**Université Rennes II**
**Haute Bretagne**
Laboratoire de Statistique

**Universität Dortmund**
Fachbereich Statistik

N° attribué par la bibliothèque :

## Thèse / Dissertation

pour obtenir le grade de

Docteur
de l'Université Rennes II
Discipline : Statistique

Doktor
der Naturwissenschaften

présentée et soutenue publiquement
par

**Karin SAHMER**

le 30 octobre 2006

# Propriétés et extensions de la classification de variables autour de composantes latentes. Application en évaluation sensorielle.

# Eigenschaften und Erweiterungen der Methode CLV zum Clustern von Variablen. Anwendungen in der Sensometrie.

**Jury / Prüfungskommission :**

**Jacques BENASSENI**, professeur, Université Rennes II
(Président / Vorsitzender)
**Pierre CAZES**, professeur, Université Paris Dauphine
(Rapporteur / Gutachter)
**El Mostafa QANNARI**, professeur, ENITIAA / INRA Nantes
(Rapporteur / Gutachter)
**Joachim KUNERT**, professeur, Universität Dortmund
(Directeur de thèse et rapporteur / Betreuer und Gutachter)
**Michel CARBON**, professeur, Université Rennes II
(Directeur de thèse / Betreuer)
**Claus WEIHS**, professeur, Universität Dortmund

# Table des matières

# Zusammenfassung

Clustermethoden bieten eine Möglichkeit, Einblick in die Struktur von Daten zu gewinnen. Normalerweise dienen sie dazu, Gruppen von Beobachtungen zu bilden. In der Regel basieren sie auf der Analyse einer Distanzmatrix, zum Beispiel der euklidischen Distanz zwischen den Beobachtungen. Wenn sehr viele Variablen erhoben werden, kann es aber auch Ziel sein, Gruppen von Variablen zu bilden. Dies ist zum Beispiel der Fall in der sensorischen Profilprüfung, bei der verschiedene Produkte gemäß unterschiedlicher sensorischer Deskriptoren von mehreren Prüfpersonen beurteilt werden. Um eine umfassende sensorische Beschreibung der Produkte zu gewährleisten, sind zunächst möglichst viele Deskriptoren in die Studie einzubeziehen. Ein Clustern der Deskriptoren kann anschließend genutzt werden, um die Anzahl der Deskriptoren in zukünftigen Studien zu reduzieren, indem aus jeder Gruppe nur ein oder zwei Deskriptoren verwendet werden.

Für das Clustern von Variablen ist es möglich, eine Distanz zwischen Variablen zu definieren und anschließend das Clustern anhand der resultierenden Distanzmatrix mit denselben Methoden wie beim Clustern von Beobachtungen durchzuführen. Es gibt aber auch Methoden, die direkt zum Clustern von Variablen entwickelt wurden. Zu nennen ist hier die Prozedur Varclus des Programmpaketes SAS. Als Alternative bietet sich die Methode CLV an, die von Vigneau und Qannari [22] und Vigneau et al. [23] entwickelt wurde. Die vorliegende Arbeit beinhaltet eine statistische Analyse der Methode CLV, um diese Methode besser zu verstehen, ihre Leistungsfähigkeit zu beurteilen und sie mit anderen Methoden zum Clustern von Variablen zu vergleichen.

Die Methode CLV verbindet ein agglomeratives hierarchisches Clustern mit einem partitionierenden Algorithmus. Seien $x_1, \ldots, x_p$ die $p$ zu clusternden Variablen. Wir bezeichnen mit $\mathbf{x}^{(k)}$ den Vektor, der sich aus den Variablen zusammensetzt, die zur Gruppe $G^{(k)}$ gehören, und mit $\mathbf{\Sigma}^{(k)}$ die Kovarianzmatrix dieser Variablen. In jeder Gruppe $G^{(k)}$, $k = 1, \ldots, K$ (wobei $K$ die Anzahl der Gruppen bezeichnet), wird eine latente Variable $c^{(k)}$ definiert, die eine Linearkombination $\mathbf{d}^{(k)\prime}\mathbf{x}^{(k)}$ der Variablen dieser Gruppe ist. Es wird die Maximierung des Kriteriums

$$T^{(K)} = \sum_{k=1}^{K} \sum_{j \in G^{(k)}} \mathrm{Cov}^2\left(x_j, c^{(k)}\right) = \sum_{k=1}^{K} \mathbf{d}^{(k)\prime} \mathbf{\Sigma}^{(k)2} \mathbf{d}^{(k)}$$

unter der Nebenbedingung

$$\mathrm{Var}\left(c^{(k)}\right) = \mathbf{d}^{(k)\prime}\mathbf{\Sigma}^{(k)}\mathbf{d}^{(k)} = 1$$

angestrebt. Für eine gegebene Partition $G^{(1)}, \ldots, G^{(K)}$ wird $T^{(K)}$ maximiert, wenn in jeder Gruppe als Vektor $\mathbf{d}^{(k)}$ der zum grössten Eigenwert $\lambda_1^{(k)}$ von $\mathbf{\Sigma}^{(k)}$ gehörende Eigenvektor gewählt wird. Es ergibt sich

$$T^{(K)} = \sum_{k=1}^{K} \lambda_1^{(k)}.$$

Wenn zwei Gruppen $G^{(k)}$ und $G^{(l)}$ vereinigt werden, verkleinert sich das Kriterium $T$ um

$$\Delta T = \lambda_1^{(k)} + \lambda_1^{(l)} - \lambda_1^{(G^{(k)} \cup G^{(l)})},$$

wobei $\lambda_1^{(G^{(k)} \cup G^{(l)})}$ der grösste Eigenwert der neu gebildeten Gruppe ist. Zu Beginn des Algorithmus bildet jede Variable eine eigene Gruppe. In jedem Schritt werden die beiden Gruppen vereinigt, die das kleinste $\Delta T$ erzeugen, bis schließlich alle Variablen in einer einzigen Gruppe zusammengefasst sind. Vigneau und Qannari [22] schlagen vor, die Entwicklung des Kriteriums $\Delta T$ als Entscheidungshilfe für die Anzahl der Gruppen zu nutzen. Die Entscheidung für $K$ Gruppen wird getroffen, wenn $\Delta T$ beim Übergang von $K$ auf $K-1$ Gruppen bedeutend größer ist als in den vorangegangenen Schritten. Die sich aus dem hierarchischen Clustern ergebende Partition in $K$ Gruppen wird schließlich durch einen partitionierenden Algorithmus verbessert. Dieser Algorithmus verläuft in zwei Schritten:

1. Jede Variable wird der Gruppe zugeordnet, mit deren latenter Variable ihre quadrierte Kovarianz am grössten ist.
2. Neuberechnung der latenten Variablen in jeder Gruppe.

Diese beiden Schritte werden so lange wiederholt, bis keine Variable mehr die Gruppe wechselt.

