

**Université AMADOU MAHTAR MBOW**  
**MASTER 1 MISID**  
**PROBABILITE ET STATISTIQUE POUR IA**

**Partie 1 : Méthodes Factorielles**

**Analyse en Composantes Principales**

**Algorithme**

1. Centrer le tableau  $(X_{<n,p>})$  :  $X' = X - g$  et  $g^j = \frac{\sum_{i=1}^N P_i \cdot x_{ij}}{\sum_{i=1}^N P_i}$ ;  $j = 1 \dots p$ ;  $i = 1 \dots N$  &  $p_i = \frac{1}{N}$
2. Calculer la matrice variance –covariance :  $V = \frac{1}{N} X^t \cdot X$
3. Déterminer la métrique  $M = \begin{cases} I \\ D \frac{1}{\sigma_j^2} \end{cases}$  Données homogènes/ hétérogènes.
4. Recherche des axes principaux  $U_k$  de la matrice  $(VM)$ 
  - ✚ Calculer les valeurs propres :  $\det(VM - \lambda I) = 0$
  - ✚ Trier les valeurs propres par ordre décroissant :  $\lambda_1 > \lambda_2 > \dots > \lambda_p$ .
5. Calculer la qualité de représentation :  $Q_j = \frac{\sum_{i=1}^J \lambda_i}{\sum_{i=1}^p \lambda_i} \geq 80\%$ .
6. Calculer les vecteurs propres  $U_k$  de la matrice  $(VM)$  en utilisant la formule :  $VMU_k = \lambda_k U_k$ .
7. Calculer les composantes principales :  $C_k^i = \langle X_i, U_k \rangle_M = X_i^t M U_k$  et  $C_k = X M U_k$ .
8. Représenter graphiquement les individus dans l'espace réduit en utilisant les composantes principales.
9. Les contributions aux inerties :

✚ Part d'inertie de  $X_i$  prise en compte par l'axe  $U_k$  :  $\cos^2(\theta_{ik}) = \frac{(C_k^i)^2}{\|X_i\|_M^2}$ .

✚ Contribution relative de l'individu  $X_i$  à l'inertie expliquée de l'axe  $U_k$  :

$$\rho_{ik} = \frac{P_i \cdot (C_k^i)^2}{\sum_{i=1}^n P_i \cdot (C_k^i)^2} = \frac{P_i \cdot (C_k^i)^2}{Var(C_k)} = \frac{P_i \cdot (C_k^i)^2}{\lambda_k}$$

10. Représentation des variables à l'aide du coefficient de corrélation :

$$Cor(X^j, C_k) = \frac{Cov(X^j, C_k)}{\sigma_{X^j} \cdot \sigma_{C_k}} = \frac{\sum_{i=1}^N P_i \cdot X_i^j \cdot C_k^i}{\sigma_{X^j} \cdot \sqrt{\lambda_k}} = \frac{\frac{1}{N} (X^j)^t \cdot C_k}{\sigma_{X^j} \cdot \sqrt{\lambda_k}}$$

## AFC & ACM

### Algorithme

1. Tableau [Variable/Variable]  $\rightarrow$  deux tableaux de profils [Individus/ Variables]

2. Application de deux ACP  $\rightarrow N(I)$

$\downarrow N(J)$

3. Les valeurs propres significatives du nuage  $\lambda_k \in ]0,1[$

$$N(I) \mapsto \lambda_k, U_k$$

$$N(J) \mapsto \lambda_k, V_k$$

4. Calculer les composantes principales :  $N(I) \mapsto C_k = F_J^I . D_{\substack{1 \\ f, j}} U_k$  &  $N(J) \mapsto d_k = F_I^J . D_{\substack{1 \\ f, i}} V_k$

$$C_k^i = \frac{1}{\sqrt{\lambda_K}} \sum_{j=1}^P F_J^I . d_k^j \quad \& \quad d_k^j = \frac{1}{\sqrt{\lambda_K}} \sum_{i=1}^n F_I^J . C_k^i$$

5. Les formules de transitions :

## Partie 2 : Pratique des méthodes factorielles sur Python