseudo-inversa di una matrice  $IJ \times K$  allo svolgere la pseudo-inversa di una matrice  $K \times K$ . Per ulteriori approfondimenti e dettagli su questi prodotti matriciali e sulle loro proprietà si rimanda a Gloub & Van Loan (1996).

### 2.1.5 Norma di Frobenius, prodotto interno e prodotto esterno

La norma di Frobenius del tensore  $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$  è definita come la radice quadrata della somma dei quadrati di tutti i suoi elementi

$$\|\mathcal{X}\|_F = \sqrt{\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \dots \sum_{i_N=1}^{I_N} x_{i_1, i_2, \dots, i_N}^2} . \quad (2.4)$$

Il prodotto interno di due tensori delle stesse dimensioni  $\mathcal{X}, \mathcal{Y} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$  è definito come la somma dei prodotti di ogni elemento di  $\mathcal{X}$  per l'elemento corrispondente di  $\mathcal{Y}$ :

$$\langle \mathcal{X}, \mathcal{Y} \rangle = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \dots \sum_{i_N=1}^{I_N} x_{i_1, i_2, \dots, i_N} y_{i_1, i_2, \dots, i_N} . \quad (2.5)$$

Di conseguenza il prodotto interno di un tensore per sé stesso equivale alla sua norma di Frobenius al quadrato

$$\langle \mathcal{X}, \mathcal{X} \rangle = \|\mathcal{X}\|_F^2 .$$

Dati due vettori  $a, b$  di dimensioni  $m \times 1$  e  $n \times 1$  rispettivamente, il loro prodotto esterno (denotato con  $a \circ b$ ) è definito come una matrice  $m \times n$  in cui ogni elemento di  $a$  è moltiplicato per ogni elemento di  $b$

$$a \circ b = \begin{bmatrix} a_1 b_1 & a_1 b_2 & \dots & a_1 b_n \\ a_2 b_1 & a_2 b_2 & \dots & a_2 b_n \\ \vdots & \vdots & \ddots & \vdots \\ a_m b_1 & a_m b_2 & \dots & a_m b_n \end{bmatrix} .$$

Il prodotto esterno tra  $N$  vettori  $a^{(1)}, a^{(2)}, \dots, a^{(N)}$  di dimensioni rispettivamente  $I_1, I_2, \dots, I_N$  sarà definito come un tensore di  $N$ -esimo ordine

$$a^{(1)} \circ a^{(2)} \circ \dots \circ a^{(N)} = \mathcal{X}, \text{ con } \mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N} . \quad (2.6)$$

Questo significa che ogni elemento del tensore  $\mathcal{X}$  è il prodotto dei corrispondenti elementi dei vettori  $a^{(n)}$  con  $n = 1, \dots, N$

$$x_{i_1, i_2, \dots, i_N} = a_{i_1}^{(1)} a_{i_2}^{(2)} \dots a_{i_N}^{(N)}, \quad \forall 1 \leq i_n \leq I_n .$$

### 2.1.6 Rango tensoriale

Un tensore  $\mathcal{X}$  a  $N$  vie è di rango 1 se può essere scritto nella forma (2.6), cioè come prodotto esterno di  $N$  vettori. La Figura 2.4 illustra graficamente un tensore a 3 vie di rango 1 e la sua rappresentazione come prodotto esterno di 3 vettori.

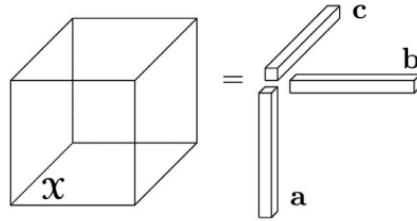


FIGURA 2.4: Tensore a 3 vie di rango 1,  $\mathcal{X} = a \circ b \circ c$ . Figura presa da Kolda & Bader (2009).

Il rango di un tensore  $\mathcal{X}$ , indicato con  $rank(\mathcal{X})$ , è definito come il più piccolo numero di tensori di rango 1 la cui somma genera  $\mathcal{X}$ . Tale definizione deriva da Hitchcock (1927) che introdusse l'idea della forma poliadica di un tensore, ovvero la rappresentazione di un tensore come somma di tensori di rango 1 e quindi rappresentabili a loro volta come prodotti esterni di vettori (la Figura 2.5 rappresenta graficamente la forma poliadica di un tensore di terzo ordine).

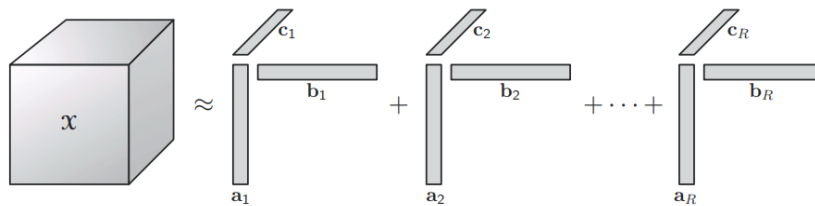


FIGURA 2.5: Forma poliadica di un tensore di terzo ordine. Figura presa da Kolda & Bader (2009).

### 2.1.7 Prodotto $n$ -modale

Il prodotto  $n$  modale tra un tensore  $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_n}$  e una matrice  $A \in \mathbb{R}^{J \times I_N}$ , indicato con  $\mathcal{X} \times_n A$ , risulta essere un tensore di dimensione  $I_1 \times \dots \times I_{n-1} \times J \times I_{n+1} \times \dots \times I_N$  con

$$(\mathcal{X} \times_n A)_{i_1, \dots, i_{n-1}, j, i_{n+1}, \dots, i_N} = \sum_{i_n=1}^{I_n} x_{i_1, i_2, \dots, i_N} \cdot a_{j, i_n}.$$

Si illustrano due proprietà del prodotto  $n$  modale che saranno utili in seguito:



- Il prodotto  $n$ -modale può anche essere espresso in termini di *matricizzazione*

$$\mathcal{Y} = \mathcal{X} \times_n A \iff Y_{(n)} = AX_{(n)} .$$

- In una serie di prodotti modali su diverse mode l'ordine della moltiplicazione è irrilevante

$$\mathcal{X} \times_m A \times_n B = \mathcal{X} \times_n B \times_m A \quad (m \neq n) .$$

Si rimanda a Bader & Kolda (2006) per una trattazione più estesa sui diversi concetti di moltiplicazione nei contesti tensoriali.

### 2.1.8 Rango $n$ -esimo

Sia  $\mathcal{X}$  un tensore di  $N$ -esimo ordine di dimensioni  $I_1 \times I_2 \times \dots \times I_N$  allora il *rango  $n$ -esimo* di  $\mathcal{X}$ , indicato con  $rank_n(\mathcal{X})$ , è il *rango di colonna* di  $X_{(n)}$ . In altre parole il *rango  $n$ -esimo* è la dimensione dello spazio vettoriale attraversato dalle fibre di moda  $n$ .

Se  $R_n = rank_n(\mathcal{X})$  per ogni  $n = 1, \dots, N$  allora si dice che  $\mathcal{X}$  è un tensore di rango  $(R_1 \times R_2 \times \dots \times R_N)$ . Si noti però che il rango  $n$ -esimo non va confuso con il rango tensoriale definito nella sezione 2.1.6.

## 2.2 Scomposizione CANDECOMP/PARAFAC

Il concetto della scomposizione *CP* (*CANDECOMP/PARAFAC*) deriva dall'idea di analisi proporzionale parallela (Cattell (1944)) e dall'idea di forma poliadica di un tensore (Hitchcock (1927)). Questa scomposizione tensoriale è però diventata popolare nel 1970 come *CANDECOMP* (*canonical decomposition*) tramite Carroll & Chang (1970) e come *PARAFAC* (parallel factors) tramite Harshman et al. (1970). Il nome "CP", che viene utilizzato in questo elaborato, è stato introdotto da Kiers (2000).

La scomposizione *CP* fornisce un'approssimazione intuitiva del tensore  $\mathcal{X} \in \mathbb{R}^{I_1 \times \dots \times I_N}$  usando la definizione di rango del tensore (vedi la sezione 2.1.6) e troncando la sommatoria ad un valore  $r$ , approssimando quindi  $\mathcal{X}$  con un tensore  $\hat{\mathcal{X}}$  di rango  $r$ , dove  $r$  è scelto a priori

$$\hat{\mathcal{X}} = \sum_{l=1}^r \hat{a}_l^{(1)} \circ \hat{a}_l^{(2)} \circ \dots \circ \hat{a}_l^{(N)} .$$

Normalizzando gli  $\hat{a}_l$  con  $a_l = \frac{\hat{a}_l}{\|\hat{a}_l\|}$  e memorizzando l'informazione della loro norma in  $\lambda \in \mathbb{R}^r$  allora

$$\hat{\mathcal{X}} = \sum_{l=1}^r \lambda_l a_l^{(1)} \circ a_l^{(2)} \circ \dots \circ a_l^{(N)} .$$

Indicando poi con  $A_n = \begin{bmatrix} a_1^{(n)} & a_2^{(n)} & \dots & a_r^{(n)} \end{bmatrix} \in \mathbb{R}^{I_n \times r}$  le matrici fattoriali per  $n = 1 \dots N$  e con  $\Lambda = \text{diag}(\lambda)$ , allora la versione con *matricizzazione* sulla moda  $n$  risulta essere

$$\hat{X}_{(n)} = A_n \Lambda (A_N \circ \dots \circ A_{n+1} \circ A_{n-1} \circ \dots \circ A_1)^T .$$

La Figura 2.6 mostra la rappresentazione di un tensore di terzo ordine prima come somma di tensori di rango uno e poi con la rappresentazione tramite matrici fattoriali.

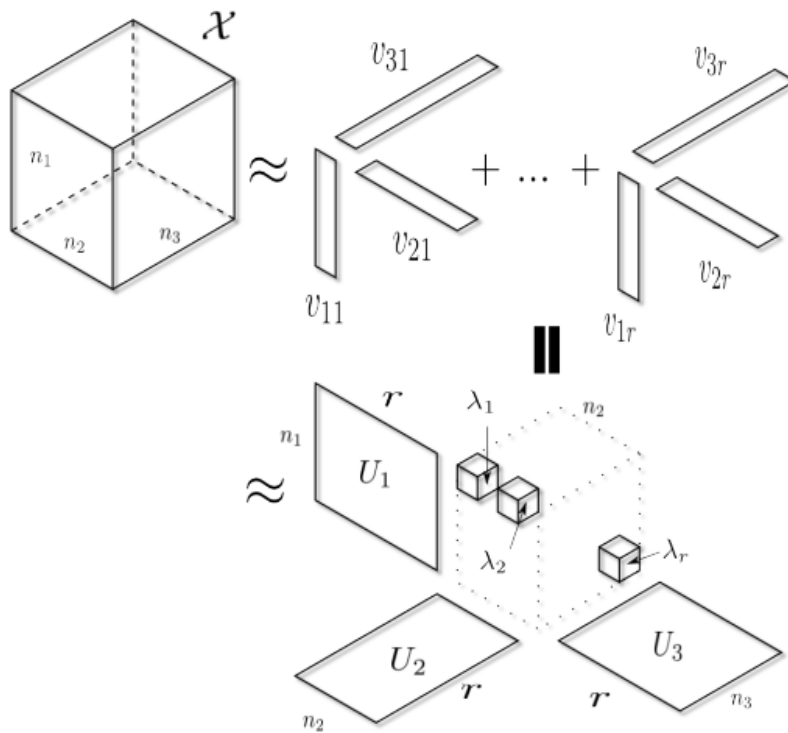


FIGURA 2.6: Scomposizione *CP* per un tensore  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ . Figura presa da Li et al. (2018).

Non essendoci un algoritmo per determinare la scelta ottimale del rango  $r$  relativo alla miglior approssimazione a rango ridotto, in Bro & Kiers (2003) è stata proposta una diagnostica consistente chiamata *CONCORDIA* (“*core consistency diagnostic*”) per

comparare diverse scelte del rango ridotto. L'obiettivo è quello di minimizzare la norma di Frobenius (vedi (2.4)) delle differenze tra  $\mathcal{X}$  e  $\hat{\mathcal{X}}$ . Formalizzando, si cerca

$$\min_{\hat{\mathcal{X}}} \|\mathcal{X} - \hat{\mathcal{X}}\|_F . \quad (2.7)$$

Si calcola quindi  $\hat{\mathcal{X}}$  per diversi valori di  $r$  e li si confronta tramite la norma di Frobenius sulla differenza con il tensore originale. Nello stesso articolo si suggerisce come regola generale che  $\hat{\mathcal{X}}$  è un buon approssimatore di  $\mathcal{X}$  se

$$100 \cdot \left(1 - \frac{\|\mathcal{X} - \hat{\mathcal{X}}\|_F^2}{\|\mathcal{X}\|_F^2}\right) \approx 90\% . \quad (2.8)$$

Questa regola impone che la norma di Frobenius del tensore approssimato sia circa il 90% della norma di Frobenius del tensore originale, perciò ha un legame con l'analisi delle componenti principali classica in cui si cerca quel numero di componenti principali con cui si spiega la maggior parte della variabilità dei dati e spesso tale scelta si basa sul numero di componenti che spiega una percentuale superiore ad una certa soglia di interesse (e.g. 90%).

Fissato il rango  $r$  esistono diversi algoritmi per calcolare la scomposizione  $CP$ , in questo lavoro viene utilizzato l'algoritmo  $ALS$  (*alternating least squares*) proposto nei due articoli originali (Harshman et al. (1970) e Carroll & Chang (1970)). L'approccio  $ALS$  consiste nel fissare a turno tutte le matrici fattoriali tranne una (riducendo così il problema ad uno di minimi quadrati), per poi iterare la stima delle matrici finché un certo criterio di convergenza non viene soddisfatto. Ad esempio riprendendo il problema di minimo in equazione (2.7) e fissando oltre al rango  $r$  tutte le matrici fattoriali tranne la prima avremo

$$\min_{\hat{A}_1} \|X_{(1)} - \hat{A}_1 (A_N \odot A_{N-1} \odot \cdots \odot A_2)^T\|_F , \quad \text{dove } \hat{A}_1 = A_1 \Lambda .$$

La cui soluzione è data da

$$\hat{A}_1 = X_{(1)} [(A_N \odot A_{N-1} \odot \cdots \odot A_2)^T]^\dagger .$$

Usando le proprietà (2.1), (2.2), (2.3) si riscrive la soluzione in modo da semplificare la pseudo-inversa:

$$\hat{A}_1 = X_{(1)} (A_N \odot A_{N-1} \odot \cdots \odot A_2) (A_N^T A_N \star A_{N-1}^T A_{N-1} \star \cdots \star A_2^T A_2)^\dagger .$$

Infine si normalizzano le colonne di  $\hat{A}_1$  per ottenere  $A_1$  con  $a_l = \frac{\hat{a}_l}{\lambda_l}$ , dove  $\lambda_l = \|\hat{a}_l\|$ .

Si noti che *ALS* non garantisce la convergenza ad un minimo globale, per tale motivo normalmente si indica un numero massimo di iterazioni per il quale se l'algoritmo non trova un minimo si arresta lo stesso. Lo *pseudo-codice* della scomposizione *CP* con metodo *ALS* è riportato nell'algoritmo 1 dell'Appendice.