Für die Analyse der Methode CLV wird zunächst ein statistisches Modell formuliert. Die Methode CLV steht in enger Beziehung zur Hauptkomponentenanalyse, da die in jeder Gruppe definierte latente Variable proportionell zur ersten Hauptkomponente ist. Dennoch wurde ein faktorenanalytisches Modell bevorzugt, in dem sich jede Variable aus einem durch das Modell erklärten Term und einem Fehlerterm zusammensetzt und die Fehlerterme verschiedener Variablen unabhängig voneinander sind. Es ist möglich, dieses Modell der Analyse der Methode CLV zugrunde zu legen, weil die erste Hauptkomponente und der Faktor miteinander korreliert sind, falls ein faktorenanalytisches Modell mit einem Faktor vorliegt. Das formulierte Modell geht von der Existenz von $K$ Variablen-Gruppen aus. Die Variablen in jeder Gruppe wiederum folgen einem faktorenanalytischen Modell mit einem Faktor. Die sich aus dem Modell ergebende Kovarianzmatrix lässt sich

schreiben als

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}^{(1)} & \mathbf{\Sigma}^{(12)} & \dots & \mathbf{\Sigma}^{(1K)} \\ \mathbf{\Sigma}^{(12)'} & \mathbf{\Sigma}^{(2)} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{\Sigma}^{(K-1,K)} \\ \mathbf{\Sigma}^{(1K)'} & \dots & \mathbf{\Sigma}^{(K-1,K)'} & \mathbf{\Sigma}^{(K)} \end{pmatrix}$$

wobei

$$\mathbf{\Sigma}^{(k)} = \mathbf{b}^{(k)}\mathbf{b}^{(k)'} + \psi^{(k)}\mathbf{I}$$

und

$$\mathbf{\Sigma}^{(kl)} = \phi^{(kl)}\mathbf{b}^{(k)}\mathbf{b}^{(l)'}.$$

Der Vektor $\mathbf{b}^{(k)}$ enthält die Ladungen der Variablen der Gruppe $G^{(k)}$ auf den Faktor der Gruppe, $\psi^{(k)}$ ist die Fehlervarianz der Variablen der Gruppe $G^{(k)}$, und $\phi^{(kl)}$ bezeichnet die Korrelation des Faktors der Gruppe $G^{(k)}$ mit dem Faktor der Gruppe $G^{(l)}$. Es wird gezeigt, dass dieses Modell sensorische Profildaten angemessen darstellt, und zwar sowohl bei einer Prüfung mit fest vorgegebenen Deskriptoren als auch beim Free-Choice-Profiling.

Um den hierarchischen Algorithmus unter dem vorgeschlagenen Modell zu analysieren, werden zunächst Gruppen mit unkorrelierten Faktoren ($\phi^{(kl)} = 0$ für alle $k \neq l$) betrachtet. Für diesen Fall ergibt sich, dass $\Delta T$ gleich der Fehlervarianz ist, wenn zwei Teilmengen derselben Gruppe $G^{(k)}$ vereinigt werden, also

$$\Delta T = \psi^{(k)}.$$

Wenn zwei verschiedene Gruppen $G^{(k)}$ und $G^{(l)}$ vereinigt werden, erhält man

$$\Delta T = \min(\mathbf{b}^{(k)'}\mathbf{b}^{(k)} + \psi^{(k)}, \mathbf{b}^{(l)'}\mathbf{b}^{(l)} + \psi^{(l)}).$$

Hieraus ergibt sich, dass der Algorithmus zunächst Variablen derselben Gruppe vereinigt und die $K$ Gruppen korrekt bildet, bevor unterschiedliche Gruppen zusammengefasst werden. Voraussetzung hierfür ist lediglich, dass $\psi^{(k)} < \mathbf{b}^{(l)'}\mathbf{b}^{(l)} + \psi^{(l)}$ für alle $k, l = 1, \dots, K$.

Auch im Fall korrelierter Faktoren ($\phi^{(kl)} \neq 0$) kann das Kriterium $\Delta T$ mithilfe der Parameter des faktorenanalytischen Modells ausgedrückt werden. Im Spezialfall gleicher Fehlervarianzen $\psi$ in den verschiedenen Gruppen ergibt sich

$$\psi \leq \Delta T \leq \min(\mathbf{b}^{(k)'}\mathbf{b}^{(k)} + \psi, \mathbf{b}^{(l)'}\mathbf{b}^{(l)} + \psi).$$

Die Extremwerte dieser Abschätzung entsprechen den oben beschriebenen Ergebnissen: Die untere Schranke wird angenommen, wenn $|\phi^{(kl)}| = 1$ und somit zwei Teilmengen derselben Gruppe vereinigt werden. Die obere Schranke wird

angenommen, wenn $\phi^{(kl)} = 0$ und damit zwei Gruppen mit unkorrelierten Faktoren zusammengelegt werden.

Die Analyse des partitionierenden Algorithmus unterstreicht die Bedeutung einer geeigneten Initialisierung. Wenn bei Gruppen mit unkorrelierten Faktoren der Algorithmus ausgehend von einer guten Partition gestartet wird, genügt ein einziger Durchlauf, um die korrekte Gruppierung zu finden. Weicht die Anfangs-Partition jedoch zu stark von der wahren Partition ab, wird die korrekte Gruppierung überhaupt nicht gefunden.

Normalerweise ist die Populations-Kovarianzmatrix nicht bekannt. Das Clustern erfolgt basierend auf der empirischen Kovarianzmatrix. Die Werte der Cluster-Kriterien werden in diesem Fall mit $\hat{T}$ und $\Delta\hat{T}$ bezeichnet, da es sich um Schätzungen handelt. Zur Untersuchung der Eigenschaften der Methode CLV in diesem Fall wird zunächst die Verteilung von $\hat{T}$ und $\Delta\hat{T}$ betrachtet. Es stellt sich heraus, dass schon die Bestimmung des Erwartungswertes problematisch ist. $\Delta\hat{T}$ ist ein verzerrter Schätzer von $\Delta T$. Die Verzerrung lässt sich zudem nicht hinreichend genau bestimmen. Unter diesem Gesichtspunkt und weil außerdem die Realisierungen von $\Delta\hat{T}$ von allen vorangegangenen Schritten des hierarchischen Algorithmus abhängen, ist es sinnvoller, die Leistungsfähigkeit der Methode CLV anhand einer Simulationsstudie zu untersuchen.

In dieser Simulationsstudie wird die Methode CLV mit anderen Methoden zum Clustern von Variablen verglichen. Es zeigt sich, dass die Leistungsfähigkeit der Methode CLV mit derjenigen von drei weiteren Methoden vergleichbar ist. Dabei handelt es sich um den Ward-Algorithmus auf der Basis von $(1 - r^2)$ als Abstandsmaß (wobei $r$ die Korrelation zwischen zwei Variablen bezeichnet), des weiteren die Prozedur Varclus des Programmpaketes SAS sowie eine Gruppierung der Variablen basierend auf den Ergebnissen einer Hauptkomponentenanalyse mit Varimax-Rotation. Es ist bemerken, dass bei all diesen Methoden in kleinen Stichproben (wie sie bei sensorischen Analysen häufig sind) nur mittelmäßige Ergebnisse zu erwarten sind.

