Correspondence Analysis

Analisi dei Dati[1]

[1] Corso di Laurea in Scienze Statistiche e Attuariali
Dipartimento di Diritto, Economia, Management e Metodi Quantitativi (DEMM)
Università degli Studi del Sannio

Prof. Pietro Amenta

Fonte: *Pietro Amenta. Appunti di Analisi dei Dati Multidimensionali*

Abbiamo visto che se riportiamo le relazioni di una tripletta statistica $(\mathbf{X}, \mathbf{Q}, \mathbf{D})$ in uno schema grafico, otteniamo lo schema di base del **Diagramma di Dualità.** Tale schema evidenzia tutte le relazioni di una Analisi in Componenti Principali della matrice $\mathbf{X}$ con metrica $\mathbf{Q}$ e pesi $\mathbf{D}$: ACP$(\mathbf{X}, \mathbf{Q}, \mathbf{D})$.

$$
\begin{tikzcd}
\Re^{p*} \arrow[r, "\mathbf{X}"] & \Re^{n} \arrow[d, "\mathbf{D}"] \\
\Re^{p} \arrow[u, "\mathbf{Q}"] & \Re^{n*} \arrow[l, "\mathbf{X}^T"]
\end{tikzcd}
$$

Possiamo ottenere le relazioni di base dell’AC utilizzando la tripletta statistica e iil diagramma di dualità. Infatti l’AC risulta essere una particolare ACP.

1. $\boxed{\text{AC come ACP}(\mathbf{F}, \mathbf{D}_I^{-1}, \mathbf{D}_J^{-1})}$ Consideriamo lo schema di dualità della tripletta $(\mathbf{F}, \mathbf{D}_I^{-1}, \mathbf{D}_J^{-1})$, con $\mathbf{u} \in F$

$$\begin{tikzcd}
F^* \arrow[r, "\mathbf{F}"] & E \arrow[d, "\mathbf{D}_I^{-1}"] \\
\boxed{\mathrm{F}} \arrow[u, "\mathbf{D}_J^{-1}"] & E^* \arrow[l, "\mathbf{F}^T"]
\end{tikzcd}$$

$$\boxed{\mathrm{ACP}(\mathbf{F}, \mathbf{D}_I^{-1}, \mathbf{D}_J^{-1})} \to \mathbf{F}^T\mathbf{D}_I^{-1}\mathbf{F}\mathbf{D}_J^{-1}\mathbf{u} = \lambda\mathbf{u}$$

$$\mathbf{D}_J^{-1/2}\mathbf{F}^T\mathbf{D}_I^{-1}\mathbf{F}\mathbf{D}_J^{-1/2}\mathbf{g} = \lambda\mathbf{g} \longleftarrow$$

dove $\mathbf{g} = \mathbf{D}_J^{-1/2}\mathbf{u}$, con $\mathbf{u} = \mathbf{D}_J^{1/2}\mathbf{g}$.

2. $\boxed{\text{AC come ACP}(\mathbf{D}_I^{-1}\mathbf{F}, \mathbf{D}_J^{-1}, \mathbf{D}_I)}$ Sviluppiamo il diagramma di dualità della tripletta $(\mathbf{D}_I^{-1}\mathbf{F}, \mathbf{D}_J^{-1}, \mathbf{D}_I)$, con $\mathbf{u} \in F$

$$
\begin{tikzcd}
F^* \arrow[r, "\mathbf{F}"] & E \arrow[r, "\mathbf{D}_I^{-\frac{1}{2}}"] \arrow[d, "\mathbf{D}_I^{-1}"] & E \arrow[r, "\mathbf{D}_I^{-\frac{1}{2}}"] \arrow[d, "\mathbf{I}"] & E \arrow[d, "\mathbf{D}_I"] \\
\boxed{\mathrm{F}} \arrow[u, "\mathbf{D}_J^{-1}"] & E^* \arrow[l, "\mathbf{F}^T"] & E^* \arrow[l, "\mathbf{D}_I^{-\frac{1}{2}}"] & E^* \arrow[l, "\mathbf{D}_I^{-\frac{1}{2}}"]
\end{tikzcd}
$$

$$\boxed{\text{ACP}(\mathbf{D}_I^{-1}\mathbf{F}, \mathbf{D}_J^{-1}, \mathbf{D}_I)} \rightarrow \mathbf{F}^T\mathbf{D}_I^{-1}\mathbf{D}_I\mathbf{D}_I^{-1}\mathbf{F}\mathbf{D}_J^{-1}\mathbf{u} = \lambda\mathbf{u}$$

$$\mathbf{D}_J^{-1/2}\mathbf{F}^T\mathbf{D}_I^{-1}\mathbf{F}\mathbf{D}_J^{-1/2}\mathbf{g} = \lambda\mathbf{g} \longleftarrow$$

dove $\mathbf{g} = \mathbf{D}_J^{-1/2}\mathbf{u}$, con $\mathbf{u} = \mathbf{D}_J^{1/2}\mathbf{g}$.