## 2.3 Scomposizione di Tucker

La scomposizione di Tucker (Tucker (1966)) è una forma di *PCA* per tensori di ordine superiore al secondo in quanto scompone un tensore in un tensore nucleo moltiplicato lungo ogni moda per una matrice diversa. Nel caso di un tensore ad  $N$  vie  $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$  la scomposizione di Tucker è definita nel seguente modo

$$\mathcal{X} = \mathcal{G} \times_1 A^{(1)} \times_2 A^{(2)} \times_3 \dots \times_N A^{(N)} = [[\mathcal{G}; A^{(1)}, \dots, A^{(N)}]]. \quad (2.9)$$

Qui  $A^{(1)} \in \mathbb{R}^{I_1 \times R_1}$ ,  $A^{(2)} \in \mathbb{R}^{I_2 \times R_2}$ ,  $\dots$ ,  $A^{(N)} \in \mathbb{R}^{I_N \times R_N}$  sono le matrici fattoriali e possono essere pensate come componenti principali della relativa moda. Il tensore  $\mathcal{G} \in \mathbb{R}^{R_1 \times R_2 \times \dots \times R_N}$  è chiamato tensore nucleo e ogni suo elemento mostra il livello di interazione tra le differenti componenti. Se  $R_1, \dots, R_N$  sono più piccoli degli  $I_1, \dots, I_N$ , il tensore  $\mathcal{G}$  può essere pensato come una versione compressa del tensore  $\mathcal{X}$ . Di fatto la scomposizione *CP* può essere vista come un caso speciale di scomposizione di Tucker dove il tensore nucleo è superdiagonale e  $R_1 = R_2 = \dots = R_N$ . La Figura 2.7 mostra graficamente l'idea della scomposizione di Tucker per un tensore  $\mathcal{X}$  a 3 vie.

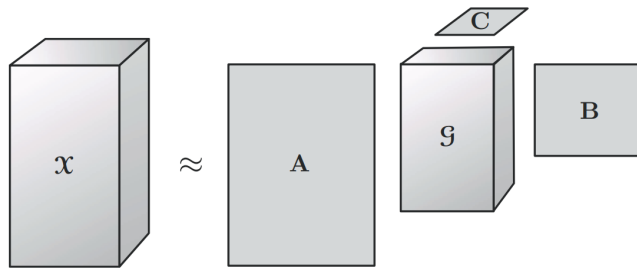


FIGURA 2.7: Scomposizione di Tucker di un tensore a 3 vie,  $\mathcal{X} = \mathcal{G} \times_1 A \times_2 B \times_3 C$ . Figura presa da Kolda & Bader (2009).

Per un dato tensore  $\mathcal{X}$  è facile trovare un'esatta scomposizione di Tucker di rango  $(R_1, R_2, \dots, R_N)$  se  $R_n = \text{rank}_n(\mathcal{X})$  per ogni  $n = 1, \dots, N$  (si veda la sezione 2.1.8 per la definizione di  $\text{rank}_n$ ), risulta invece più difficile e non necessariamente esatto il calcolo di una scomposizione di Tucker di rango  $(R_1, R_2, \dots, R_N)$  se per almeno un  $n$   $R_n < \text{rank}_n(\mathcal{X})$  (detta anche scomposizione di Tucker troncata). La Figura 2.8 è una rappresentazione grafica della scomposizione di Tucker troncata per un tensore di terzo

ordine, infatti le parti non colorate mostrano che la scomposizione non riproduce esattamente il tensore originale. Questa approssimazione tramite scomposizione di Tucker

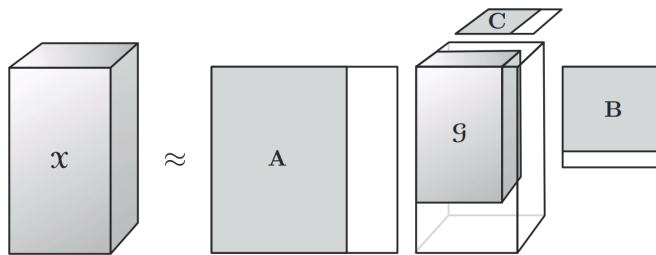


FIGURA 2.8: Scomposizione di Tucker troncata di un tensore a 3 vie.  $\mathcal{X} \approx \mathcal{G} \times_1 A \times_2 B \times_3 C$ . Figura presa da Kolda & Bader (2009).

è simile a quella che fa la scomposizione  $CP$  solo che rilassa il vincolo per la quale il tensore originario debba essere esprimibile come somma di tensori di rango 1.

Sono stati proposti diversi metodi per calcolare la scomposizione di Tucker, in questo lavoro ci si concentra però solo su due. Il primo (il cui *pseudo-codice* è riportato nell'algoritmo 2 dell'Appendice) introdotto da Tucker (1966) e conosciuto come “*high order SVD*” (HOSVD) si basa sull'idea di cercare quelle componenti che colgono meglio la variabilità di ogni moda  $n$  indipendentemente dalle altre mode.

*HOSVD* impone dei vincoli sul tensore nucleo  $\mathcal{G}$ :

- $\mathcal{G}$  deve essere *tutto-ortogonale*, nel senso che ogni sotto tensore a  $N - 1$  vie lungo una moda  $n$  ( $\forall n = 1, \dots, N$ ) deve essere ortogonale rispetto agli altri (il prodotto interno tra i diversi sotto tensori deve essere pari a 0)

$$\langle \mathcal{G}_{i_n=\alpha}, \mathcal{G}_{i_n=\beta} \rangle = 0 \quad \forall \alpha \neq \beta,$$

- $\mathcal{G}$  deve avere i sotto-tensori lungo ogni via ordinati rispetto la norma di Frobenius

$$\|\mathcal{G}_{i_n=1}\|_F \geq \|\mathcal{G}_{i_n=2}\|_F \geq \dots \geq \|\mathcal{G}_{i_n=R_n}\|_F.$$

Per via di questi vincoli, in particolare della *tutto-ortogonalità* imposta al tensore nucleo, la *HOSVD* troncata non è un approssimazione ottima in termini di norma di Frobenius sulla differenza tra tensore originale e tensore approssimato. Tuttavia, *HOSVD* è un buon punto di partenza per un algoritmo *ALS* iterativo (secondo metodo di calcolo della scomposizione di Tucker che si vede in questo lavoro) chiamato “*higher order orthogonal iteration*” (HOOI), introdotto solo per tensori a 3-vie da Kroonenberg & De Leeuw (1980), esteso per tensori a  $N$ -vie da Kapteyn et al. (1986) e reso più efficiente da De Lathauwer et al. (2000).

Sia  $\mathcal{X}$  un tensore di dimensioni  $I_1 \times I_2 \times \cdots \times I_N$ ; il problema di ottimizzazione che *HOOI* vuole risolvere è

$$\min_{\mathcal{G}, A^{(1)}, \dots, A^{(N)}} \|\mathcal{X} - (\mathcal{G} \times_1 A^{(1)} \times_2 A^{(2)} \times_3 \cdots \times_N A^{(N)})\|_F. \quad (2.10)$$

Sotto i vincoli che  $\mathcal{G} \in \mathbb{R}^{R_1 \times R_2 \times \cdots \times R_N}$ ,  $A^{(n)} \in \mathbb{R}^{I_n \times R_n}$  e che  $A^{(n)}$  sia a colonne ortogonali per  $n = 1, \dots, N$ , è dimostrabile che il tensore nucleo  $\mathcal{G}$  debba soddisfare

$$\mathcal{G} = \mathcal{X} \times_1 A^{(1)T} \times_2 A^{(2)T} \times_3 \cdots \times_N A^{(N)T}.$$

Si può riscrivere il quadrato della funzione obiettivo 2.10 come

$$\begin{aligned} \|\mathcal{X} - (\mathcal{G} \times_1 A^{(1)} \times_2 A^{(2)} \times_3 \cdots \times_N A^{(N)})\|_F^2 &= \\ &= \|\mathcal{X}\|_F^2 - \|(\mathcal{X} \times_1 A^{(1)T} \times_2 A^{(2)T} \times_3 \cdots \times_N A^{(N)T})\|_F^2. \end{aligned}$$

Per la dimostrazione e i dettagli di questa trasformazione si rimanda a De Lathauwer et al. (2000). Si può quindi usare un approccio *ALS* per risolvere il problema di minimo in (2.10), in quanto essendo  $\|\mathcal{X}\|_F^2$  una costante la minimizzazione in (2.10) può essere rivista come una serie di sotto problemi che coinvolgono la massimizzazione di ogni *n-esima* componente

$$\max_{A^{(n)}} \|\mathcal{X} \times_1 A^{(1)T} \times_2 A^{(2)T} \times_3 \cdots \times_N A^{(N)T}\|_F^2, \quad (2.11)$$

con  $A^{(n)} \in \mathbb{R}^{I_n \times R_n}$  a colonne ortogonali. La funzione obiettivo in (2.11) può essere riscritta in forma matriciale come  $\|A^{(n)T} W\|_2$  dove  $W = X_{(n)}(A^{(N)} \otimes \cdots \otimes A^{(n+1)} \otimes A^{(n-1)} \otimes \cdots \otimes A^{(1)})$  perciò la soluzione per ogni  $A^{(n)}$  sono i primi  $R_n$  vettori singolari sinistri della decomposizione a valori singolari (SVD) di  $W$ .

Questo metodo (il cui *pseudo-codice* è riportato nell'algoritmo 3 dell'Appendice) converge ad una soluzione ma non garantisce la convergenza ad un minimo globale dell'equazione (2.10).

In Timmerman & Kiers (2000) viene proposta una procedura chiamata *DIFFIT* per la scelta del rango in una scomposizione di Tucker troncata:

1. Si svolge la scomposizione di Tucker per diverse combinazioni di rango escludendo i casi in cui la dimensione di una moda è maggiore del prodotto delle dimensioni delle altre mode ( $R_n > R_1 \cdots R_{n-1} \cdot R_{n+1} \cdots R_N$ ) perché danno gli stessi risultati di quando  $R_n = R_1 \cdots R_{n-1} \cdot R_{n+1} \cdots R_N$ .

2. Si comparano le soluzioni con lo stesso numero totale di componenti ( $s = R_1 + \dots + R_N$ ) tenendo solo la miglior soluzione per ogni  $s$  sulla base della percentuale di adattamento calcolata come

$$100 \cdot \frac{\|(\mathcal{G} \times_1 A^{(1)} \times_2 A^{(2)} \times_3 \dots \times_N A^{(N)})\|_F^2}{\|\mathcal{X}\|_F^2}.$$

3. Si calcola  $diff_s$  come differenza di percentuale di adattamento tra la miglior soluzione con  $s$  componenti e la miglior soluzione con  $s - 1$  componenti. Si tengono solo i  $diff_s$  tali che  $diff_s > diff_{s+1}$ . Queste soluzioni sono indicate con  $m = 1, \dots, M$  e il numero di componenti  $s$  associato alla  $m$ -esima soluzione è indicato con  $t(m)$ , quindi il suo  $diff$  sarà  $diff_{t(m)}$ .
4. Si calcola  $b_{t(m)} = diff_{t(m)} / diff_{t(m+1)}$  che indica il tasso di miglioramento nel passare da  $t(m)$  componenti a  $t(m + 1)$ .
5. Si mantengono solo quei numeri di componenti totali ( $t(m)$ ) tali che  $diff_{t(m)} > \frac{\|\mathcal{X}\|_F^2}{s_{max} - N}$  così da non considerare le soluzioni con un miglioramento troppo piccolo.
6. Si sceglie la soluzione con il più alto valore di  $b_{t(m)}$  tra le soluzioni rimaste.

## 2.4 Algoritmo EM per clustering tensoriale

Le scomposizioni tensoriali illustrate possono essere utili anche come base di partenza per modelli mistura tensoriali e relativi metodi di *clustering* tensoriale *model based*. In Mai et al. (2022) infatti si introduce il “*doubly enhanced EM algorithm*” (DEEM) per *clustering* tensoriale *model based* che è un algoritmo basato sul *tensor normal mixture model* (TNMM). Il TNMM esplora la struttura di covarianza tensoriale e grazie alla semplicità delle distribuzioni normali tensoriali (TN) il relativo *clustering* è basato sul algoritmo di *expectation-maximization* (EM). Nel DEEM però a differenza del EM standard si impone sparsità nella regola ottimale di *clustering* che viene massimizzata tramite un particolare stimatore di correlazione tensoriale non iterativo. La sparsità viene utilizzata come alternativa all’assunzione di rango ridotto del tensore di coefficienti permettendo l’utilizzo anche in caso di analisi di dati tensoriali ad alta dimensionalità. Lo stimatore non iterativo per la struttura di correlazione tensoriale permette invece un guadagno in termini computazionali e di proprietà teoriche.

Prima di addentrarsi nel vero e proprio algoritmo è utile definire la distribuzione normale tensoriale (TN) e il modello mistura di distribuzioni normali tensoriali (TNMM).

Sia  $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$  un tensore casuale tale che

$$\mathcal{X} = \mu + (\mathcal{Z} \times_1 \Sigma_1^{1/2} \times_2 \Sigma_2^{1/2} \times_3 \dots \times_N \Sigma_N^{1/2}),$$

con  $\mu \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ ,  $\Sigma_n \in \mathbb{R}^{I_n \times I_n} \forall n = 1, \dots, N$  e  $\mathcal{Z} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$  tale che ogni suo elemento sia indipendente dagli altri e si distribuisca come una normale di media nulla e varianza unitaria ( $z_{i_1, \dots, i_N} \sim N(0, 1)$ ), allora si dice che  $\mathcal{X}$  si distribuisce come una normale tensoriale

$$\mathcal{X} \sim TN(\mu; \Sigma_1, \Sigma_2, \dots, \Sigma_N).$$

Se si considerano invece  $m$  tensori  $\mathcal{X}_i \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$  (con  $i = 1, \dots, m$ ) provenienti da  $K$  cluster e con distribuzione normale tensoriale comune all'interno di ogni cluster, allora per individuare il gruppo di provenienza di ogni tensore si assume che  $\mathcal{X}_i$  segua un modello mistura di  $K$  distribuzioni tensoriali normali (TNMM)

$$\mathcal{X}_i \sim \sum_{k=1}^K \pi_k \cdot TN(\mu_k; \Sigma_1, \Sigma_2, \dots, \Sigma_N), \quad (2.12)$$

con  $\mu_k \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$  tensore medio del cluster  $k$ ,  $\Sigma_n \in \mathbb{R}^{I_n \times I_n}$  covarianza comune entro-cluster lungo la moda  $n$  e  $0 < \pi_k < 1$  probabilità a priori che il tensore provenga dal  $k$ -esimo cluster con  $\sum_{k=1}^K \pi_k = 1$ . Per assicurare l'identificabilità delle matrici di covarianza si assume che l'elemento (1,1) di  $\Sigma_n$  sia pari a 1 se  $n > 1$  ( $\sigma_{n,1,1} = 1 \forall n > 1$ ) e che l'elemento (1,1) di  $\Sigma_1$  sia pari alla varianza entro cluster del elemento  $(1, \dots, 1)$  dei tensori  $\mathcal{X}_i$ .