Abschließend werden zwei Verfahren vorgeschlagen, die eine automatische Bestimmung der Gruppenanzahl erlauben. Diese beiden Verfahren werden mittels einer Simulationsstudie verglichen. Auch hier ist zu beachten, dass die Leistungsfähigkeit bei kleinen Stichproben nicht zufriedenstellend ist. Aufgrund der Ergebnisse der Simulationsstudien kann man raten, bei sensorischen Profilprüfungen möglichst viele Produkte zu verwenden, falls die Analyse auch zur Auswahl von Deskriptoren für zukünftige Studien dienen soll.

Ein Vorteil der Methode CLV im Vergleich zu anderen Methoden zum Clustern von Variablen besteht in der Möglichkeit, externe Variablen in die Analyse einzube-

ziehen. Nachdem in dieser Arbeit die Gleichwertigkeit der Methode CLV mit bekannten Methoden festgestellt wurde, wäre eine sinnvolle Erweiterung eine statistische Analyse der externe Variablen zulassenden Optionen der Methode CLV.

# Remerciements

Mes remerciements vont tout d'abord à mes directeurs de thèse de l'université de Rennes 2, le Professeur Michel Carbon, et de l'université de Dortmund, le Professeur Joachim Kunert. Je remercie ensuite l'équipe du laboratoire de sensométrie et de chimiométrie de l'ENITIAA / INRA Nantes qui m'a accueillie. Tout particulièrement, je remercie les Professeurs El Mostafa Qannari et Evelyne Vigneau pour leur encadrement, et Mohamed Hanafi, Stéphanie Ledauphin et Véronique Cariou pour les multiples échanges et conseils.

# Notation

Dans tout le texte, les notations suivantes sont utilisées :
$p$ : le nombre de variables.
$K$ : le nombre de groupes.
$G^{(1)}, G^{(2)}, \ldots, G^{(K)}$ : les groupes de variables.
$p^{(k)}$ : le nombre de variables du groupe $G^{(k)}$.
$n$ : le nombre d'individus.

$x_j$, $j = 1, \ldots, p$ : la $j^{ème}$ variable aléatoire.
$\mathbf{x} = (x_1, \ldots, x_p)'$ : le vecteur aléatoire.
$x_j^{(k)}$ $\left(k = 1, \ldots, K,\ j = 1, \ldots, p^{(k)}\right)$ : la $j^{ème}$ variable du groupe $G^{(k)}$.
$\mathbf{x}^{(k)} = \left(x_1^{(k)}, \ldots, x_{p^{(k)}}^{(k)}\right)'$ : le vecteur aléatoire des variables du groupe $G^{(k)}$.

$\mathbf{X}$ $(n \times p)$ : la matrice des données observées.
$\mathbf{x}_j$ : la $j^{ème}$ colonne de $\mathbf{X}$ (les $n$ réalisations de $x_j$).
$\mathbf{X}^{(k)}$ $\left(n \times p^{(k)}\right)$ : les colonnes de $\mathbf{X}$ qui correspondent aux variables du groupe $G^{(k)}$.

$\mathbf{\Sigma}$ : la matrice de variance-covariance.
$\lambda_1$ : la plus grande valeur propre de $\mathbf{\Sigma}$.
$\mathbf{\Sigma}^{(k)}$ : la matrice de variance-covariance des variables du groupe $G^{(k)}$.
$\lambda_1^{(k)}$ : la plus grande valeur propre de la matrice $\mathbf{\Sigma}^{(k)}$.
$\mathbf{\Sigma}^{(kl)}$ : la matrice des covariances des variables du groupe $G^{(k)}$ avec celles du groupe $G^{(l)}$.

$\mathbf{S}$ : la matrice de variance-covariance empirique.
$l_1$ : la plus grande valeur propre de $\mathbf{S}$.
$S^{(k)}$ : la matrice de variance-covariance empirique des variables du groupe $G^{(k)}$.
$l_1^{(k)}$ : la plus grande valeur propre de la matrice $\mathbf{S}^{(k)}$.
$S^{(kl)}$ : la matrice des covariances empiriques des variables du groupe $G^{(k)}$ avec celles du groupe $G^{(l)}$.

$\mathbf{I_p}$ (ou $\mathbf{I}$ s'il n'y a pas d'ambiguïté sur la dimension de $\mathbf{I}$) : matrice identité.
$\mathbf{1_p}$ (ou $\mathbf{1}$ s'il n'y a pas d'ambiguïté sur la dimension de $\mathbf{1}$) : vecteur formé de 1.
$\mathbf{0_p}$ (ou $\mathbf{0}$ s'il n'y a pas d'ambiguïté sur la dimension de $\mathbf{0}$) : vecteur formé de 0.

$\mathbf{0_{p\times q}}$ (ou $\mathbf{0}$ s'il n'y a pas d'ambiguïté sur la dimension de $\mathbf{0}$) : matrice formé de 0.

Paramètres du modèle factoriel :
$\mathbf{b}$ : le vecteur des saturations.
$\mathbf{b}^{(k)}$ : le vecteur des saturations des variables du groupe $G^{(k)}$.
$\psi$ : la variance de l'erreur.
$\psi^{(k)}$ : la variance de l'erreur des variables du groupe $G^{(k)}$.
$\xi^{(k)}$ : la variable latente du groupe $G^{(k)}$.
$\phi^{(kl)}$ : la corrélation entre $\xi^{(k)}$ et $\xi^{(l)}$.

# Chapitre 1

# Introduction

La classification est une méthode d'investigation de la structure des données. Généralement, elle est utilisée pour une classification des individus. Cependant, dans des études où beaucoup de variables sont évaluées, l'intérêt peut porter sur la classification de variables. Ceci est, par exemple, le cas en analyse sensorielle où une classification de variables peut servir à déterminer des groupes de descripteurs reflétant les mêmes sensations. Par la suite, l'utilisateur peut se servir des résultats de la classification pour sélectionner une liste réduite de descripteurs à raison d'un ou plusieurs descripteurs par groupe.

Pour effectuer la classification de variables, il y a plusieurs approches possibles. D'abord, il est possible d'utiliser une approche similaire à celle poursuivie dans le cadre de la classification d'individus en proposant un indice de dissimilarité entre variables. Une telle dissimilarité peut, par exemple, être $(1 - r^2)$ où $r$ est le coefficient de corrélation. Il est aussi possible de baser un groupement des variables sur les résultats d'une analyse en composantes principales en considérant les coefficients des variables sur les axes principaux. Il y a enfin des méthodes de classification qui sont spécialement conçues pour le groupement de variables. En particulier, nous pouvons citer la procédure Varclus qui est intégrée dans le logiciel SAS. Une alternative est proposée par Vigneau et Qannari [22] et Vigneau *et al.* [23] avec la méthode de classification de variables autour de composantes latentes (CLV). Cette méthode est relativement simple à programmer. Elle a pour but de former des groupes de variables ; chaque groupe étant représenté par une variable latente. La procédure comprend une classification hiérarchique ascendante suivie d'un algorithme de partitionnement. Elle permet plusieurs options en utilisant des critères différents. Pour la classification de descripteurs sensoriels, il est approprié d'utiliser l'option de la méthode CLV qui consiste à regrouper dans un même groupe les variables redondantes sans tenir compte du signe de corrélation. Dans cette option, la variable latente du groupe est la première composante principale des variables de ce groupe.