3. $\boxed{\text{AC come ACP}(\mathbf{F}\mathbf{D}_J^{-1}, \mathbf{D}_J, \mathbf{D}_I^{-1})}$ Consideriamo ora il diagramma di dualità della tripletta $(\mathbf{F}\mathbf{D}_J^{-1}, \mathbf{D}_J, \mathbf{D}_I^{-1})$

$$
\begin{tikzcd}
F^* \arrow[r, "\mathbf{D}_J^{-\frac{1}{2}}"] & F^* \arrow[r, "\mathbf{D}_J^{-\frac{1}{2}}"] & F^* \arrow[r, "\mathbf{F}"] & E \arrow[d, "\mathbf{D}_I^{-1}"] \\
F \arrow[u, "\mathbf{D}_J"] & F \arrow[u, "\mathbf{I}"'] \arrow[l, "\mathbf{D}_J^{-\frac{1}{2}}"] & \boxed{F} \arrow[u, "\mathbf{D}_J^{-1}"] \arrow[l, "\mathbf{D}_J^{-\frac{1}{2}}"] & E^* \arrow[l, "\mathbf{F}^T"]
\end{tikzcd}
$$

$$\boxed{\text{ACP}(\mathbf{F}\mathbf{D}_J^{-1}, \mathbf{D}_J, \mathbf{D}_I^{-1})} \rightarrow \mathbf{D}_J^{-1}\mathbf{F}^T\mathbf{D}_I^{-1}\mathbf{F}\mathbf{u} = \lambda\mathbf{u}$$

$$\mathbf{D}_J^{-1/2}\mathbf{F}^T\mathbf{D}_I^{-1}\mathbf{F}\mathbf{D}_J^{-1/2}\mathbf{g} = \lambda\mathbf{g} \longleftarrow$$

dove $\mathbf{g} = \mathbf{D}_J^{1/2}\mathbf{u}$, con $\mathbf{u} = \mathbf{D}_J^{-1/2}\mathbf{g}$.

4. $\boxed{\text{AC come ACP}(\mathbf{D}_I^{-1}\mathbf{F}\mathbf{D}_J^{-1}, \mathbf{D}_J, \mathbf{D}_I)}$ Consideriamo infine il diagramma di dualità della tripletta $(\mathbf{D}_I^{-1}\mathbf{F}\mathbf{D}_J^{-1}, \mathbf{D}_J, \mathbf{D}_I)$

$$\begin{tikzcd}
F^* \arrow[r, "\mathbf{D}_J^{-\frac{1}{2}}"] & F^* \arrow[r, "\mathbf{D}_J^{-\frac{1}{2}}"] & F^* \arrow[r, "\mathbf{F}"] & E \arrow[r, "\mathbf{D}_I^{-\frac{1}{2}}"] \arrow[d, "\mathbf{D}_I^{-1}"] & E \arrow[r, "\mathbf{D}_I^{-\frac{1}{2}}"] \arrow[d, "\mathbf{I}"] & E \arrow[d, "\mathbf{D}_I"] \\
F \arrow[u, "\mathbf{D}_J"] & F \arrow[l, "\mathbf{D}_J^{-\frac{1}{2}}"] \arrow[u, "\mathbf{I}"] & \boxed{F} \arrow[l, "\mathbf{D}_J^{-\frac{1}{2}}"] \arrow[u, "\mathbf{D}_J^{-1}"] & E^* \arrow[l, "\mathbf{F}^T"] & E^* \arrow[l, "\mathbf{D}_I^{-\frac{1}{2}}"] & E^* \arrow[l, "\mathbf{D}_I^{-\frac{1}{2}}"]
\end{tikzcd}$$