Indicando con  $C_i \in \{1, \dots, K\}$  il cluster di appartenenza dell'osservazione  $i$ -esima e definendo  $f_k$  come la funzione di densità di probabilità condizionata del tensore dato il suo cluster di appartenenza ( $\Pr(\mathcal{X}|C = k)$ ) che quindi si assume essere una normale tensoriale, la regola di clustering ottima sarebbe

$$\hat{C}_i^{opt} = \arg \max_k \pi_{i,k}^* f_k(\mathcal{X}_i), \quad (2.13)$$

dove  $\pi_{i,k}^*$  è la probabilità marginale che  $\mathcal{X}_i$  appartenga al  $k$ -esimo cluster.

Se  $\theta = \{\pi_1^*, \dots, \pi_K^*, \mu_1, \dots, \mu_K, \Sigma_1, \dots, \Sigma_N\}$  è il vettore di parametri del modello mistura, allora se si osservassero le variabili  $C_i$  basterebbe massimizzare la funzione di log-verosimiglianza per ottenere lo stimatore  $\theta$ . Siccome però le variabili  $C_i$  sono latenti non si riesce a massimizzare direttamente la funzione di log-verosimiglianza per ottenere lo stimatore di  $\theta$ .



Il *DEEM* assume che la probabilità di appartenenza ad un cluster per l'*i-esima* osservazione dato il suo tensore  $\mathcal{X}_i$  e i parametri  $\theta$  ( $\xi_{ik} = Pr(C_i = k | \mathcal{X}_i, \theta)$ ) sia determinata da un sottoinsieme di elementi in  $\mathcal{X}_i$  (quindi assume che può essere valutata con un numero ridotto di parametri), in quanto se non ci sono differenze tra i tensori in una determinata posizione allora questa risulterà inutile ai fini della determinazione del cluster di appartenenza  $C_i$ .

Sia  $B_k$  per  $k = 2, \dots, K$  la differenza tra il *k-esimo* e il primo tensore medio, moltiplicata lungo ogni moda  $n$  per  $(\Sigma_n)^{-1}$  cioè

$$B_k = [[\mu_k - \mu_1; (\Sigma_1)^{-1}, \dots, (\Sigma_N)^{-1}] \in \mathbb{R}^{I_1 \times \dots \times I_N}, k = 2, \dots, K,$$

allora per ottenere le stime sparse dei  $B_k$ , e quindi per individuare le posizioni utili a discriminare i cluster, si inseriscono nella stima dei  $B_k$  le stime correnti per  $\mu_k$  e  $\Sigma_n$  aggiungendo la penalità *group lasso* (che consiste nel mettere una penalità L2 su gruppi di parametri e per i cui dettagli si può fare riferimento a Yuan & Lin (2006)) sulle stesse posizioni dei diversi  $B_k$ .

Il parametro di regolazione relativo alla penalità *group lasso* varia da iterazione ad iterazione e viene definito con  $\lambda^{(t+1)} > 0$ . Il problema di minimo per il calcolo delle stime sparse dei  $B_k$  alla  $(t+1)$ -esima iterazione (indicate con  $\hat{B}_2^{(t+1)}, \dots, \hat{B}_K^{(t+1)}$ ) può essere risolto facilmente con un algoritmo di discesa delle coordinate a blocchi (*blockwise coordinate descent algorithm*), per ulteriori dettagli si rimanda a Mai et al. (2022).

Dopo aver ottenuto  $\hat{B}_2^{(t+1)}, \dots, \hat{B}_K^{(t+1)}$  se si definisce  $P_k$  come una misura di distanza tra il tensore e un aggiornamento del tensore medio per il gruppo  $k$

$$P_k = \exp\left\{\left\langle \mathcal{X}_i - \frac{1}{2}(\hat{\mu}_k^{(t)} + \hat{\mu}_1^{(t)}), \hat{B}_k^{(t+1)} \right\rangle\right\},$$

allora la probabilità condizionata di appartenenza al cluster  $k$  per l'osservazione  $i$  ( $\hat{\xi}_{ik}^{(t+1)}$ ) sarà pari alla probabilità marginale di appartenere al *k-esimo* cluster calcolata alla *t-esima* iterazione ( $\hat{\pi}_k^{(t)}$ ) normalizzata con pesi dati dai  $P_k$  in modo che le  $k$  probabilità dell'osservazione  $i$  sommino ad 1

$$\hat{\xi}_{i1}^{(t+1)} = \frac{\hat{\pi}_1^{(t)}}{\hat{\pi}_1^{(t)} + \sum_{j=2}^K \hat{\pi}_j^{(t)} \cdot P_j}, \quad \hat{\xi}_{ik}^{(t+1)} = \frac{\hat{\pi}_k^{(t)} \cdot P_k}{\hat{\pi}_1^{(t)} + \sum_{j=2}^K \hat{\pi}_j^{(t)} \cdot P_j}.$$

Risulta importante notare che siccome  $B_1 = 0$  allora  $P_1 = 0$  e che per questo motivo la probabilità marginale di appartenere al primo cluster ( $\hat{\pi}_1^{(t)}$ ) non viene moltiplicata per la misura di distanza. Il primo *cluster* di fatto viene tenuto come riferimento per l'aggiornamento delle misure di distanza degli altri cluster, considerando quindi la sua

misura di distanza pari a 1.

Nel  $(t+1)$ -esimo  $E$ -step si calcolerà quindi il valore atteso della log-verosimiglianza di  $\theta$  dati gli  $m$  tensori  $\mathcal{X}_i$  e la stima al passo precedente  $\hat{\theta}^{(t)}$

$$Q^{DEEM}(\theta|\hat{\theta}^{(t)}) = E[l(\theta)|\{\mathcal{X}_i\}_{i=1}^m, \hat{\theta}^{(t)}]. \quad (2.14)$$

Nello  $step$   $M$  invece si aggiornano le stime per  $\pi_k, \mu_k, \Sigma_n$  tramite la massimizzazione della funzione (2.14). Risulta facile ottenere le stime per  $\pi_k$  e  $\mu_k$  alla  $(t+1)$ -esima iterazione, mentre gli aggiornamenti di  $\Sigma_1, \dots, \Sigma_N$  risultano molto più complessi da ottenere, in quanto  $\hat{\Sigma}_n^{(t+1)}$  dipende da tutte le altre stime di covarianza  $\hat{\Sigma}_{n'}^{(t+1)}$  con  $n' \neq n$ . Perciò, sarebbe necessario un algoritmo iterativo secondario che aumenterebbe notevolmente il costo computazionale. Inoltre, questa procedura in caso di alta dimensionalità non permetterebbe facilmente di stabilire la consistenza della regola di *clustering*.

Siccome sotto il  $TNMM$  è costruibile uno stimatore con soluzione esplicita tramite il metodo dei momenti, in Mai et al. (2022) si propone uno stimatore non iterativo per i parametri delle matrici di covarianza che consiste dei seguenti passi:

1. Dati gli  $\hat{\xi}_{ik}^{(t+1)}$  si calcolano le stime intermedie  $\check{\Sigma}_n^{(t+1)}$  tramite il metodo dei momenti.
2. Si calcola lo stimatore  $DEEM$  date le stime intermedie
  - per  $n > 1$ :  $\hat{\Sigma}_n^{(t+1)} = \check{\Sigma}_n^{(t+1)} / \check{\sigma}_{n,1,1}^{(t+1)}$ ,
  - per  $n = 1$ :  $\hat{\Sigma}_1^{(t+1)} = \check{\Sigma}_1^{(t+1)} \cdot (\hat{\sigma}_{1;1,1}^{(t+1)} / \check{\sigma}_{1;1,1}^{(t+1)})$ ,
dove  $\hat{\sigma}_{1;1,1}^{(t+1)}$  è il valore atteso della stima della varianza entro-cluster e  $\check{\sigma}_{n,1,1}^{(t+1)}$  è l'elemento  $(1, 1)$  di  $\check{\Sigma}_n^{(t+1)}$  per  $n = 1, \dots, N$ .

Questo stimatore ha soluzione esplicita (non serve quindi l'utilizzo di un algoritmo iterativo secondario) con un guadagno computazionale notevole, inoltre è dimostrabile che queste stime portino ad un *clustering* consistente anche quando la dimensione di ciascuna moda del tensore cresce esponenzialmente rispetto alla numerosità campionaria ( $m$ ). Per la dimostrazione della consistenza della regola di *clustering* e ulteriori dettagli fare riferimento a Mai et al. (2022).

L'algoritmo del  $DEEM$  (il cui *pseudo-codice* è riportato nell'algoritmo 4 dell'Appendice) restituisce come risultato le stime di  $\hat{\xi}_{ik}, \hat{\mu}_k, \hat{\pi}_k, \hat{\Sigma}_n$ , da cui la regola di *clustering* assegnerà ogni  $\mathcal{X}_i$  al cluster  $\hat{C}_i^{DEEM} = \arg \max_k \hat{\xi}_{ik}$ . Nell'algoritmo vanno inizializzati  $\hat{\pi}_k^{(0)}, \hat{\mu}_k^{(0)}, \hat{\Sigma}_n^{(0)}$ , in Mai et al. (2022) si suggerisce di inizializzarli tramite un  $K$ -means *clustering* sulla *vettorizzazione* dei tensori  $\mathcal{X}_i$  che permette di trovare i cluster iniziali ( $\hat{C}_i^{(0)}$ ). Tramite i  $\hat{C}_i^{(0)}$  si ottengono facilmente i  $\hat{\pi}_k^{(0)}$  come proporzione di osservazioni

che l'algoritmo K-medie assegna al  $k$ -esimo cluster,  $\hat{\mu}_k^{(0)}$  come media dei tensori appartenenti al  $k$ -esimo cluster secondo  $\hat{C}_i^{(0)}$  e  $\hat{\Sigma}_n^{(0)}$  come matrice di varianza e covarianza entro-cluster del  $n$ -esima moda

$$\begin{aligned}\hat{\pi}_k^{(0)} &= \frac{\sum_{i=1}^m I(\hat{C}_i^{(0)} = k)}{m}, \\ \hat{\mu}_k^{(0)} &= \frac{1}{m\hat{\pi}_k^{(0)}} \sum_{\hat{C}_i^{(0)}=k} \mathcal{X}_i, \\ \hat{\Sigma}_n^{(0)} &\propto \frac{1}{mq_n} \sum_{i=1}^m \sum_{\hat{C}_i^{(0)}=k} \left[ (\mathcal{X}_i - \hat{\mu}_k^{(0)})_{(n)} (\mathcal{X}_i - \hat{\mu}_k^{(0)})_{(n)}^T \right].\end{aligned}$$

Per l'identificabilità delle matrici di covarianza  $\sigma_{n;1,1}^{(0)} = 1 \forall n > 1$  e  $\sigma_{1;1,1}^{(0)}$  sarà la varianza entro-cluster del primo elemento del tensore.

Il criterio di convergenza dell'algoritmo è basato sulla somma dei quadrati della media delle differenze tra due iterazioni successive ( $\sum_k \|\hat{\mu}_k^{(t+1)} - \hat{\mu}_k^{(t)}\|_F^2 \leq 0.1$ ). In Mai et al. (2022) si mostra come numerosi studi hanno evidenziato che quando  $t$  è grande il parametro di regolazione  $\lambda^{(t)}$  è approssimativamente costante, e che il problema della selezione consistente di  $K$  per tensori ad alta dimensionalità è una questione ancora aperta in letteratura. Per questo motivo in questo lavoro si considerano diversi valori di  $K$  valutandone i risultati e regolando  $\lambda^{(t)} = \lambda$  tramite un criterio simile al criterio di informazione di Bayes (BIC). Per ulteriori dettagli sulla scelta dei parametri di regolazione fare riferimento a Mai et al. (2022).

## 2.5 Clustering multi-direzionale tramite il modello tensoriale a blocchi

Il *Clustering* tramite *Tensor Block Model* (TBM) introdotto in Wang & Zeng (2019) è una procedura unificata di stima ai minimi quadrati per identificare una struttura a gruppi multi-direzionali di un tensore. A differenza di molti metodi di *clustering* multi-direzionali che si basano su una procedura a due passi dove prima si cerca una rappresentazione a rango ridotto del tensore e poi si applicano algoritmi di *clustering*, il TBM usa un unico passo portando ad una miglior accuratezza ed interpretabilità.

Il TBM assume che il tensore osservato  $\mathcal{X} \in \mathbb{R}^{I_1 \times \dots \times I_N}$  sia una realizzazione rumorosa di un tensore sottostante che ha una struttura a blocchi, dove ogni moda  $n$  ( $n = 1, \dots, N$ ) consiste di  $R_n$  cluster. Quindi, se  $x_{i_1, \dots, i_N}$  deriva dal blocco determinato dal  $(r_n)$ -esimo cluster della moda  $n$  con  $r_n \in \{1, \dots, R_n\}$  allora si assume che

$$x_{i_1, \dots, i_N} = c_{r_1, \dots, r_N} + \epsilon_{i_1, \dots, i_N},$$

dove  $c_{r_1, \dots, r_N}$  è la media del blocco tensoriale indicizzato da  $(r_1, \dots, r_N)$  e gli  $\epsilon_{i_1, \dots, i_N}$  sono termini di errore di media 0 e indipendenti.

Il *clustering* tramite *TBM* quindi raggruppa lungo ogni moda e trova le medie di gruppo  $c_{r_1, \dots, r_N}$ . Perciò, si può riscrivere il *TBM* come una particolare scomposizione di Tucker (Sezione 2.3) del tensore  $\mathcal{X}$  con l'aggiunta di un termine di errore

$$\mathcal{X} = \mathcal{C} \times_1 M_1 \times_2 \cdots \times_N M_N + \mathcal{E},$$

dove  $\mathcal{C} \in \mathbb{R}^{R_1 \times \cdots \times R_N}$  è il tensore nucleo che corrisponde ad un particolare tensore  $\mathcal{G}$  della scomposizione di Tucker (2.9) contenente le medie dei gruppi  $(c_{r_1, \dots, r_N})$ .  $M_n \in \{0, 1\}^{I_n \times R_n}$  è la matrice fattoriale corrispondente ad una particolare matrice  $A^{(n)}$  di equazione (2.9) che prende solo  $\{0, 1\}$  come valori e indica quindi l'allocazione nei gruppi lungo la moda  $n$ . Infine,  $\mathcal{E}$  è il tensore di errore contenente gli  $\epsilon_{i_1, \dots, i_N}$  indipendenti a cui si impone un'ulteriore assunzione detta di  *$\sigma$ -sub-gaussianità* (per i cui dettagli si rimanda a Wang & Zeng (2019)).