L'objectif de la thèse est d'entreprendre une analyse statistique de la méthode CLV afin de mieux la comprendre, d'évaluer sa pertinence et de comparer sa performance avec celle d'autres méthodes. Cette analyse est faite avec une attention particulière pour l'application aux données issues d'un profil sensoriel puisque la méthode CLV a été développée principalement dans ce cadre (voir Vigneau et Qannari [22]). Afin de répondre à l'objectif que nous nous sommes fixés, un modèle statistique est proposé pour la méthode CLV. Bien que la méthode CLV soit, à la base, conceptuellement proche de l'analyse en composantes principales (ACP), il nous a semblé plus judicieux d'adopter un modèle basé sur l'analyse en facteurs communs et spécifiques (AFCS). Après une brève description de l'ACP et l'AFCS dans le chapitre 2, où nous soulignons en particulier les liens entre l'ACP et l'AFCS dans le cas d'un seul facteur, nous formulons, dans le chapitre 3, un modèle factoriel qui est particulièrement approprié pour des données issues d'un profil sensoriel. Une étude de cas sert à illustrer le modèle et à donner des indications sur les valeurs des paramètres du modèle.

Dans le chapitre 4, la méthode CLV est brièvement décrite et ces propriétés théoriques sont analysées sur la base du modèle postulé. Les critères de classification ($T$ et $\Delta T$) sont exprimés en fonction des paramètres du modèle. Cette analyse nécessite la détermination des valeurs propres d'une matrice partitionnée. Le développement original concernant cette partie est reporté en annexe afin de ne pas encombrer le lecteur de détails techniques.

En pratique, la matrice de variance-covariance théorique et, donc, les critères $T$ et $\Delta T$ ne sont pas connus. La classification est basée sur la matrice de variance-covariance empirique pour obtenir des estimateurs $\hat{T}$ et $\Delta\hat{T}$. Pour analyser les propriétés de cette classification, il faut d'abord connaître la distribution de $\hat{T}$ et $\Delta\hat{T}$. Il s'avère qu'il n'est pas possible d'approcher avec suffisamment d'exactitude les espérances mathématiques des estimateurs (voir le chapitre 5). Pour cette raison, une évaluation de la performance de la méthode CLV est étudiée au moyen d'une étude de simulations. Celle-ci permet, en plus, la comparaison avec d'autres méthodes. Elle est décrite dans le chapitre 6.

Dans la procédure CLV, il est préconisé de déterminer le nombre de groupes par un examen visuel du graphique indiquant l'évolution du critère $\Delta T$. Dans le chapitre 7, une procédure de permutations est proposée pour obtenir une décision automatique. Cette procédure est comparée à une méthode développée par Sahmer *et al.* [18].

Finalement, dans le chapitre 8, les diverses méthodes de classification de variables qui se sont révelées les plus pertinentes à l'issue de l'étude de simulations sont illustrées à l'aide d'un ensemble de données.

# Chapitre 2

# Analyse en composantes principales et analyse en facteurs

La classification de variables autour de composantes latentes (CLV) utilise la première composante principale comme variable latente du groupe. Il est donc approprié d'analyser la méthode CLV dans le cadre de l'analyse en composantes principales (ACP). Cependant, pour une analyse statistique, le modèle de l'analyse en facteurs communs et spécifiques (AFCS) présente des avantages. Contrairement à l'ACP, le modèle de l'AFCS suppose que le vecteur observé est formé d'une partie systématique et d'une partie d'erreur (Anderson [1]). Pour cette raison, Bentler et Kano [2] estiment que le modèle de l'AFCS est presque toujours à préférer au modèle de l'ACP. Quand le modèle factoriel avec un facteur est vrai, il y a une correspondance entre les deux modèles (voir le paragraphe 2.3). Il semble donc possible d'analyser la méthode CLV, qui est à l'origine basée sur l'ACP en considérant un modèle factoriel relevant de l'analyse en facteurs communs et spécifiques. De plus, le modèle factoriel semble très approprié pour la classification de variables autour de composantes latentes. En effet, chaque groupe de variables reflète une variable latente, et la variance non expliquée par la variable latente est la variance de l'erreur. Dans les paragraphes suivants, les modèles de l'ACP et de l'AFCS sont brièvement décrits et comparés. Pour plus de détails sur les méthodes voir, par exemple, Anderson [1] et Morrison [15].

## 2.1 L'analyse en composantes principales

Le but de l'analyse en composantes principales (ACP) est de remplacer un ensemble de $p$ variables corrélées $\mathbf{x} = (x_1, \ldots, x_p)'$ par un ensemble de $m$ variables non corrélées $\mathbf{z} = (z_1, \ldots, z_m)'$, $m < p$. Le vecteur aléatoire $\mathbf{x}$ est supposé avoir une espérance égale à $\mathbf{0}$. Nous définissons d'abord $p$ variables non corrélées qui sont les $p$ composantes principales, $z_1, \ldots, z_p$. Ce sont des combinaisons linéaires des variables

d'origine :

$$z_j = \sum_{i=1}^{p} a_{ij} x_i,\ j = 1, \ldots, p$$

ou sous forme matricielle :

$$\mathbf{z} = \mathbf{A}'\mathbf{x} \tag{2.1}$$

avec $\mathbf{A}'\mathbf{A} = \mathbf{A}\mathbf{A}' = \mathbf{I}$, $\mathrm{Var}(z_1) \geq \mathrm{Var}(z_2) \geq \ldots \geq \mathrm{Var}(z_p)$ et $\mathrm{Cor}(z_i, z_j) = 0$ $(i \neq j)$. Les colonnes de $\mathbf{A}$ sont les vecteurs propres de $\mathbf{\Sigma}$, la première colonne étant associée à la plus grande valeur propre, la seconde à la deuxième valeur propre et ainsi de suite. En multipliant les deux côtés de (2.1) par $\mathbf{A}$, $\mathbf{x}$ s'écrit comme une combinaison linéaire des $p$ composantes principales :

$$\mathbf{x} = \mathbf{A}\mathbf{z}. \tag{2.2}$$

Les composantes principales $z_j,\ j = 1, \ldots, p$ sont ordonnées selon l'importance de leurs variances. Les $m$ premières composantes $(m < p)$ $\mathbf{z}_m = (z_1, \ldots, z_m)'$ sont censées restituer une part importante de la variabilité de $\mathbf{x}$. Si nous considérons les autres $(p - m)$ composantes $\mathbf{z}_{-m} = (z_{m+1}, \ldots, z_p)'$ comme non significatives, nous réduisons le nombre de variables de $p$ à $m$ en ne choisissant que $z_1, \ldots, z_m$. Ecrivant $\mathbf{A} = [\mathbf{A}_m \mathbf{A}_{-m}]$ (où $\mathbf{A}$ est décomposé de la même manière que $\mathbf{z}$), $\mathbf{x}$ peut s'écrire :

$$\mathbf{x} = \mathbf{A}_m \mathbf{z}_m + \mathbf{A}_{-m} \mathbf{z}_{-m}. \tag{2.3}$$