$$\boxed{\text{ACP}(\mathbf{D}_I^{-1}\mathbf{F}\mathbf{D}_J^{-1}, \mathbf{D}_J, \mathbf{D}_I)} \rightarrow \mathbf{D}_J^{-1}\mathbf{F}^T\mathbf{D}_I^{-1}\mathbf{D}_I\mathbf{D}_I^{-1}\mathbf{F}\mathbf{D}_J^{-1}\mathbf{D}_J\mathbf{u} = \lambda\mathbf{u}$$

$$\mathbf{D}_J^{-1}\mathbf{F}^T\mathbf{D}_I^{-1}\mathbf{F}\mathbf{u} = \lambda\mathbf{u}$$

$$\mathbf{D}_J^{-1/2}\mathbf{F}^T\mathbf{D}_I^{-1}\mathbf{F}\mathbf{D}_J^{-1/2}\mathbf{g} = \lambda\mathbf{g} \longleftarrow$$

dove $\mathbf{g} = \mathbf{D}_J^{1/2}\mathbf{u}$, con $\mathbf{u} = \mathbf{D}_J^{-1/2}\mathbf{g}$.

Consideriamo le formule di transizione dell'AC

$$\mathbf{v}_\alpha = \frac{1}{\sqrt{\lambda_\alpha}} \mathbf{F} \mathbf{D}_J^{-1} \mathbf{u}_\alpha$$

$$\mathbf{u}_\alpha = \frac{1}{\sqrt{\lambda_\alpha}} \mathbf{F}^T \mathbf{D}_I^{-1} \mathbf{v}_\alpha$$

Le coordinate delle modalità riga e colonna si possono scrivere

$$\begin{aligned}
\Psi_\alpha &= \mathbf{D}_I^{-1} \mathbf{F} \mathbf{u}_\alpha^* = \mathbf{D}_I^{-1} \mathbf{F} \mathbf{D}_J^{-1} \mathbf{u}_\alpha = \mathbf{D}_I^{-1} \mathbf{F} \mathbf{D}_J^{-1} \frac{1}{\sqrt{\lambda_\alpha}} \mathbf{F}^T \mathbf{D}_I^{-1} \mathbf{v}_\alpha \\
&= \frac{1}{\sqrt{\lambda_\alpha}} \mathbf{D}_I^{-1} \mathbf{F} \mathbf{D}_J^{-1} \mathbf{F}^T \mathbf{D}_I^{-1} \mathbf{v}_\alpha = \sqrt{\lambda_\alpha} \mathbf{D}_I^{-1} \mathbf{v}_\alpha \\
\Phi_\alpha &= \mathbf{D}_J^{-1} \mathbf{F}^T \mathbf{v}_\alpha^* = \mathbf{D}_J^{-1} \mathbf{F}^T \mathbf{D}_I^{-1} \mathbf{v}_\alpha = \mathbf{D}_J^{-1} \mathbf{F}^T \mathbf{D}_I^{-1} \frac{1}{\sqrt{\lambda_\alpha}} \mathbf{F} \mathbf{D}_J^{-1} \mathbf{u}_\alpha \\
&= \frac{1}{\sqrt{\lambda_\alpha}} \mathbf{D}_J^{-1} \mathbf{F}^T \mathbf{D}_I^{-1} \mathbf{F} \mathbf{D}_J^{-1} \mathbf{u}_\alpha = \sqrt{\lambda_\alpha} \mathbf{D}_J^{-1} \mathbf{u}_\alpha
\end{aligned}$$

Consideriamo le coordinate riga $\Psi_\alpha = \mathbf{D}_I^{-1}\mathbf{F}\mathbf{u}^*_\alpha$ con $\mathbf{Q} = \mathbf{D}_J^{-1}$