Sia  $\Theta \in \mathbb{R}^{I_1 \times \cdots \times I_N}$  il tensore del segnale medio con una struttura a blocchi

$$\Theta = \mathcal{C} \times_1 M_1 \times_2 \cdots \times_N M_N,$$

allora se si conoscessero le dimensioni dei *cluster*  $R = (R_1, \dots, R_N)$  lo stimatore dei minimi quadrati del *TBM* sarebbe

$$\begin{aligned} \hat{\Theta} &= \arg \min_{\Theta} \{ \|\mathcal{X} - \Theta\|_F^2 \} = & (2.15) \\ &= \arg \min_{\Theta} \{ \|\mathcal{X}\|_F^2 - 2\langle \mathcal{X}, \Theta \rangle + \|\Theta\|_F^2 \} = \\ &= \arg \min_{\Theta} \{ -2\langle \mathcal{X}, \Theta \rangle + \|\Theta\|_F^2 \}. \end{aligned}$$

Visto che l'equazione (2.15) richiede che  $R = R_1, \dots, R_N$  (ovvero il numero di *cluster* in ogni moda) sia fissato, lo si seleziona tramite il criterio di informazione di Bayes (BIC) che mette una penalità sul numero di parametri moltiplicato per una quantità che dipende dalle dimensioni delle mode del tensore (per ulteriori dettagli si faccia riferimento a Wang & Zeng (2019)).

Questo metodo di *clustering* può essere visto come un'estensione per tensori di ordine

superiore al secondo del classico algoritmo  $k$ -medie, dove il tensore nucleo  $\mathcal{C}$  ha il ruolo che hanno i centroidi nel  $k$ -means. La convergenza dell'algoritmo (il cui *pseudo-codice* è riportato nell'algoritmo 5 dell'Appendice) è garantita e il costo computazionale cresce linearmente con la produttoria delle dimensioni delle mode del tensore ( $\prod_n I_n$ ).

Per tenere conto del compromesso varianza-distorsione (fare riferimento a Azzalini & Scarpa (2012) per approfondimenti) e migliorare l'interpretabilità si può estendere il TBM ad una versione sparsa tramite metodi di regolarizzazione sulle medie dei blocchi tensoriali. Lo *sparse tensor block model* viene quindi stimato tramite minimi quadrati regolarizzati

$$\hat{\Theta}^{sparse} = \arg \min_{\Theta} \{ \|\mathcal{X} - \Theta\|_F^2 + \lambda \|\mathcal{C}\|_{\rho} \}.$$

In questo lavoro ci si concentra su  $\rho = 1$  cioè sulla penalità *Lasso*. Si seleziona anche il parametro di regolazione  $\lambda$  tramite il criterio di informazione *BIC* che mette una penalità sul numero di parametri diversi da 0 moltiplicati per una quantità che dipende dalle dimensioni del tensore.



# Capitolo 3

## Applicazione ai dati

### 3.1 Creazione dei tensori

Trattando i dati visti nella sezione 1.2 come 12 tabelle (una per paese) a 3 entrate (età, periodo e sesso) da un punto di vista matematico possono essere visti come 12 tensori a 3 vie/mode. La prima via rappresenta i 9 decenni, la seconda rappresenta le 111 età (da 0 a 110 anni, dove le morti da 110 anni in poi sono state accorpate nell'ultima età considerata), e la terza rappresenta il sesso. La scelta di utilizzare 12 tensori è dovuta al fatto che uno degli obiettivi dell'analisi è cogliere le differenze tra paesi sia in termini di adattamento delle approssimazioni del tensore tramite scomposizioni che in termini di gruppi individuati dai *clustering* tensoriali.

Sia  $\mathcal{X}_i$  il tensore relativo all'*i-esimo* paese allora  $\mathcal{X}_i \in [0, 1]^{9 \times 111 \times 2}$ , cioè per ogni fissato valore di ogni via si ha la proporzione di decessi per l'*i-esimo* paese nell'età fissata, per il sesso fissato e nel periodo fissato. Quindi, se si fissano la prima e la terza via si ottiene la forma *vettorizzata* di una curva della distribuzione dell'età alla morte.

### 3.2 Riduzione della complessità tramite scomposizione tensoriale CP

Il problema della riduzione della complessità del tensore può essere affrontato tramite la scomposizione tensoriale approssimando il tensore originale con un tensore a rango ridotto. Si è ridotto quindi il tensore di ogni paese tramite la scomposizione *CP* troncata. La valutazione della miglior approssimazione è stata fatta prendendo in considerazione diversi troncamenti a partire dal tensore di approssimazione troncato a rango 1 fino a quello troncato a rango 10. Le diverse approssimazioni (indicate con  $\hat{X}_r^{(CP)}$  per  $r =$

1, ..., 10) sono state valutate sia in base alla diagnostica CONCORDIA (introdotta nella sezione 2.2) che in base ad un'analisi grafica delle curve risultanti dal tensore approssimato, cercando di individuare cosa si coglie man mano che si aumenta il rango di approssimazione.

In Figura 3.1 si possono vedere gli *scree-plot* per ogni paese, la prima cosa che si nota è che tutte le curve non sembrano avere un vero e proprio punto di gomito ma che dal rango pari a 4 o 5 non sembra migliorare di molto l'approssimazione. Inoltre, si notano due gruppi di curve di cui uno composto da curve che inizialmente scendono meno velocemente delle altre. Questo gruppo risulta essere composto da Belgio, Finlandia, Francia, Italia, Spagna e Svizzera, ovvero i paesi del centro-sud Europeo (si veda Figura 3.2).

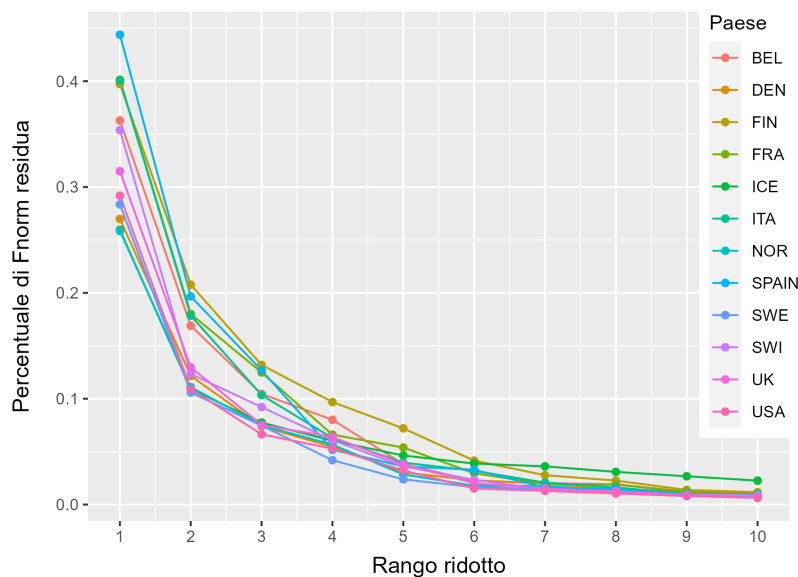


FIGURA 3.1: *Scree plots* delle scomposizioni CP troncate colorate per paese.

Analizzando le curve di questi 6 paesi (Figura 3.3 e Figura 3.4) si nota un picco sui 20 anni nelle curve dei maschi del V decennio di Finlandia Francia e Italia. Il Belgio e la Svizzera presentano un innalzamento delle curve relative al IV e V decennio del 1900 ma senza picchi particolari (entrambi i paesi hanno avuto una posizione neutrale durante la II guerra mondiale ma essendo al centro della zona di guerra hanno subito invasioni e bombardamenti). Infine, per la Spagna in Figura 3.4 si evidenzia come le curve relative ai decenni precedenti al 1970 siano molto irregolari (la Spagna è stata sotto la dittatura franchista fino al 1975 perciò risulta facile immaginare che la raccolta dati sotto tale dittatura non sia stata agevolata e che sia stata sottoposta a censura).



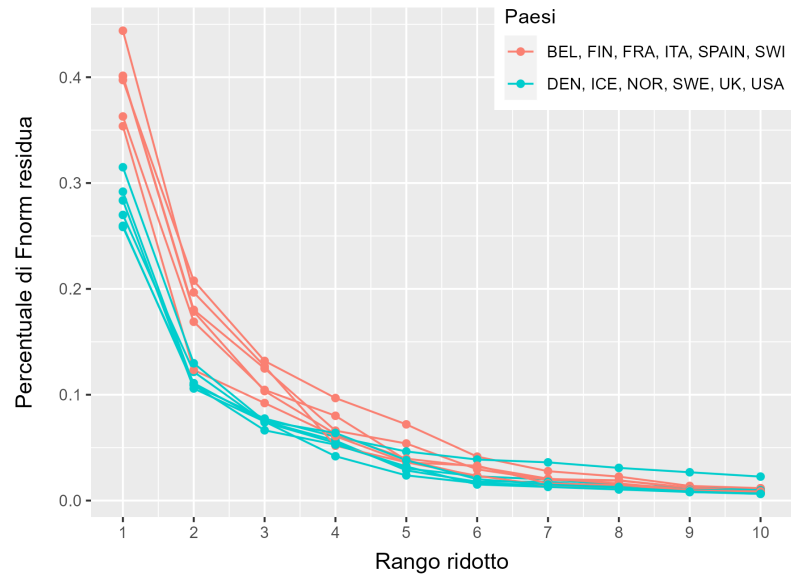


FIGURA 3.2: Scree plots delle scomposizioni CP troncate colorate in base ai due gruppi di paesi individuati.

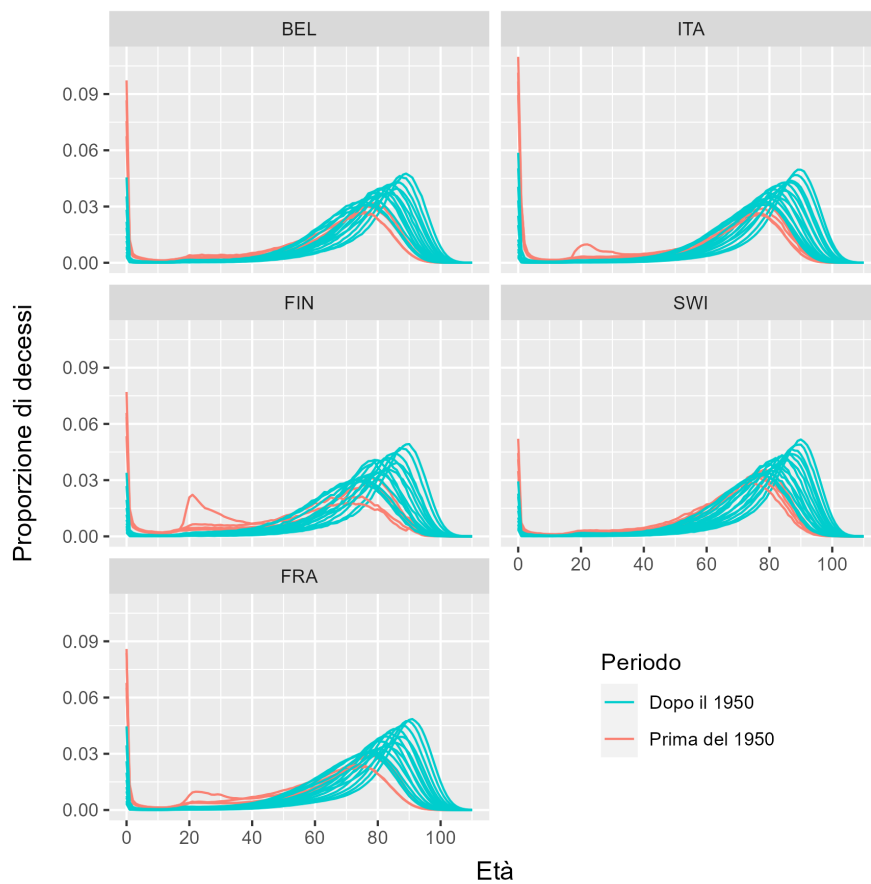


FIGURA 3.3: Distribuzioni dell'età alla morte originali per i 5 paesi che hanno curve particolari durante i primi due decenni, colorate in base al periodo.

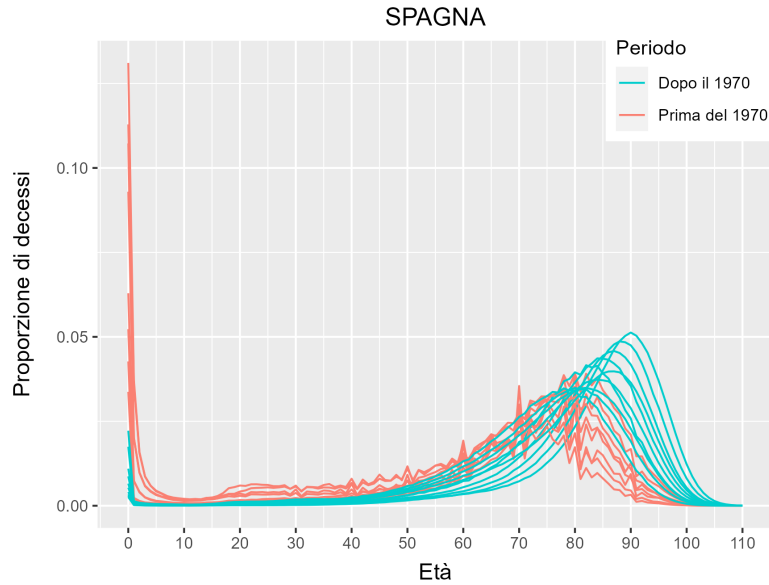


FIGURA 3.4: Distribuzioni dell'età alla morte originali per la Spagna colorate in base al periodo.

Tramite un'analisi grafica si è poi cercato di capire cosa si riesce a spiegare in più con l'aumento del rango del tensore di approssimazione. Per tutti i paesi l'aumento del rango sembra portare agli stessi effetti. In Figura 3.5 si mostrano le curve del tensore approssimato per l'Italia al variare del rango. Il tensore di approssimazione di rango 1 sembra cogliere l'andamento medio senza discriminare le differenze tra curve. Aumentando il rango a 2 il tensore inizia a cogliere le differenze tra curve nelle età anziane e nei primi anni di età dove però approssima alcune curve assegnando valori negativi. Il tensore di approssimazione a rango 3 sembra cogliere meglio le differenze, con un'approssimazione accettabile. I massimi locali come quello sui 20 anni per i maschi italiani degli anni della guerra sembrano iniziare ad essere approssimati più o meno bene dal tensore di rango pari a 6/7.