L'ACP représente aussi une décomposition de la matrice de variance-covariance de $\mathbf{x}$ :

$$\mathbf{\Sigma} = \mathbf{A}_m \mathbf{\Lambda}_m \mathbf{A}'_m + \mathbf{A}_{-m} \mathbf{\Lambda}_{-m} \mathbf{A}'_{-m} \tag{2.4}$$

où $\mathbf{\Lambda} = \begin{pmatrix} \mathbf{\Lambda}_m & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_{-m} \end{pmatrix}$ est la matrice diagonale ayant pour éléments diagonaux les valeurs propres de $\mathbf{\Sigma}$ rangées par ordre décroissant. Puisque les valeurs de $\mathbf{\Lambda}_{-m}$ sont plus petites que celles de $\mathbf{\Lambda}_m$, le premier terme de la décomposition (2.4) contient des valeurs plus importantes que le deuxième terme. Ainsi cette décomposition qui maximise la variance restitue aussi une part importante des covariances (Jolliffe [10]). Cependant, contrairement à la décomposition selon l'AFCS, le deuxième terme de l'expression (2.4) n'est pas une matrice diagonale.

## 2.2 L'analyse en facteurs communs et spécifiques

L'analyse en facteurs communs et spécifiques (AFCS) est basée sur un modèle statistique qui relie les variables manifestes (observables) aux variables latentes (non observables). Ces dernières sont appelées facteurs. La relation entre les variables manifestes $\mathbf{x}$ et les variables latentes $\boldsymbol{\xi}$ est une relation linéaire :

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{B}\boldsymbol{\xi} + \boldsymbol{\epsilon} \tag{2.5}$$

où $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)'$ est un vecteur de paramètres et $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_q)'$ est un vecteur aléatoire de $q$ variables latentes avec $\mathrm{E}(\boldsymbol{\xi}) = \mathbf{0}$, $\mathrm{Cov}(\boldsymbol{\xi}) = \mathrm{E}(\boldsymbol{\xi}\boldsymbol{\xi}') = \boldsymbol{\Phi}$ et $\mathrm{Var}(\xi_j) = 1$ $(j = 1, \ldots, q)$. $\mathbf{B}$ est une matrice $(p \times q)$. L'entrée $b_{ij}$ est appelée la saturation de la variable $i$ dans le facteur $j$ (Dickes [5]). Les entrées du vecteur $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_p)'$ sont les erreurs de mesures. Elles sont parfois considérées comme facteurs spécifiques. Nous imposons $\mathrm{E}(\boldsymbol{\epsilon}) = \mathbf{0}$ et $\mathrm{Cov}(\boldsymbol{\epsilon}) = \boldsymbol{\Psi}$, où $\boldsymbol{\Psi}$ est une matrice diagonale, ce qui implique que les erreurs de mesure sont considérées comme non corrélées. De plus, il est supposé que $\mathrm{Cov}(\epsilon_i, \xi_j) = 0$ $(i = 1, \ldots, p,\ j = 1, \ldots, q)$, ce qui signifie que les corrélations entre les erreurs de mesure et les facteurs sont égales à zéro. L'exigence de corrélations égales à zéro entre les erreurs d'un côté et entre les erreurs et les facteurs d'un autre côté est essentielle pour le modèle. Par contre, exiger que les facteurs aient une espérance nulle et une variance de 1 est un choix. Tout autre choix changerait simplement les paramètres $\boldsymbol{\mu}$ et $\mathbf{B}$ et non le modèle. Nous obtenons :

$$\mathrm{E}(\mathbf{x}) = \boldsymbol{\mu}$$

et

$$\boldsymbol{\Sigma} = \mathbf{B}\boldsymbol{\Phi}\mathbf{B}' + \boldsymbol{\Psi}. \tag{2.6}$$

Si $\boldsymbol{\Phi}$ est une matrice diagonale, les facteurs sont orthogonaux, sinon ils sont obliques. Pour des facteurs orthogonaux, nous obtenons :

$$\boldsymbol{\Sigma} = \mathbf{B}\mathbf{B}' + \boldsymbol{\Psi}. \tag{2.7}$$

Pour les modèles décrits dans les paragraphes suivants, $\boldsymbol{\mu}$ est supposé être égal à zéro. Ceci facilite la notation. Les résultats trouvés sont également vrais pour $\boldsymbol{\mu} \neq \mathbf{0}$.

Il faut noter qu'il y a une indétermination dans le modèle. En effet, si $\mathbf{B}$ est remplacé par $\mathbf{B}^* = \mathbf{B}\mathbf{C}'$ (où $\mathbf{C}$ est une matrice orthonormale) et $\boldsymbol{\xi}$ est remplacé par $\boldsymbol{\xi}^* = \mathbf{C}\boldsymbol{\xi}$, le modèle ne change pas, car $\mathbf{B}^*\boldsymbol{\xi}^* = \mathbf{B}\mathbf{C}'\mathbf{C}\boldsymbol{\xi} = \mathbf{B}\boldsymbol{\xi}$ et $\mathbf{B}^*\mathbf{B}^{*'} = \mathbf{B}\mathbf{C}'\mathbf{C}\mathbf{B}' = \mathbf{B}\mathbf{B}'$. Dans le cas oblique, une multiplication de $\boldsymbol{\xi}$ par une matrice $\mathbf{C}$ non singulaire (pas forcément orthonormale) et une multiplication de $\mathbf{B}$ par $\mathbf{C}^{-1}$ conduit au même constat. En pratique, cette possibilité de rotation est utilisée pour chercher des matrices de saturations qui sont facilement interprétables et qui reflètent une structure simple. Il y a plusieurs définitions de structures simples. Elles ont en commun la recherche d'une matrice $\mathbf{B}$ qui contient beaucoup de zéros et quelques valeurs importantes, mais peu de valeurs moyennes. L'idée consiste à trouver une solution dans laquelle chaque variable a une saturation non nulle dans un seul facteur ou dans très peu de facteurs.

Contrairement à l'ACP, les variables latentes $\xi_i$ ne sont pas des combinaisons linéaires des variables observées. Cela rend les estimations plus difficiles. Il y a plusieurs méthodes pour l'estimation de $\mathbf{B}$ et $\boldsymbol{\Psi}$. Maxwell [14] cite entre autres la méthode centroïde (centroid method), la méthode des facteurs principaux (principal factor method), la méthode des moindres carrés généralisés (generalized least

squares method) et la méthode de maximum de vraisemblance (maximum likelihood method). Il n'existe pas de solution analytique du maximum de vraisemblance. Il faut recourir à un algorithme itératif, par exemple l'algorithme EM (expectation - maximization).