$$\begin{aligned}
\mathbf{D}_I^{-1}\mathbf{F}\mathbf{u}^*_\alpha &= \sqrt{\lambda_\alpha}\mathbf{D}_I^{-1}\mathbf{v}_\alpha \\
\mathbf{D}_I^{-1}\mathbf{F}\mathbf{u}^*_\alpha\mathbf{u}^{*T}_\alpha\mathbf{Q}^{-1} &= \sqrt{\lambda_\alpha}\mathbf{D}_I^{-1}\mathbf{v}_\alpha\mathbf{u}^{*T}_\alpha\mathbf{Q}^{-1} \\
\mathbf{D}_I^{-1}\mathbf{F}\sum_\alpha \mathbf{u}^*_\alpha\mathbf{u}^{*T}_\alpha\mathbf{Q}^{-1} &= \sum_\alpha \sqrt{\lambda_\alpha}\mathbf{D}_I^{-1}\mathbf{v}_\alpha\mathbf{u}^{*T}_\alpha\mathbf{Q}^{-1} \\
\mathbf{D}_I^{-1}\mathbf{F}\sum_\alpha \mathbf{u}^*_\alpha\mathbf{u}^{*T}_\alpha\mathbf{D}_J &= \sum_\alpha \sqrt{\lambda_\alpha}\mathbf{D}_I^{-1}\mathbf{v}_\alpha\mathbf{u}^{*T}_\alpha\mathbf{D}_J \\
\mathbf{D}_I^{-1}\mathbf{F} &= \sum_\alpha \Psi_\alpha\mathbf{u}^{*T}_\alpha\mathbf{D}_J \\
\mathbf{D}_I^{-1}\mathbf{F} &= \sum_\alpha \frac{1}{\sqrt{\lambda_\alpha}}\Psi_\alpha\Phi_\alpha^T\mathbf{D}_J
\end{aligned}$$

dove $\sum_\alpha \mathbf{u}^*_\alpha\mathbf{u}^{*T}_\alpha\mathbf{D}_J = \mathbf{I}$ e $\Phi_\alpha = \sqrt{\lambda_\alpha}\mathbf{u}^*_\alpha$

L'elemento generico della matrice $\mathbf{D}_I^{-1}\mathbf{F}$ si può allora scrivere

$$\frac{f_{ij}}{f_{i.}} = f_{.j} \sum_{\alpha} \frac{\Psi_{i\alpha}\Phi_{j\alpha}}{\sqrt{\lambda_\alpha}}$$

e quindi

$$f_{ij} = f_{i.} f_{.j} \sum_{\alpha} \frac{\Psi_{i\alpha}\Phi_{j\alpha}}{\sqrt{\lambda_\alpha}} = f_{i.} f_{.j} \left(\frac{\Psi_{i1}\Phi_{j1}}{\sqrt{\lambda_1}} + \sum_{\alpha>1} \frac{\Psi_{i\alpha}\Phi_{j\alpha}}{\sqrt{\lambda_\alpha}} \right)$$

Si ricordi che $\lambda_1 = 1$ (autovalore banale) e che $\Psi_{i1} = \Phi_{j1} = 1$ $\forall (i, j)$ poichè $\mathbf{u}_1$ coincide con il baricentro (profilo medio $f_{.j}$)

$$\Psi_1 = \mathbf{D}_I^{-1}\mathbf{F}\mathbf{D}_J^{-1}\mathbf{u}_1 = \mathbf{D}_I^{-1}\mathbf{F}\mathbf{D}_J^{-1} \begin{bmatrix} f_{.1} \\ \vdots \\ f_{.J} \end{bmatrix} = \mathbf{D}_I^{-1}\mathbf{F} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

Possiamo quindi ricostruire l'elemento generico della matrice **F** attraverso le coordinate delle modalità riga e colonna

$$f_{ij} = f_{i.}f_{.j}\left(1 + \sum_{\alpha>1} \frac{\Psi_{i\alpha}\Phi_{j\alpha}}{\sqrt{\lambda_\alpha}}\right)$$

Se invece consideriamo le le coordinate delle modalità riga e colonna normalizzate ad 1 $\hat{\Psi}_{i\alpha} = \Psi_{i\alpha}/\sqrt{\lambda_\alpha}$ e $\hat{\Phi}_{j\alpha} = \Phi_{j\alpha}/\sqrt{\lambda_\alpha}$, allora abbiamo che $\Psi_{i\alpha} = \hat{\Psi}_{i\alpha}\sqrt{\lambda_\alpha}$ e $\Phi_{i\alpha} = \hat{\Phi}_{j\alpha}\sqrt{\lambda_\alpha}$. Possiamo scrivere allora sostituendo

$$\begin{aligned} f_{ij} &= f_{i.}f_{.j}\left(1 + \sum_{\alpha>1} \lambda_\alpha \frac{\hat{\Psi}_{i\alpha}\hat{\Phi}_{j\alpha}}{\sqrt{\lambda_\alpha}}\right) \\ &= f_{i.}f_{.j}\left(1 + \sum_{\alpha>1} \sqrt{\lambda_\alpha}\hat{\Psi}_{i\alpha}\hat{\Phi}_{j\alpha}\right) \end{aligned}$$