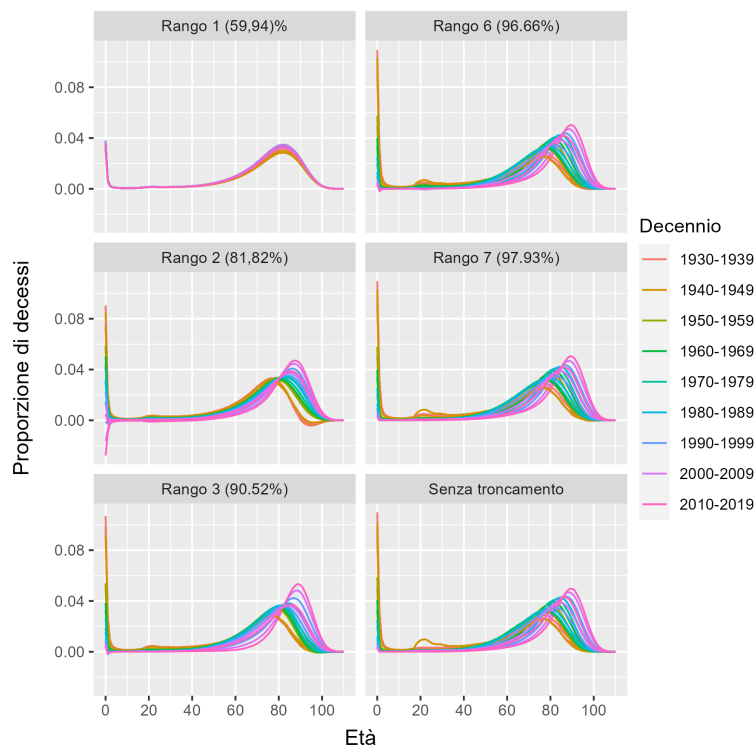


FIGURA 3.5: Curve relative al tensore approssimato con scomposizione CP per l'Italia con diversi troncamenti del rango  $r$  e relativa percentuale della norma di Frobenius spiegata scritta tra parentesi. Le curve sono colorate in base al decennio di appartenenza.

Nella Tabella 3.1 si riportano alcuni risultati relativi alle percentuali della norma di Frobenius spiegata da alcuni dei tensori di approssimazione trovati tramite la scomposizione  $CP$  troncata sui diversi paesi. In particolare, si riportano le percentuali della norma di Frobenius spiegata dal tensore di approssimazione di rango 3 i cui tensori sembrano rispettare la regola generale suggerita dalla diagnostica CONCORDIA (riportata in equazione (2.8)) per la quale per esser dei buon approssimatori devono spiegare circa il 90% della norma del tensore originale. Coerentemente con quanto detto in precedenza si nota anche che per i 6 paesi con delle particolarità nelle curve dei primi decenni (Belgio, Finlandia, Francia, Italia, Spagna e Svizzera) il tensore di rango 3 sembra spiegare un po' meno (tra l'85% e il 90%) rispetto gli altri paesi per i quali il tensore approssimato spiega più del 92% della norma del tensore originario. Dalla terza colonna della tabella (che mostra la norma di Frobenius spiegata dal tensore di rango 4) si nota che per la Finlandia la regola in equazione (2.8) suggerirebbe di approssimare il tensore con almeno rango pari a 4. Infine, nell'ultima colonna della tabella si riporta la norma di Frobenius spiegata dal tensore di rango 10 e si vede come per tutti i paesi spieghi circa il 99% del tensore originale, a parte per l'Islanda che spiega il 97,66% cosa che si notava anche dalla Figura 3.1.

Paese	Norma spiegata con $CP$ troncata ( $r = 3$ )	Norma spiegata con $CP$ troncata ( $r = 4$ )	Norma spiegata con $CP$ troncata ( $r = 10$ )
BEL	89.55%	91.99%	99.10%
DEN	92.61%	94.55%	98.91%
FIN	86.81%	90.31%	98.83%
FRA	87.52%	93.40%	99.14%
ICE	92.25%	93.97%	97.74%
ITA	89.65%	93.71%	99.16%
NOR	92.51%	94.34%	99.02%
SPAIN	87.31%	94.82%	99.17%
SWE	92.54%	95.81%	99.25%
SWI	90.79%	93.92%	99.13%
UK	92.48%	93.61%	99.36%
USA	93.36%	94.72%	99.37%

TABELLA 3.1: Norma di Frobenius spiegata dalle scomposizioni  $CP$  troncate a rango 3,4 e 10.

### 3.3 Riduzione della complessità tramite scomposizione di Tucker

Si è poi cercata anche un'approssimazione per ogni paese tramite la scomposizione di Tucker troncata (il cui tensore approssimato viene indicato con  $\hat{\mathcal{X}}^{(Tucker)}$ ). In particolare, per ogni numero totale di componenti da 3 a 17 ( $s = R_1 + R_2 + R_3 = \{3, \dots, 17\}$ ) si è cercata la combinazione di ranghi  $R_1, R_2, R_3$  tale che  $\hat{\mathcal{G}} \in \mathbb{R}^{R_1 \times R_2 \times R_3}$  per la quale  $\mathcal{G} \times_1 A^{(1)} \times_2 A^{(2)} \times_3 A^{(3)}$  ( $= \hat{\mathcal{X}}^{(Tucker)}$ ) spieghi maggiormente la norma di Frobenius del tensore originale, escludendo i casi in cui la dimensione di una moda di  $\mathcal{G}$  è maggiore del prodotto delle altre due (come suggerito da Timmerman & Kiers (2000)).

Successivamente sono state confrontate le soluzioni per i diversi valori di  $s$  tramite tre criteri: il primo riguarda il punto di gomito dello “*scree-plot*” sulla base della norma di Frobenius residua, il secondo riguarda la scomposizione scelta tramite procedura *DIFFIT* (illustrata in 2.3) e l'ultimo riguarda l'analisi grafica delle curve ottenute dal tensore approssimato.

In Figura 3.6 si riportano gli *scree-plot* al variare di  $s$  per ogni paese. Si nota che  $s = 4$  da gli stessi risultati di  $s = 3$  questo perché in un tensore a 3 vie  $s = 4$  impone che per una moda la dimensione sia pari a 2 e per le altre due sia pari a 1, questo significa che una moda avrà dimensione superiore al prodotto delle altre due e quindi come visto nella sezione 2.3 produce gli stessi risultati del tensore con  $s = 3$  in termini di norma di

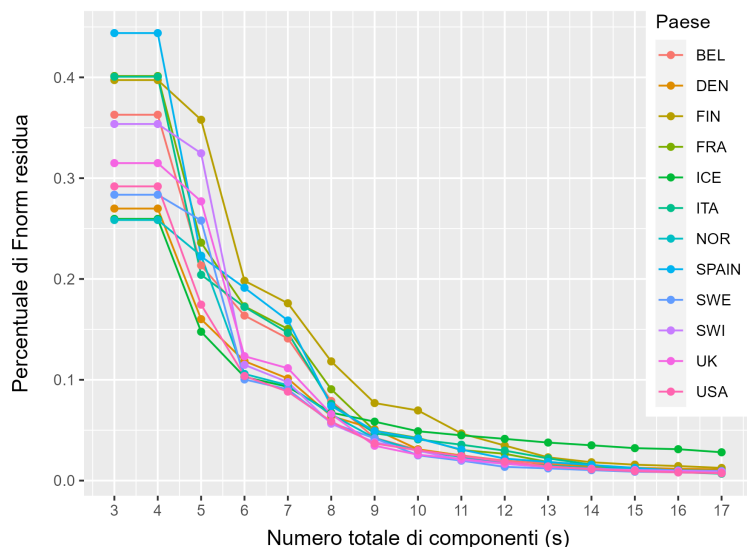


FIGURA 3.6: *Scree plots* colorati per paese delle scomposizioni di Tucker troncate.

Frobenius spiegata. Le curve non sembrano avere un punto di gomito ma da  $s$  pari a 9 o 10 non sembra migliorare di molto l'approssimazione, l'unica eccezione sembra essere la Finlandia per la quale la curva fino a  $s = 13$  scende più lentamente delle altre. Si nota infine che come nella scomposizione *CP* per l'Islanda il tensore troncato da un certo  $s$  in poi sembra spiegare meno rispetto ai tensori troncati degli altri paesi con stessa  $s$ .

La tabella 3.2 riporta per ogni paese la scomposizione di Tucker scelta tramite la procedura DIFFIT. Per Belgio, Islanda, Spagna, Regno Unito e Stati Uniti la procedura seleziona un  $s$  pari a 8 o 9 come intuito precedentemente tramite gli *scree-plot*, per Norvegia, Svezia e Svizzera seleziona  $s = 6$  con una percentuale della norma di Frobenius spiegata pari a circa il 90%; per Finlandia e Italia invece seppur la selezione sia comunque pari a 5 e 6 la percentuale della norma di Frobenius spiegata risulta pari al 80%; infine, per Danimarca e Francia la procedura seleziona un  $s$  maggiore di 10 con un adattamento di circa il 99% della norma di Frobenius del tensore originale.

Paese	Numero totale di componenti ( $s$ )	Ranghi $R1, R2, R3$	Norma spiegata da $\mathcal{X}^{(Tucker)}$
BEL	9	4,2,3	95.76%
DEN	12	5,2,5	98.26%
FIN	6	2,2,2	80.18%
FRA	16	8,2,6	99.09%
ICE	8	3,2,3	93.25%
ITA	5	2,1,2	79.59%
NOR	6	2,2,2	89.42%
SPAIN	8	3,2,3	92.60%
SWE	6	2,2,2	89.97%
SWI	6	2,2,2	88.52%
UK	9	4,2,3	96.54%
USA	9	4,2,3	96.32%

TABELLA 3.2: Norma di Frobenius spiegata dalla miglior scomposizione di Tucker troncata secondo la procedura *DIFFIT* per ogni paese.

In Figura 3.7 si mostrano gli *scree-plot* divisi secondo i gruppi individuati dalla procedura *DIFFIT* e si nota come le curve all'interno dei gruppi siano molto simili mentre siano diverse tra i gruppi (ad eccezione del gruppo di Danimarca e Francia che sembra essere simile al gruppo più numeroso). Gli *scree-plot* di Norvegia Svezia e Svizzera sembrano avere un punto di gomito su  $s = 6$  mentre Finlandia in  $s = 6$  e Italia in  $s = 5$ , nonostante il salto, non sembrano avere un punto di gomito in quanto la curva continua a scendere in maniera abbastanza evidente. Viste le scelte prese per Danimarca, Finlandia, Francia e Italia tramite la procedura *DIFFIT*, che non risultano essere totalmente appropriate, si è deciso di valutare graficamente le curve risultanti dalle diverse approssimazioni per ogni paese. Come per le approssimazioni ottenute dalla scomposizione CP anche per quelle ottenute dalla scomposizione di Tucker i risultati tra i diversi paesi al variare della riduzione sembrano essere simili, si riportano dunque i risultati solo per l'Italia.

In Figura 3.8 si riportano quindi le curve relative al tensore approssimato per l'Italia con 9 diversi troncamenti. Il tensore approssimato con  $s = 3$  mostra una relazione evidente con il tensore approssimato dalla scomposizione CP troncata a rango 1 (si ricorda che come detto nella sezione 2.3 la scomposizione di Tucker può essere vista come una generalizzazione della scomposizione CP), cogliendo l'andamento medio senza discriminare le curve. La scomposizione di Tucker con  $s$  tra 5 e 7 sembra avere una relazione con la scomposizione CP a rango 2, infatti coglie le differenze nelle età anziane e nei primi anni di età dove però approssima alcune curve con valori negativi. Il tensore

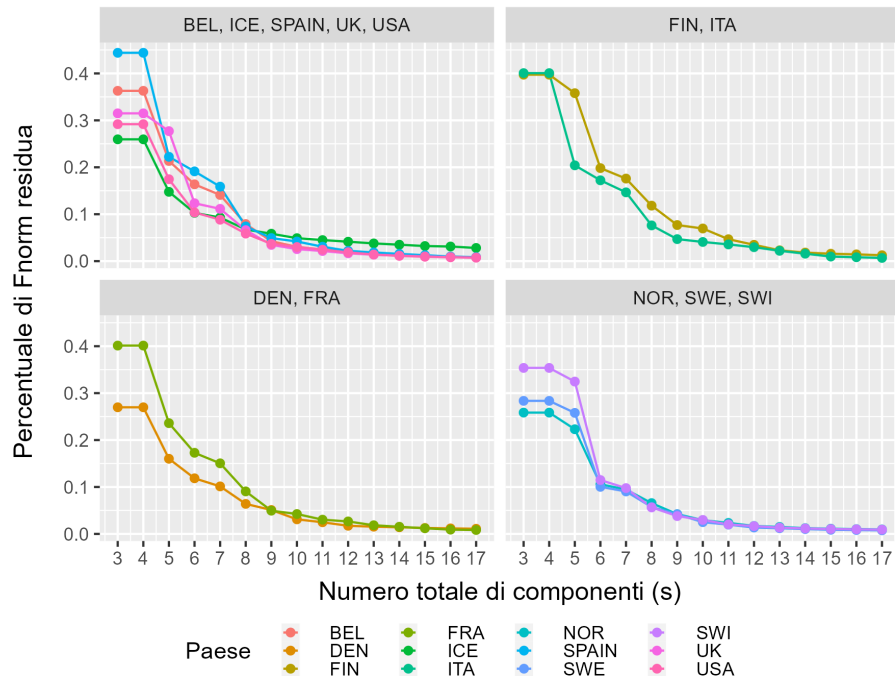


FIGURA 3.7: Scree plots delle scomposizioni di Tucker troncate. Curve colorate per paese e divise in base alla scelta di  $s$  secondo la procedura *DIFFIT*.

con  $s$  tra 8 e 9 sembra approssimare in maniera accettabile le curve senza cogliere però alcuni dettagli come il picco sui 20 anni per i maschi nel quinto decennio.

Questi risultati per  $s$  pari a 8 o 9 sono confrontabili con quelli della scomposizione *CP* troncata a rango 3 infatti anche il valore della norma spiegata risulta essere intorno al 90%. I dettagli come il picco sui 20 anni nel periodo del 1940-1949 sembrano iniziare ad essere colti con  $s = 12$ , il cui tensore spiega circa il 97% della norma di Frobenius del tensore originale esattamente come nel caso della scomposizione *CP* con rango 7 di Figura 3.5. L'approssimazione con  $s = 5$  scelta dalla procedura *DIFFIT* per l'Italia non sembra pertanto essere accettabile.