## 2.3 Comparaison de l'ACP et l'AFCS

Une différence évidente entre l'ACP et l'AFCS est donnée par l'approche même qui est utilisée. L'ACP est une méthode géométrique de réduction de la dimension alors que l'AFCS est basée sur un modèle statistique. Cependant, si le modèle de l'AFCS est vrai, il y a quand-même des similarités entre les deux méthodes. Ici, nous nous intéressons au cas d'un seul facteur ou d'une seule composante, car la variable latente dans chacun des groupes fournis par la classification CLV, en est la première composante principale standardisée. S'il y a une équivalence entre la première composante principale et le facteur dans un modèle factoriel à un facteur, il est possible de formuler un modèle statistique qui sera utilisé pour l'analyse de la méthode CLV. En fait, une telle équivalence existe si le modèle suivant est vrai :

$$\mathbf{x} = \mathbf{b}\xi + \boldsymbol{\epsilon}$$

avec la matrice de variance-covariance :

$$\boldsymbol{\Sigma} = \mathbf{b}\mathbf{b}' + \psi\,\mathbf{I}, \tag{2.8}$$

où $\mathbf{b}$ est un vecteur et $\psi$ un réel avec $\psi > 0$. Il s'agit du modèle avec un seul facteur et des variances de l'erreur égales. Les vecteurs propres de $\mathbf{b}\mathbf{b}'$ sont aussi des vecteurs propres de $\boldsymbol{\Sigma}$. Il est facile de vérifier que, dans ce cas, $\boldsymbol{\Sigma}$ admet comme valeur propre $\lambda_1 = \mathbf{b}'\mathbf{b} + \psi$ associée au vecteur propre $\mathbf{b}$ et que les autres valeurs propres de $\boldsymbol{\Sigma}$ sont égales à $\psi$. En ACP, on choisit usuellement la contrainte $\mathbf{a}'\mathbf{a} = 1$. $\mathbf{a}_1 = \frac{1}{\sqrt{\mathbf{b}'\mathbf{b}}}\mathbf{b}$ vérifie cette contrainte. Si le modèle (2.8) est vrai, les pondérations de la première composante principale sont donc proportionelles aux saturations dans le facteur. La première composante principale est donnée par

$$z_1 = \mathbf{a}_1'\mathbf{x} = \frac{1}{\sqrt{\mathbf{b}'\mathbf{b}}}\mathbf{b}'(\mathbf{b}\xi + \boldsymbol{\epsilon}) = \sqrt{\mathbf{b}'\mathbf{b}}\xi + \frac{1}{\sqrt{\mathbf{b}'\mathbf{b}}}\mathbf{b}'\boldsymbol{\epsilon}.$$

Sa variance est égale à $\lambda_1 = \mathbf{b}'\mathbf{b} + \psi$. Sa covariance avec le facteur est donnée par

$$\mathrm{Cov}(z_1, \xi) = \mathrm{Cov}(\sqrt{\mathbf{b}'\mathbf{b}}\xi + \frac{1}{\sqrt{\mathbf{b}'\mathbf{b}}}\mathbf{b}'\boldsymbol{\epsilon}, \xi) = \sqrt{\mathbf{b}'\mathbf{b}}$$

et la corrélation par

$$\mathrm{Cor}(z_1, \xi) = \frac{\sqrt{\mathbf{b}'\mathbf{b}}}{\sqrt{\mathbf{b}'\mathbf{b} + \psi}}.$$

Lorsque $\psi$ tend vers 0, la corrélation tend vers 1.

Si les variances de l'erreur ne sont pas égales pour les différentes variables, les resultats ci-dessus ne s'appliquent pas. Cependant, Bentler et Kano [2] ont démontré un résultat asymptotique. Pour cela, ils ont considéré le modèle à un facteur :

$$\mathbf{x} = \mathbf{b}\xi + \boldsymbol{\epsilon}$$

avec

$$\boldsymbol{\Sigma} = \mathbf{b}\mathbf{b}' + \boldsymbol{\Psi}$$

où $\boldsymbol{\Psi}$ est une matrice diagonale avec les valeurs $\psi_1, \ldots, \psi_p$ sur la diagonale. Soit $\lambda_1$ la plus grande valeur propre de $\boldsymbol{\Sigma}$ et $\mathbf{a}_1$ ($\mathbf{a}_1'\mathbf{a}_1 = 1$) le vecteur propre associé. Si

$$\mathbf{b}'\mathbf{b} \to \infty \qquad \text{pour} \qquad p \to \infty$$

et s'il existe un $\psi_0 > 0$ tel que

$$\psi_i < \psi_0, \ i = 1, \ldots, p,$$

alors pour $p \to \infty$

$$\text{Cor}(\mathbf{a}_1'\mathbf{x}, \xi) \to 1$$

et

$$\sqrt{\lambda_1}\mathbf{a}_1 \to \mathbf{b}.$$

Cela signifie que la corrélation entre la première composante principale et le facteur converge vers 1, et que les pondérations de la première composante principale sont asymptotiquement proportionnelles aux saturations.

Nous pouvons conclure que même si l'ACP ne permet pas de déterminer les paramètres du modèle factoriel à un facteur, la première composante principale est fortement corrélée avec le facteur. Ceci nous permet d'utiliser la méthode CLV sur des variables qui sont supposées suivre un modèle factoriel. Comme nous allons le démontrer dans le chapitre 4, CLV permet de trouver des groupes d'un tel modèle. Dans le chapitre 3 nous décrivons un modèle factoriel approprié pour les descripteurs sensoriels.

# Chapitre 3

# Un modèle factoriel pour les données de profils sensoriels

## 3.1 Le modèle général

En analyse sensorielle, et plus précisément dans l'épreuve des profils sensoriels, des produits sont évalués selon différents descripteurs sensoriels par plusieurs juges. Souvent, il y a une redondance entre les descripteurs, c'est-à-dire qu'un ou plusieurs descripteurs mesurent la même sensation. Nous pouvons considérer cette sensation comme une variable latente qui est refletée par les descripteurs. Par exemple, les descripteurs "pimenté", "piquant", "épicé" peuvent refléter une même variable latente et former un groupe. Nous formulons un modèle statistique qui décrit ces redondances. Dans ce qui suit, l'indice $i = 1, \ldots, I$ est utilisé pour les produits, l'indice $j = 1, \ldots, J$ pour les juges et l'indice $l$ pour les descripteurs. $p^{(k)}$ est le nombre de descripteurs dans le groupe $G^{(k)}$ et $p := \sum_{k=1}^{K} p^{(k)}$.

**Modèle pour les descripteurs**

Nous désignons la variable latente du groupe $G^{(k)}$ par $\xi^{(k)}$. La valeur de la variable aléatoire qui représente le $l^{ème}$ descripteur du groupe $G^{(k)}$ pour le $i^{ème}$ produit est donnée par :

$$y_{il}^{(k)} = a_l^{(k)} \xi_i^{(k)} + z_{il}^{(k)}$$

où $a_l^{(k)}$ est un paramètre fixé (mais inconnu). $z_{il}^{(k)}$ est une variable aléatoire représentant la spécificité du descripteur $l$ (le facteur spécifique). Nous exigeons que pour $i = 1, \ldots, n$, les $\xi_i^{(k)}$, les $z_{il}^{(k)}$ et, donc, les $y_{il}^{(k)}$ sont indépendantes et identiquement distribuées (*i.i.d.*). Ceci signifie en particulier que toutes les corrélations entre les variables aléatoires concernant des produits différents sont égales à zéro, et que la distribution de $y_{il}^{(k)}$ est égale à celle de $y_{i'l}^{(k)}$. De plus, nous supposons que :

$$\mathrm{E}\left(\xi_i^{(k)}\right) = 0, \qquad i = 1, \ldots, I, \quad k = 1, \ldots, K$$

et

$$\mathrm{E}\left(z_{il}^{(k)}\right) = 0, \qquad i = 1, \ldots, I, \quad k = 1, \ldots, K, \quad l = 1, \ldots, p^{(k)}.$$

Pour inclure le cas où les espérances sont différentes de zéro, il suffit d'ajouter un paramètre additif $\mu_l^{(k)}$. Cependant, dans l'analyse statistique, un tel paramètre ne joue pas de rôle puisque nous ne considérons que la matrice de variance-covariance. Nous supposons également que, pour tout $k$ :

$$\mathrm{Var}\left(\xi_i^{(k)}\right) = 1.$$

Ceci ne représente pas une réelle contrainte puisque le paramètre $a_l^{(k)}$ peut s'adapter en conséquence. La corrélation entre $\xi_i^{(k)}$ et $\xi_i^{(k')}$ sera désignée par :

$$\mathrm{Cor}\left(\xi_i^{(k)}, \xi_i^{(k')}\right) = \phi^{(kk')}.$$

Les variables $\xi_i^{(k)}$ ne sont pas corrélées avec les variables $z_{il}^{(k)}$. Nous supposons aussi que

$$\mathrm{Cor}\left(z_{il}^{(k)}, z_{il'}^{(k)}\right) = 0 \quad \text{pour} \quad l \neq l'$$

et

$$\mathrm{Cor}\left(z_{il}^{(k)}, z_{il'}^{(k')}\right) = 0 \quad \text{pour} \quad k \neq k'.$$

En d'autres termes, cela signifie que les corrélations entre les descripteurs sont entièrement expliquées par les variables latentes. Les facteurs spécifiques sont indépendants les uns des autres. La variance de $z_{il}^{(k)}$ sera désignée par :

$$\mathrm{Var}\left(z_{il}^{(k)}\right) = {\sigma_l^{(k)}}^2, \qquad i = 1, \ldots, I, \quad k = 1, \ldots, K, \quad l = 1, \ldots, p^{(k)}.$$

La matrice de variance-covariance des $p$ descripteurs est égale à :

$$\mathbf{\Sigma}_{descr} = \begin{pmatrix} \mathbf{\Sigma}^{(1)} & \mathbf{\Sigma}^{(12)} & \ldots & \mathbf{\Sigma}^{(1K)} \\ \mathbf{\Sigma}^{(12)'} & \mathbf{\Sigma}^{(2)} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{\Sigma}^{(K-1,K)} \\ \mathbf{\Sigma}^{(1K)'} & \ldots & \mathbf{\Sigma}^{(K-1,K)'} & \mathbf{\Sigma}^{(K)} \end{pmatrix} \tag{3.1}$$

avec :

$$\mathbf{\Sigma}^{(k)} = \mathbf{a}^{(k)}\mathbf{a}^{(k)'} + \begin{pmatrix} {\sigma_1^{(k)}}^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & {\sigma_{p^{(k)}}^{(k)}}^2 \end{pmatrix}$$

et

$$\mathbf{\Sigma}^{(kk')} = \phi^{(kk')}\mathbf{a}^{(k)}\mathbf{a}^{(k')'}$$

où

$$\mathbf{a}^{(k)} = \begin{pmatrix} a_1^{(k)} \\ a_2^{(k)} \\ \vdots \\ a_{p^{(k)}}^{(k)} \end{pmatrix}.$$

Nous pouvons considérer deux cas particuliers :

1. des variances égales pour tous les descripteurs : ${\sigma_l^{(k)}}^2 = \sigma^2,\ k = 1, \ldots, K,\ l = 1, \ldots, p^{(k)}$,
2. des variances proportionnelles aux paramètres $a_l^{(k)}$ : ${\sigma_l^{(k)}}^2 = {a_l^{(k)}}^2 \sigma^2,\ k = 1, \ldots, K,\ l = 1, \ldots, p^{(k)}$.

**Modèle qui prend en compte les juges**

Les descripteurs sont mesurés par des juges. Ceci conduit à une erreur de mesure. Nous n'observons pas $y_{il}^{(k)}$, mais le jugement $x_{ijl}^{(k)}$ du juge $j$ pour $y_{il}^{(k)}$. Nous pouvons postuler le modèle :

$$x_{ijl}^{(k)} = b_j y_{il}^{(k)} + \epsilon_{ijl}^{(k)} = b_j \left( a_l^{(k)} \xi_i^{(k)} + z_{il}^{(k)} \right) + \epsilon_{ijl}^{(k)}.$$

Dans ce modèle, $b_j > 0$ désigne un facteur d'échelle propre au juge $j$. Les erreurs $\epsilon_{jl}^{(k)}$ ne sont pas corrélées avec les $\xi_i^{(k)}$ et les $z_{il}^{(k)}$, et nous avons :

$$\mathrm{E}\left(\epsilon_{ijl}^{(k)}\right) = 0 \quad \forall\, i, j, k, l,$$

$$\mathrm{Var}\left(\epsilon_{ijl}^{(k)}\right) = \psi_j, \quad i = 1, \ldots, I,\ j = 1, \ldots, J,\ k = 1, \ldots, K,\ l = 1, \ldots, p^{(k)}.$$

Les corrélations entre toutes les variables $\epsilon_{ijl}^{(k)}$ sont égales à zéro. Ceci implique que les erreurs d'un juge pour l'évaluation de deux produits différents sont indépendantes. Dans ce modèle, les effets de l'ordre de présentation des produits sont donc négligés. Par la suite, nous ne considérons plus l'indice $i$ puisque les $x_{ijl}^{(k)}$ sont distribués de manière *iid*. De nouveau, nous pouvons considérer deux cas particuliers :

1. des variances égales pour toutes les erreurs : $\psi_j = \psi,\ j = 1, \ldots, J$,
2. des variances proportionnelles aux paramètres $b_j$ : $\psi_j = b_j^2\, \psi,\ j = 1, \ldots, J$.