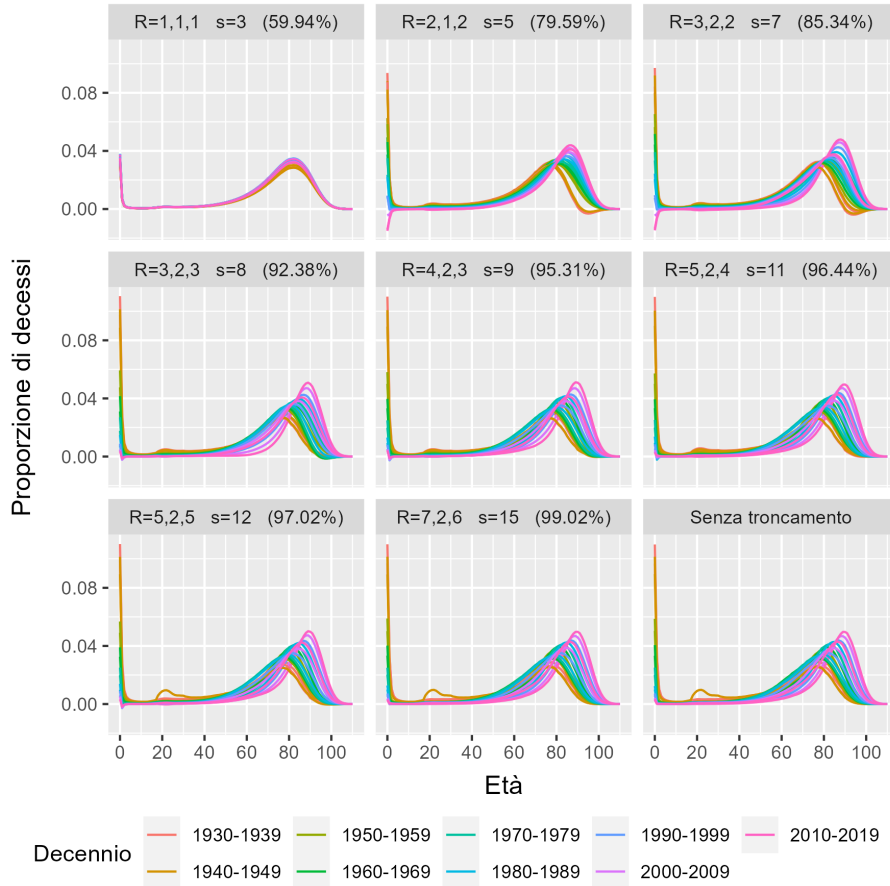


FIGURA 3.8: Curve colorate per decennio relative al tensore approssimato per l'Italia con diversi troncamenti e relativa percentuale della norma di Frobenius spiegata scritta tra parentesi.

Nelle tabelle 3.3 e 3.4 sono riportati i risultati della procedura *DIFFIT* rispettivamente per Italia e Francia. Si nota che nonostante la scelta della procedura non sia adeguata la seconda miglior soluzione risulta essere esattamente la scomposizione con  $s$  pari a 9 che è la soluzione in linea con gli altri paesi.



ITALIA

Ranghi $R1, R2, R3$	Numero totale di componenti ( $s$ )	Norma spiegata da $\hat{\mathcal{X}}^{(Tucker)}$	Tasso di miglioramento ( $b$ )
2,1,2	5	79.59%	6.17
4,2,3	9	95.31%	4.82
7,2,6	15	99.02%	4.11
3,2,3	8	92.38%	2.41
6,2,5	13	97.80%	1.26
8,2,6	16	99.17%	1.02
7,2,5	14	98.41%	1.01
4,2,4	10	95.91%	0.78
2,2,2	6	82.77%	0.45
8,2,7	17	99.32%	0.21

TABELLA 3.3: Risultati della procedura *DIFFIT* sul tensore relativo all'Italia con ordinamento dei tensori secondo il tasso di miglioramento  $b$  definito dalla procedura

FRANCIA

Ranghi $R1, R2, R3$	Numero totale di componenti ( $s$ )	Norma spiegata da $\hat{\mathcal{X}}^{(Tucker)}$	Tasso di miglioramento ( $b$ )
8,2,6	16	99.09%	3.51
4,2,3	9	95.05%	3.41
2,1,2	5	76.40%	2.62
6,2,5	13	98.15%	2.49
5,2,4	11	96.96%	1.47
3,2,3	8	90.94%	1.46
7,2,6	15	98.80%	1.14
2,2,2	6	82.71%	1.06
8,2,7	17	99.17%	0.10

TABELLA 3.4: Risultati della procedura *DIFFIT* sul tensore relativo alla Francia con ordinamento dei tensori secondo il tasso di miglioramento  $b$  definito dalla procedura.

### 3.4 Ricerca di gruppi tra i tensori tramite DEEM clustering

Come prima analisi di *clustering* si sono cercati gruppi di paesi che presentano tensori simili tramite un approccio *model based* tensoriale basato sul *DEEM*. Sono stati selezionati  $\lambda$  (posto costante per ogni iterazione per quanto detto nella sezione 2.4) e  $K$  tramite il criterio di informazione *BIC* (*bayesian information criterion*), mentre le

probabilità iniziali di appartenere ad ogni *cluster* sono state inizializzate tramite l'algoritmo "k-medie" svolto sulla *vettorizzazione* dei tensori. Il parametro  $K$  selezionato è risultato pari a 2, ma viste le considerazioni fatte nella sezione 2.4 sulla non consistenza della scelta di  $K$  si è deciso di provare a fissarlo pari a 2, 3 e 4 regolando soltanto  $\lambda$ .

Nelle tabelle 3.5, 3.6, 3.7 si riportano fissato  $K$  le probabilità associate ai *cluster* per ogni paese. Dalle tabelle emerge che ci sono dei paesi con le stesse probabilità di appartenere ai diversi *cluster* indipendentemente da  $K$ . In particolare: Spagna e Italia hanno sempre probabilità assegnata pari a 1 di stare nello stesso cluster e 0 negli altri; Belgio e Francia condividono sempre probabilità quasi uguali di appartenenza ai diversi cluster e con  $K$  pari a 3 o 4 hanno probabilità associata ad uno dei cluster pari a 1 (con  $K=2$  questo non accade probabilmente perché sono diversi da Italia e Spagna ma nel caso di due soli cluster sono più simili a questi paesi piuttosto che ai paesi nordici); infine, Islanda, Norvegia e Svezia risultano avere sempre le stesse probabilità, più precisamente risultano avere probabilità associata ad uno dei cluster di circa 1. Risulta interessante notare come questi gruppi che sembrano emergere siano tutti composti da paesi vicini geograficamente con conseguenti storie molto simili, basti pensare a Italia e Spagna che oltre ad essere vicine geograficamente condividono una storia recente molto simile influenzata principalmente dal fascismo italiano, dal franchismo spagnolo e dalle crisi degli ultimi decenni.

Per quanto riguarda la Svizzera si nota come venga sempre accostata con maggior probabilità al *cluster* composto da Islanda, Norvegia e Svezia, ma con probabilità diverse dagli altri paesi associati allo stesso *cluster*; questo probabilmente perché seppur abbia una storia simile ai paesi nordici, geograficamente e culturalmente non risulta essere accostabile a quei paesi. La Finlandia al contrario essendo un paese nordico con una storia recente caratterizzata dal coinvolgimento durante la seconda guerra mondiale che ha portato alle particolarità delle curve come il picco sui 20 anni per i maschi del 1940 (curve illustrate in Figura 3.3), risulta avere sempre probabilità divise tra due *cluster*.

Quindi viste le considerazioni fatte il *clustering* con  $K = 4$  risulta il più adeguato, perciò si riporta in tabella 3.8 la divisione dei paesi trovata tramite DEEM con  $K = 4$  e in Figura 3.9 le curve medie ricavate dal tensore medio di ogni cluster. Risultano 18 curve medie per cluster (ognuna delle quali è relativa ad un sesso in un decennio), dalle quali si vede come i cluster si differenzino in maniera particolare per le mortalità infantili e per i massimi locali sulle età intorno ai 20 anni.

Paese	$Pr(Y_i = k_1 \mathcal{X}_i, \theta)$	$Pr(Y_i = k_2 \mathcal{X}_i, \theta)$
BEL	0.17	0.83
DEN	0.94	0.06
FIN	0.68	0.32
FRA	0.18	0.82
ICE	0.98	0.02
ITA	0.00	1.00
NOR	0.99	0.01
SPAIN	0.00	1.00
SWE	0.99	0.01
SWI	0.95	0.05
UK	0.91	0.09
USA	0.93	0.07

TABELLA 3.5: Probabilità individuate dal *DEEM* (con  $K = 2$ ) di appartenere ai due *cluster* per ogni paese.

Paese	$Pr(Y_i = k_1 \mathcal{X}_i, \theta)$	$Pr(Y_i = k_2 \mathcal{X}_i, \theta)$	$Pr(Y_i = k_3 \mathcal{X}_i, \theta)$
BEL	0.00	1.00	0.00
DEN	0.83	0.17	0.00
FIN	0.22	0.78	0.00
FRA	0.00	1.00	0.00
ICE	0.99	0.01	0.00
ITA	0.00	0.00	1.00
NOR	0.99	0.01	0.00
SPAIN	0.00	0.00	1.00
SWE	1.00	0.00	0.00
SWI	0.77	0.23	0.00
UK	0.49	0.51	0.00
USA	0.15	0.85	0.00

TABELLA 3.6: Probabilità individuate dal *DEEM* (con  $K = 3$ ) di appartenere ai tre *cluster* per ogni paese.

Paese	$Pr(Y_i = k_1 \mathcal{X}_i, \theta)$	$Pr(Y_i = k_2 \mathcal{X}_i, \theta)$	$Pr(Y_i = k_3 \mathcal{X}_i, \theta)$	$Pr(Y_i = k_4 \mathcal{X}_i, \theta)$
BEL	0.00	0.00	1.00	0.00
DEN	0.02	0.98	0.00	0.00
FIN	0.00	0.64	0.36	0.00
FRA	0.00	0.00	1.00	0.00
ICE	0.99	0.01	0.00	0.00
ITA	0.00	0.00	0.00	1.00
NOR	1.00	0.00	0.00	0.00
SPAIN	0.00	0.00	0.00	1.00
SWE	1.00	0.00	0.00	0.00
SWI	0.75	0.25	0.00	0.00
UK	0.03	0.97	0.00	0.00
USA	0.01	0.98	0.01	0.00

TABELLA 3.7: Probabilità individuate dal *DEEM* (con  $K = 4$ ) di appartenere ai quattro *cluster* per ogni paese.

Cluster 1	Cluster 2	Cluster 3	Cluster 4
Svizzera	Danimarca	Belgio	Italia
Svezia	Finlandia	Francia	Spagna
Norvegia	UK		
Islanda	USA		

TABELLA 3.8: Gruppi di paesi formati dal *DEEM* con  $K = 4$ .

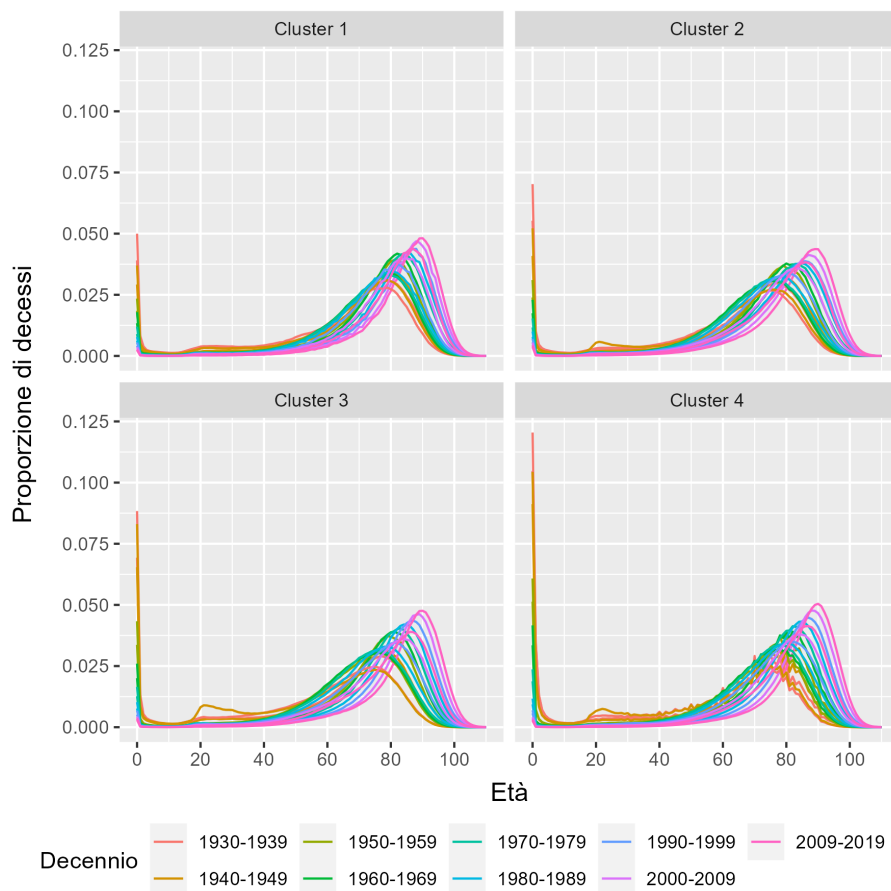


FIGURA 3.9: Curve di distribuzione dell'età alla morte derivanti dai tensori medi dei 4 cluster individuati con *DEEM*, divise per cluster e colorate per periodo.

### 3.5 Ricerca di gruppi all'interno dei tensori tramite TBM clustering

Come seconda analisi di *clustering* sono stati cercati gruppi multi-direzionali (blocchi) all'interno dei tensori relativi ai paesi tramite lo *sparse tensor block model*, per individuare anche possibili gruppi di età e periodo.

Dalle analisi sulla scomposizione di Tucker (che ha una relazione con il *TBM clustering*) si è notato che in quasi tutte le soluzioni la seconda moda non veniva ridotta ma tenuta con rango *n-esimo* pari alla dimensione della moda. Nella tabella 3.9 si riportano le norme di Frobenius spiegate dalle scomposizioni di Tucker troncate con rango della seconda moda pari a 1 per l'Italia. I ranghi per la prima e per la terza moda vengono aumentati in contemporanea perché altrimenti una moda avrebbe rango più grande del prodotto delle altre due non migliorando l'approssimazione per quanto detto nella sezione 2.3. Si nota come il massimo della norma di Frobenius che si riesce a spiegare

Ranghi $R_1, R_2, R_3$	Numero totale di componenti ( $s$ )	Norma spiegata da $\hat{\mathcal{X}}^{(Tucker)}$
1,1,1	3	59.94 %
2,1,2	5	79.59 %
3,1,3	7	83.79 %
4,1,4	9	83.89 %
5,1,5	11	83.96 %
6,1,6	13	83.97 %
7,1,7	15	83.98 %
8,1,8	17	83.98 %
9,1,9	19	83.98 %

TABELLA 3.9: Norma di Frobenius spiegata dalle scomposizioni di Tucker troncate con rango della seconda moda pari a 1 per il tensore relativo all'Italia.

è l'84% della norma del tensore originario, perciò la riduzione della dimensione relativa al sesso sembra comportare un'approssimazione non molto buona del tensore originario. Questi risultati sono probabilmente dovuti alle differenze di sesso mostrate nelle analisi preliminari di sezione 1.2.

Viste queste differenze di genere e visto che il *TBM* fa un *clustering* su ogni moda assumendo che ci sia un raggruppamento possibile (non considerando perciò il caso in cui il rango  $n$ -esimo del tensore nucleo sia pari alla dimensione originaria), per vedere se la divisione in gruppi dei paesi cambi considerando separatamente i sessi in ogni paese (quindi per vedere se il sesso ha un'influenza sul raggruppamento dei paesi) e per evitare di perdere l'informazione sul sesso (tenendo in considerazione un unico gruppo), si è deciso di applicare il *TBM* considerando come unità statistica un sesso in un paese ed inserendo l'informazione relativa all'unità statistica come terza moda di un unico tensore in cui le prime due mode rimangono l'età (di dimensione 111) e il decennio (di dimensione pari a 9). Perciò, il tensore su cui si è applicato il *TBM* ha dimensione  $111 \times 9 \times 24$ .

La dimensione del *clustering*  $R = R_1, R_2, R_3$  e in seguito il parametro di regolazione relativo alla penalità lasso sono stati selezionati tramite il criterio di informazione *BIC* accennato nella sezione 2.5. Questi criteri hanno portato ad un  $R = R_1, R_2, R_3 = 8, 24, 8$ , per il quale l'algoritmo per il calcolo del *TBM* è stato inizializzato tramite un *k-medie* su ogni moda indipendentemente dalle altre.

I risultati hanno portato a raggruppare soltanto due periodi (1950-1959 e 1960-1969) come prevedibile in quanto avendo preso in considerazione decenni i periodi sono già molto raggruppati. Per le età invece si sono formati 24 gruppi rappresentati nelle seguenti

classi di età:

[0], [1], [2 - 3], [4 - 18], [19 - 34], [35 - 43], [44 - 49], [50 - 54], [55 - 58], [59 - 62], [63 - 66], [67 - 70], [71 - 74], [75 - 78], [79 - 81], [82 - 84], [85 - 87], [88 - 90], [91 - 93], [94 - 95], [96 - 97], [98 - 99], [100 - 102], [103 - 110].

Si nota che i primi anni di età non vengono raggruppati, infatti nella sezione 1.2 si notava come la mortalità infantile in genere presenti un picco che scende molto velocemente differenziando molto i valori tra i primi anni di età. Le età giovanili (da 4 a 34 anni) vengono raggruppate in 2 *cluster*, mentre dai 44 a 102 anni si raggruppano le modalità in classi da 3/4 età per gruppo, con un ultimo *cluster* che rappresenta le età da 103 anni in poi. Risulta interessante notare come nelle età giovanili e nelle età da 103 anni in poi le distribuzioni dell'età alla morte presentino quasi sempre valori intorno allo 0, mentre dai 44 ai 100 anni si hanno valori più alti che variano a seconda della curva.

La tabella 3.10 mostra invece la divisione delle unità negli 8 *cluster*, in cui si nota che in ogni *cluster* è presente un solo sesso. I paesi vengono divisi nello stesso modo per maschi e per le femmine a parte che per la svizzera le cui femmine stanno con i paesi nordici mentre i maschi stanno con USA, UK e Danimarca. I gruppi formati si portano a considerazioni identiche a quanto visto in precedenza con il *DEEM*, infatti risultano negli stessi *cluster* paesi geograficamente vicini o comunque con una storia recente molto simile. Si evidenzia inoltre che la Svizzera con il *DEEM* risultava sempre essere classificata con i paesi nordici ma con probabilità notevoli di essere associata al gruppo degli Stati Uniti, infatti con il *TBM* cambia *cluster* in base al sesso, questo significa che forse l'indecisione vista nel *DEEM* può essere dovuta ad una somiglianza tra le curve dei maschi degli USA, UK e Danimarca con quelle dei maschi Svizzeri. Infine, è utile notare che il *TBM* raggruppa la Finlandia con Francia e Belgio, mentre il *DEEM* la raggruppava con USA, UK e Danimarca anche se con probabilità sopra al 30% di essere nel *cluster* di Belgio e Francia.

Cluster 1	Cluster 2	Cluster 3	Cluster 4
Femmine Danimarca, Femmine UK, Femmine USA	Femmine Svizzera, Femmine Svezia, Femmine Norvegia, Femmine Islanda	Femmine Belgio, Femmine Finlandia, Femmine Francia	Femmine Italia, Femmine Spagna
Cluster 5	Cluster 6	Cluster 7	Cluster 8
Maschi Danimarca, Maschi Svizzera, Maschi UK, Maschi USA	Maschi Svezia, Maschi Norvegia, Maschi Islanda	Maschi Belgio, Maschi Finlandia, Maschi Francia	Maschi Italia, Maschi Spagna

TABELLA 3.10: Gruppi di unità formati dal *TBM*.

Dalle Figure 3.10 e 3.11 che riportano le curve derivanti dal tensore a blocchi del segnale medio individuato tramite *TBM*, divise per *cluster* di appartenenza. Si nota come per il *DEEM* che i cluster si differenziano per le mortalità infantili e per i massimi locali delle curve. Inoltre, facendo attenzione alla diversa scala dell'asse delle ordinate si nota come i cluster dei maschi risultino avere valori di mortalità infantile e dei massimi locali più elevati che nei cluster delle femmine. In generale si vede che il tensore a blocchi sembra cogliere l'andamento delle curve e che il raggruppamento delle età non porta ad una grande perdita di informazione nonostante si sia passati dal valutare 111 età al valutare 8 gruppi di età.

Dalle Figure 3.12 e 3.13 che riportano le curve originali di ogni paese ordinate in base al *cluster* di appartenenza, si nota come effettivamente i paesi all'interno dello stesso *cluster* mostrano un andamento della mortalità infantile simile. In particolare, si nota che le morti a 0 anni nel primo decennio per le femmine dei paesi nel primo *cluster* sono tra il 5% e il 6% delle morti di quel decennio, per le femmine dei paesi nel secondo gruppo sono circa il 4%, per quelle del terzo *cluster* sono tra il 6% e il 7%, infine per le femmine di Italia e Spagna sono più del 10%. Per quanto riguarda i maschi invece la mortalità infantile nel *cluster* 5 è tra il 6% e l'8%, per il sesto gruppo è circa il 5%, per il settimo *cluster* è tra l'8% e il 10%, mentre per l'ultimo gruppo è più del 10%. Infine, si nota che la Finlandia presenta una proporzione di morti infantili nel primo decennio per i maschi di circa il 7% delle morti e per le femmine di circa il 6%, questo la colloca esattamente al confine tra i *cluster* di Danimarca, UK e USA e i cluster di Belgio e Francia in termini di valori della mortalità infantile nei primi decenni, motivo per cui risultano diverse le collocazioni di questo paese con *DEEM* e con *TBM*.



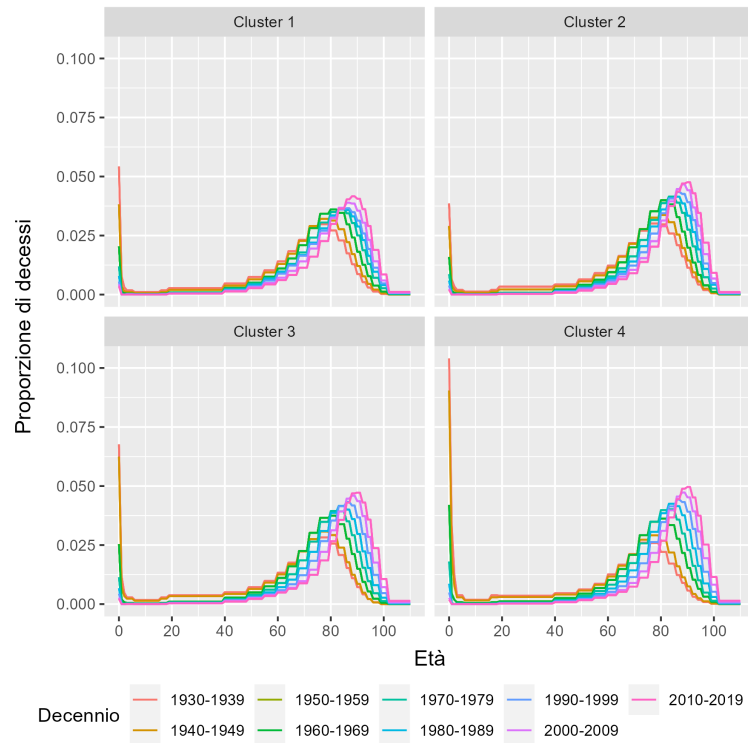


FIGURA 3.10: Distribuzioni dell'età alla morte per le femmine ricavate dal tensore del segnale medio a blocchi del *TBM*. Curve divise per cluster e colorate per decennio.

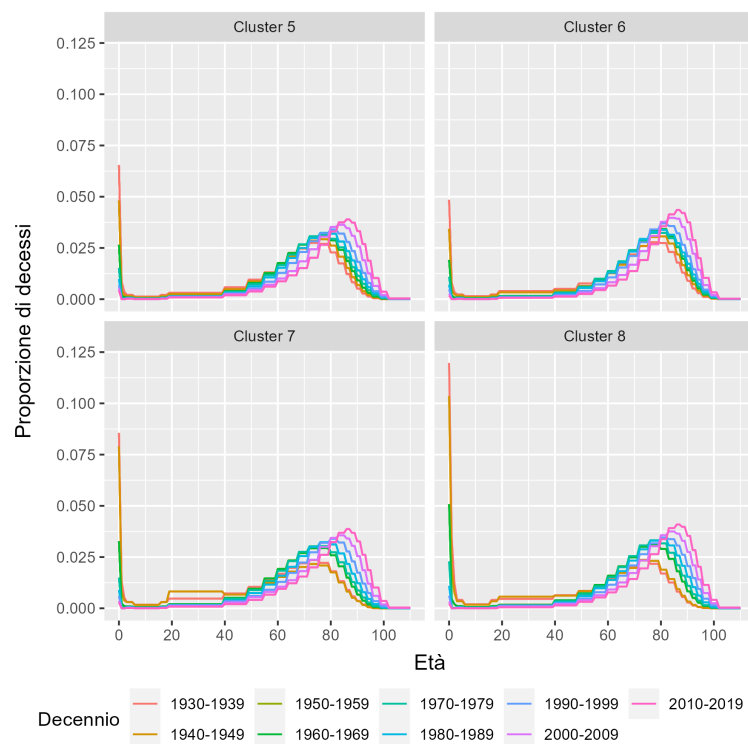


FIGURA 3.11: Distribuzioni dell'età alla morte per i maschi ricavate dal tensore del segnale medio a blocchi del *TBM*. Curve divise per cluster e colorate per decennio.

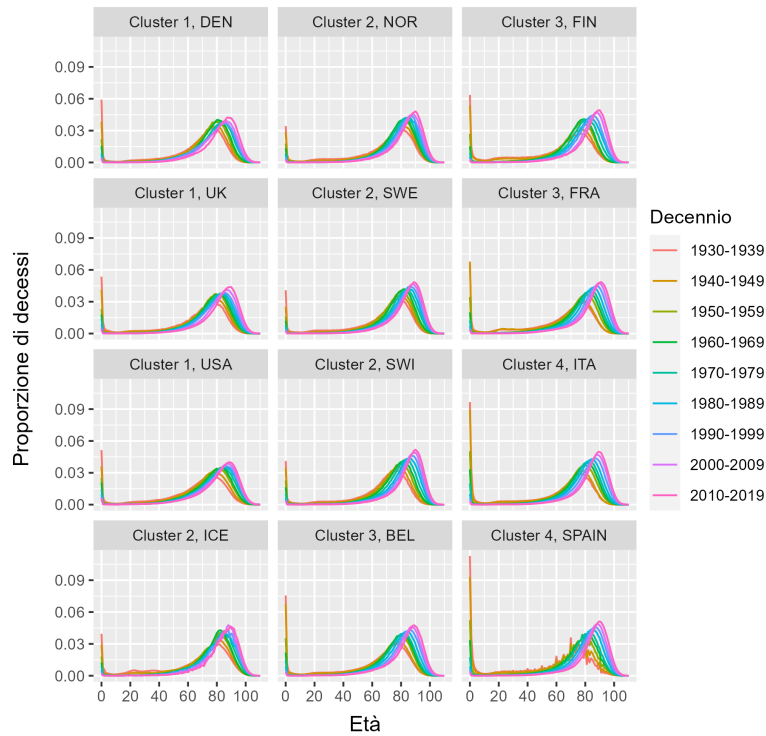


FIGURA 3.12: Distribuzioni originali dell'età alla morte delle femmine divise per paese, ordinate per i cluster di appartenenza e colorate per decennio.

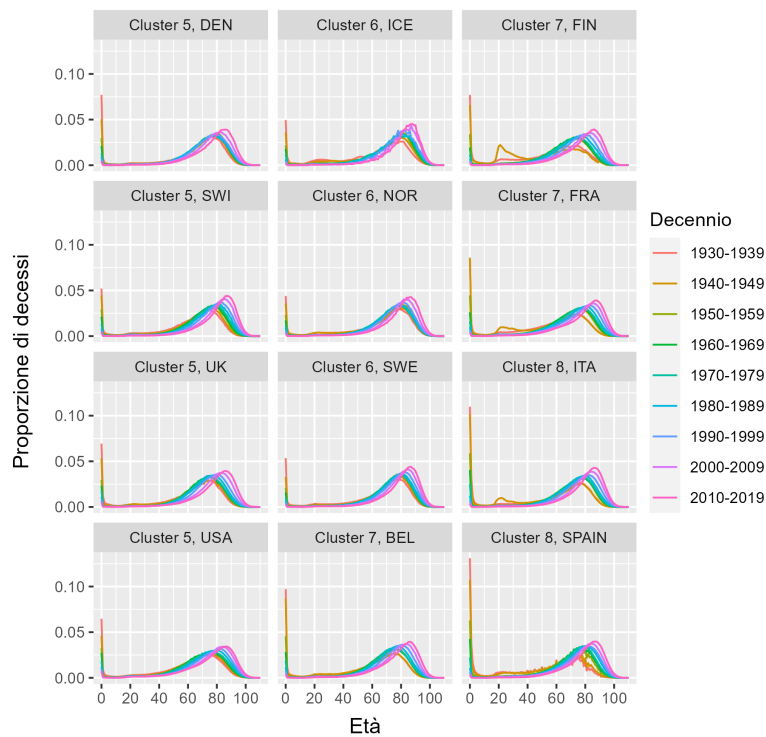


FIGURA 3.13: Distribuzioni originali dell'età alla morte dei maschi divise per paese, ordinate per i cluster di appartenenza e colorate per decennio.

# Capitolo 4

## Conclusioni

In questo lavoro di tesi si è cercato di studiare la distribuzione dell'età alla morte tramite un approccio tensoriale che tenga quindi conto della struttura multilineare dei dati.

Tramite scomposizioni tensoriali si è cercato di ridurre la complessità dei dati utilizzando come indicatore della complessità il rango tensoriale. Si è visto come a seconda della complessità del tensore (che nel caso di studio aumenta a secondo della rumorosità delle curve e del numero di massimi locali) si riescano ad approssimare più o meno bene i tensori e le curve all'interno di essi già con tensori di rango pari a 3 o 4 (trovati tramite la scomposizione *CP*) o con tensori aventi ogni rango *n-esimo* inferiore a 5 (trovati tramite scomposizione di Tucker). Inoltre, le procedure note in letteratura per valutare le diverse riduzioni a seconda della scomposizione utilizzata sembrano dare risultati ragionevoli, se accompagnate a opportune riflessioni sul obiettivo da raggiungere: ad esempio se non si vuole perdere l'informazione sui massimi locali delle curve bisognerà tenerne conto nell'utilizzo delle procedure di valutazione dell'adattamento.

Per quanto riguarda la ricerca di gruppi tra tensori, utilizzando un algoritmo *DEEM* sui 12 tensori relativi ai paesi si sono ottenuti dei *cluster* che rispecchiano la vicinanza dei paesi in termini geografici-culturali. Questa divisione risulta appropriata in quanto è noto che la mortalità dipenda molto da influenze socio culturali dovute anche alla posizione geografica dei paesi. Inoltre, si è visto come i *cluster* individuati presentino un andamento della mortalità infantile comune all'interno dei gruppi, questo probabilmente ha un'influenza importante nella costruzione dei *cluster* in quanto il valore della mortalità infantile nei primi decenni risulta avere il valore più alto per ogni paese.

Tramite un *TBM* sono stati cercati blocchi all'interno dei tensori. Come prevedibile i decenni sono risultati molto diversificati tra loro e quindi risulta poco ragionevole raggrupparli. La moda relativa alle età risulta invece avere una struttura a gruppi, infatti si potrebbero raggruppare specialmente le età giovanili in cui la proporzione di

morti è molto bassa. Infine, per quanto riguarda i paesi risultano circa gli stessi gruppi identificati con il *DEEM* raggruppando quindi paesi con caratteristiche socio-geografiche simili. Non risultano esserci differenze di genere nel raggruppamento dei paesi a parte per la Svizzera, che anche nel *DEEM* veniva classificata con una certa incertezza rispetto agli altri paesi.

In conclusione, si può dire che i risultati ottenuti da questa analisi tensoriale sembrano essere in linea con le attese e con gli altri lavori presenti in letteratura che trattano gli stessi dati con un classico approccio matriciale. Il vantaggio di questi approcci è principalmente dovuto al fatto di tener conto della struttura multilineare e quindi della relazione tra le varie mode. Lo svantaggio principale è dovuto alla complessità computazionale all'aumentare del numero di vie. Inoltre, nella scelta dei parametri di regolazione del *DEEM* si sono riscontrate difficoltà dovute al criterio di informazione utilizzato, che hanno portato ad una selezione più semplice.

Potrebbe essere interessante in futuro provare a utilizzare metodi di regressione tensoriale e di analisi di serie storiche tensoriali (come il *tensor autoregressive model*) per vedere l'effetto di variabili esplicative su questi tensori o provare a fare previsioni sull'andamento nei prossimi decenni della distribuzione dell'età alla morte.

# Appendice

---

**Algoritmo 1** *CP -ALS*

---

In ingresso va passato  $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$  e va scelto il rango  $r$ .

1. Inizializzazione:  $A_n \in \mathbb{R}^{I_n \times r}$  per  $n = 1, \dots, N$ .
2. Fino a quando non cambiano più le matrici fattoriali o viene raggiunto il numero massimo di iterazioni:  
ciclo per  $n = 1, \dots, N$ 
  - (a)  $V \leftarrow A_1^T A_1 \star \dots \star A_{n-1}^T A_{n-1} \star A_{n+1}^T A_{n+1} \star \dots \star A_N^T A_N$
  - (b)  $A_n \leftarrow X_{(1)}(A_N \odot \dots \odot A_{n+1} \odot A_{n-1} \odot \dots \odot A_1) V^\dagger$
  - (c) normalizzazione delle colonne di  $A_n$  (memorizzando le norme nella diagonale di  $\Lambda \in \mathbb{R}^{r \times r}$ )

Ritorna  $\Lambda, A_1, A_2, \dots, A_N$

---

---

**Algoritmo 2** *HOSVD*

---

In ingresso va passato il tensore  $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$  e il rango del tensore nucleo  $(R_1, R_2, \dots, R_N)$

1. Ciclo per  $n = 1, \dots, N$ 
  - (a)  $A^{(n)} \leftarrow$  i primi  $R_n$  vettori singolari sinistri della SVD di  $X_{(n)}$

Fine ciclo

2.  $\mathcal{G} \leftarrow \mathcal{X} \times_1 A^{(1)T} \times_2 A^{(2)T} \times_3 \dots \times_N A^{(N)T}$

Ritorna  $\mathcal{G}, A^{(1)}, A^{(2)}, \dots, A^{(N)}$

---

**Algoritmo 3 HOOI**

In ingresso va passato il tensore  $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$  e il rango del tensore nucleo  $(R_1, R_2, \dots, R_N)$

1. Inizializzazione:  $A_n \in \mathbb{R}^{I_n \times R_n}$  per  $n = 1, \dots, N$  usando HOSVD
2. Finché non si raggiunge un minimo di  $\|\mathcal{X} - (\mathcal{G} \times_1 A^{(1)} \times_2 A^{(2)} \times_3 \dots \times_N A^{(N)})\|_F$  oppure si raggiunge un numero massimo di iterazioni:
 

ciclo per  $n = 1, \dots, N$

  - (a)  $\mathcal{Y} \leftarrow \mathcal{X} \times_1 A^{(1)T} \times_2 \dots \times_{n-1} A^{(n-1)T} \times_{n+1} A^{(n+1)T} \times_{n+2} \dots \times_N A^{(N)T}$
  - (b)  $A^{(n)} \leftarrow$  i primi  $R_n$  vettori singolari sinistri della SVD di  $Y_{(n)}$
3.  $\mathcal{G} \leftarrow \mathcal{X} \times_1 A^{(1)T} \times_2 A^{(2)T} \times_3 \dots \times_N A^{(N)T}$

Ritorna  $\mathcal{G}, A^{(1)}, A^{(2)}, \dots, A^{(N)}$

---

**Algoritmo 4** *DEEM* ( $\mathcal{X}_i \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ ,  $K$ )
 

---

 1. Inizializzazione:  $\hat{\pi}_k^{(0)}$ ,  $\hat{\mu}_k^{(0)}$ ,  $\hat{\Sigma}_n^{(0)}$ .

 2. Ciclo per  $t = 1, \dots$ , fino a convergenza:

 (a) *E-Step* potenziato:

 i. Si trovano i  $\hat{B}_2^{(t+1)}, \dots, \hat{B}_K^{(t+1)} \in \mathbb{R}^{I_1 \times \dots \times I_N}$  come i  $B_2, \dots, B_K$  che minimizzano la seguente funzione obbiettivo

$$\sum_{k=2}^K \left( \langle B_k, [[B_k; \hat{\Sigma}_1^{(t)}, \dots, \hat{\Sigma}_N^{(t)}]] \rangle - 2 \langle B_k, \hat{\mu}_k^{(t)} - \hat{\mu}_1^{(t)} \rangle \right) + \lambda^{(t+1)} \sum_{\mathcal{J}} \sqrt{\sum_{k=2}^K B_{k;\mathcal{J}}^2},$$

 dove con l'indice  $\mathcal{J}$  si intende un elemento  $\{i_1, \dots, i_N\}$  del tensore e quindi  $B_{k;\mathcal{J}} = B_{k;[i_1, \dots, i_N]}$ .

 ii. Per  $i = 1, \dots, m$  e  $k = 1, \dots, K$  si calcolano le probabilità

$$\hat{\xi}_{i1}^{(t+1)} = \frac{\hat{\pi}_1^{(t)}}{\hat{\pi}_1^{(t)} + \sum_{k=2}^K \hat{\pi}_k^{(t)} \cdot \exp\{\langle \mathcal{X}_i - \frac{1}{2}(\hat{\mu}_k^{(t)} + \hat{\mu}_1^{(t)}) \rangle, \hat{B}_k^{(t+1)}\}},$$

$$\hat{\xi}_{ik}^{(t+1)} = \frac{\hat{\pi}_k^{(t)} \cdot \exp\{\langle \mathcal{X}_i - \frac{1}{2}(\hat{\mu}_k^{(t)} + \hat{\mu}_1^{(t)}) \rangle, \hat{B}_k^{(t+1)}\}}{\hat{\pi}_1^{(t)} + \sum_{k=2}^K \hat{\pi}_k^{(t)} \cdot \exp\{\langle \mathcal{X}_i - \frac{1}{2}(\hat{\mu}_k^{(t)} + \hat{\mu}_1^{(t)}) \rangle, \hat{B}_k^{(t+1)}\}}, \quad k > 1.$$

 (b) *M-Step* potenziato:

i. Si aggiornano

$$\hat{\pi}_k^{(t+1)} = \left( \sum_{i=1}^m \hat{\xi}_{ik}^{(t+1)} \right) / m,$$

$$\hat{\mu}_k^{(t+1)} = \left( \sum_{i=1}^m \hat{\xi}_{ik}^{(t+1)} \mathcal{X}_i \right) / \left( \sum_{i=1}^m \hat{\xi}_{ik}^{(t+1)} \right).$$

ii. Si calcolano gli stimatori di covarianza intermedi

$$\check{\Sigma}_n^{(t+1)} = \frac{I_n}{m(\prod_{j=1}^N I_j)} \sum_{i=1}^m \sum_{k=1}^K \left[ \hat{\xi}_{ik}^{(t+1)} (\mathcal{X}_i - \hat{\mu}_k^{(t+1)})_{(n)} (\mathcal{X}_i - \hat{\mu}_k^{(t+1)})_{(n)}^T \right].$$

 iii. Si ottengono gli stimatori di covarianza DEEM  $\hat{\Sigma}_n^{(t+1)}$  riscaldando  $\check{\Sigma}_n^{(t+1)}$ 

$$\hat{\Sigma}_n^{(t+1)} = \check{\Sigma}_n^{(t+1)} / \check{\sigma}_{n;1,1}^{(t+1)}, \quad \forall n > 1,$$

$$\hat{\Sigma}_1^{(t+1)} = \check{\Sigma}_1^{(t+1)} (\hat{\sigma}_{1;1,1}^{(t+1)} / \check{\sigma}_{1;1,1}^{(t+1)}),$$

$$\text{dove } \hat{\sigma}_{1;1,1}^{(t+1)} = \frac{1}{m} \sum_{i=1}^m \sum_{k=1}^K \left[ \hat{\xi}_{ik}^{(t+1)} (\mathcal{X}_{i;1,\dots,1} - \hat{\mu}_{k;1,\dots,1}^{(t+1)})^2 \right].$$

 Ritorna  $\hat{\xi}_{ik}, \hat{\mu}_k, \hat{\pi}_k, \hat{\Sigma}_n$ .
 

---

---

**Algoritmo 5** *Clustering multi-direzionale basato sul TBM*


---

In ingresso va passato  $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$  e la dimensione del *clustering*  $R = (R_1, \dots, R_N)$ .

1. Inizializzazione dei *clustering* marginali tramite  $N$  *k-means* indipendenti (uno per ogni moda).
2. Ripetizione fino a convergenza di:
  - (a) aggiornamento degli elementi del tensore nucleo  $\mathcal{C}$

$$\hat{c}_{r_1, \dots, r_N} = \frac{1}{n_{r_1, \dots, r_N}} \sum_{M_1^{-1}(r_1) \times \dots \times M_N^{-1}(r_N)} y_{i_1, \dots, i_N}, \quad (\text{A.1})$$

dove  $M_n^{-1}(r_n)$  indica gli indici che derivano dal *cluster*  $r_n$  nel  $n$ -esima moda e  $n_{r_1, \dots, r_N}$  indica il numero di entrate derivati dal gruppo (o blocco) denotato dagli indici  $(r_1, \dots, r_N)$ .

- (b) Ciclo per  $n = 1, \dots, N$ :
  - aggiornamento delle matrici di appartenenza del  $n$ -esima moda ( $\hat{M}_n$ ) assegnando ad ogni riga  $i_n$  tutti 0 a parte un 1 nella colonna  $\hat{r}_n \in \{1, \dots, R_n\}$  tale che

$$\hat{r}_n = \arg \min_{r \in \{1, \dots, R_n\}} \sum_{\{i_1, \dots, i_{n-1}, i_{n+1}, \dots, i_N\}} (\hat{c}_{\hat{r}_1, \dots, r, \dots, \hat{r}_N} - y_{i_1, \dots, i_n, \dots, i_N})^2.$$

Ritorna  $\hat{\mathcal{C}} \in \mathbb{R}^{R_1 \times \dots \times R_N}$  e le matrici  $\hat{M}_n$  ( $\forall n = 1, \dots, N$ ).

---



# Bibliografia

- AZZALINI, A. & SCARPA, B. (2012). *Data analysis and data mining: An introduction*. OUP USA.
- BADER, B. W. & KOLDA, T. G. (2006). Algorithm 862: Matlab tensor classes for fast algorithm prototyping. *ACM Transactions on Mathematical Software (TOMS)* **32**, 635–653.
- BRO, R. & KIERS, H. A. (2003). A new efficient method for determining the number of components in parafac models. *Journal of Chemometrics: A Journal of the Chemometrics Society* **17**, 274–286.
- CARROLL, J. D. & CHANG, J.-J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of “eckart-young” decomposition. *Psychometrika* **35**, 283–319.
- CATTELL, R. B. (1944). “parallel proportional profiles” and other principles for determining the choice of factors by rotation. *Psychometrika* **9**, 267–283.
- DE LATHAUWER, L., DE MOOR, B. & VANDEWALLE, J. (2000). On the best rank-1 and rank-( $r_1, r_2, \dots, r_n$ ) approximation of higher-order tensors. *SIAM journal on Matrix Analysis and Applications* **21**, 1324–1342.
- GLOUB, G. H. & VAN LOAN, C. F. (1996). Matrix computations. *Johns Hopkins Universtiy Press, 3rd edition* .
- HARSHMAN, R. A. et al. (1970). Foundations of the parafac procedure: Models and conditions for an “ explanatory” multimodal factor analysis .
- HITCHCOCK, F. L. (1927). The expression of a tensor or a polyadic as a sum of products. *Journal of Mathematics and Physics* **6**, 164–189.
- KAPTEYN, A., NEUDECKER, H. & WANSBEEK, T. (1986). An approach to n-mode components analysis. *Psychometrika* **51**, 269–275.

- KIERS, H. A. (2000). Towards a standardized notation and terminology in multiway analysis. *Journal of Chemometrics: A Journal of the Chemometrics Society* **14**, 105–122.
- KOLDA, T. G. (2006). Multilinear operators for higher-order decompositions. Tech. rep., Sandia National Laboratories (SNL), Albuquerque, NM, and Livermore, CA . . .
- KOLDA, T. G. & BADER, B. W. (2009). Tensor decompositions and applications. *SIAM review* **51**, 455–500.
- KROONENBERG, P. M. & DE LEEUW, J. (1980). Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika* **45**, 69–97.
- LI, J., BIEN, J. & WELLS, M. T. (2018). rtensor: An r package for multidimensional array (tensor) unfolding, multiplication, and decomposition. *Journal of Statistical Software* **87**, 1–31.
- MAI, Q., ZHANG, X., PAN, Y. & DENG, K. (2022). A doubly enhanced em algorithm for model-based tensor clustering. *Journal of the American Statistical Association* **117**, 2120–2134.
- TIMMERMAN, M. E. & KIERS, H. A. (2000). Three-mode principal components analysis: Choosing the numbers of components and sensitivity to local optima. *British journal of mathematical and statistical psychology* **53**, 1–16.
- TUCKER, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika* **31**, 279–311.
- WANG, M. & ZENG, Y. (2019). Multiway clustering via tensor block models. *Advances in neural information processing systems* **32**.
- YUAN, M. & LIN, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **68**, 49–67.