A partir du modèle stipulé ci-dessus, il s'ensuit :

$$\mathrm{E}\left(x_{jl}^{(k)}\right) = 0$$

et

$$\mathrm{Var}\left(x_{jl}^{(k)}\right) = b_j^2 \left( {a_l^{(k)}}^2 + {\sigma_l^{(k)}}^2 \right) + \psi_j.$$

La covariance entre deux descripteurs mesurés par le même juge est égale à :

$$\mathrm{Cov}\left(x_{jl}^{(k)}, x_{jl'}^{(k')}\right) = b_j b_j a_l^{(k)} a_{l'}^{(k')} \phi^{(kk')} = b_j^2 \,\mathrm{Cov}\left(y_l^{(k)}, y_{l'}^{(k')}\right).$$

Si les deux descripteurs appartiennent au même groupe ($k = k'$), nous obtenons :

$$\mathrm{Cov}\left(x_{jl}^{(k)}, x_{jl'}^{(k)}\right) = b_j^2 \; a_l^{(k)} a_{l'}^{(k)}.$$

La covariance associée à un même descripteur, mesuré par deux juges différents, est égale à :

$$\begin{aligned}\mathrm{Cov}\left(x_{jl}^{(k)}, x_{j'l}^{(k)}\right) &= \mathrm{Cov}\left(b_j a_l^{(k)} \xi^{(k)}, b_{j'} a_l^{(k)} \xi^{(k)}\right) + \mathrm{Cov}\left(b_j z_l^{(k)}, b_{j'} z_l^{(k)}\right) \\ &= b_j b_{j'} \left({a_l^{(k)}}^2 + {\sigma_l^{(k)}}^2\right) = b_j b_{j'} \mathrm{Var}\left(y_l^{(k)}\right).\end{aligned}$$

La covariance de deux descripteurs différents, mesurés par deux juges différents, est égale à :

$$\mathrm{Cov}\left(x_{jl}^{(k)}, x_{j'l'}^{(k')}\right) = b_j b_{j'} a_l^{(k)} a_{l'}^{(k')} \phi^{(kk')}.$$

Si les deux descripteurs appartiennent au même groupe, leur covariance est égale à :

$$\mathrm{Cov}\left(x_{jl}^{(k)}, x_{j'l'}^{(k)}\right) = b_j b_{j'} a_l^{(k)} a_{l'}^{(k)}.$$

Dans ce qui précède, nous n'avons pas supposé que les descripteurs sont les mêmes d'un juge à un autre. Le cas du profil à vocabulaire fixé (mêmes descripteurs pour tous les juges) est traité dans le paragraphe suivant. Il est aussi possible que chaque juge choisisse sa propre liste de descripteurs (profil libre). Ce cas est traité dans le paragraphe 3.3.

## 3.2 Profil conventionnel

Pour le profil conventionnel (à vocabulaire fixé), il est d'usage de considérer, pour chaque descripteur, la moyenne sur tous les juges. Pour le descripteur $l$ du groupe $G^{(k)}$, la moyenne sur tous les juges est donnée par :

$$\bar{x}_l^{(k)} = \frac{1}{J} \sum_{j=1}^{J} \left(b_j a_l^{(k)} \xi^{(k)} + b_j z_l^{(k)} + \epsilon_{jl}^{(k)}\right).$$

La variance de $\bar{x}_l^{(k)}$ est égale à :

$$\begin{aligned}\mathrm{Var}\left(\bar{x}_l^{(k)}\right) &= \frac{1}{J^2}\left(\sum_{j=1}^{J} b_j\right)^2 {a_l^{(k)}}^2 + \frac{1}{J^2}\left(\sum_{j=1}^{J} b_j\right)^2 {\sigma_l^{(k)}}^2 + \frac{1}{J^2}\sum_{j=1}^{J} \psi_j \\ &= \left(\frac{1}{J}\sum_{j=1}^{J} b_j\right)^2 \left({a_l^{(k)}}^2 + {\sigma_l^{(k)}}^2\right) + \frac{1}{J^2}\sum_{j=1}^{J} \psi_j.\end{aligned}$$

La covariance entre $\bar{x}_l^{(k)}$ et $\bar{x}_{l'}^{(k')}$ est égale à :

$$\begin{aligned}\mathrm{Cov}\left(\bar{x}_l^{(k)}, \bar{x}_{l'}^{(k')}\right) &= \mathrm{Cov}\left(\frac{1}{J}\sum_{j=1}^{J} b_j a_l^{(k)} \xi^{(k)}, \frac{1}{J}\sum_{j=1}^{J} b_j a_{l'}^{(k')} \xi^{(k')}\right) \\ &= \left(\frac{1}{J}\sum_{j=1}^{J} b_j\right)^2 a_l^{(k)} a_{l'}^{(k')} \phi^{(kk')}.\end{aligned}$$

Si le descripteur $l$ et le desripteur $l'$ appartiennent au même groupe, nous obtenons :

$$\mathrm{Cov}\left(\bar{x}_l^{(k)}, \bar{x}_{l'}^{(k)}\right) = \left(\frac{1}{J}\sum_{j=1}^{J} b_j\right)^2 a_l^{(k)} a_{l'}^{(k)}.$$

La matrice de variance-covariance de $\left(\bar{x}_1^{(1)}, \ldots, \bar{x}_{p^{(1)}}^{(1)}, \bar{x}_1^{(2)}, \ldots, \bar{x}_{p^{(K)}}^{(K)}\right)$ a donc la structure suivante :

$$\mathbf{\Sigma}_J = \left(\frac{1}{J}\sum_{j=1}^{J} b_j\right)^2 \mathbf{\Sigma}_{descr} + \frac{1}{J^2}\left(\sum_{j=1}^{J} \psi_j\right)\mathbf{I} \tag{3.2}$$

où $\mathbf{\Sigma}_{descr}$ est la matrice de variance-covariance des descripteurs (voir l'équation (3.1)). L'indice $J$ indique qu'il s'agit de la matrice qu'on obtient quand il y a $J$ juges.

**Etude asymptotique : nombre infini de juges**

La variance de l'erreur due aux juges tend, sous certaines conditions, vers zéro, si le nombre $J$ de juges tend vers l'infini. Nous exigeons pour cela que $\left(\frac{1}{J}\sum_{j=1}^{J} b_j\right)$ et $\left(\frac{1}{J}\sum_{j=1}^{J} b_j^2\right)$ sont convergents pour $J \to \infty$.

Considérons le cas de variances de l'erreur égales pour tous les juges. La variance de $\bar{x}_l^{(k)}$ est alors égale à :

$$\mathrm{Var}\left(\bar{x}_l^{(k)}\right) = \left(\frac{1}{J}\sum_{j=1}^{J} b_j\right)^2 \left(a_l^{(k)^2} + \sigma_l^{(k)^2}\right) + \frac{1}{J}\,\psi.$$

Pour $J \to \infty$, il s'ensuit :

$$\mathrm{Var}\left(\bar{x}_l^{(k)}\right) \to c^2 \left(a_l^{(k)^2} + \sigma_l^{(k)^2}\right)$$

où $c$ est la limite de $\frac{1}{J}\sum_{j=1}^{J} b_j$. Si la variance de l'erreur du juge est égale à $b_j^2\,\psi$, la variance de $\bar{x}_l^{(k)}$ est égale à :

$$\mathrm{Var}\left(\bar{x}_l^{(k)}\right) = \left(\frac{1}{J}\sum_{j=1}^{J} b_j\right)^2 \left(a_l^{(k)^2} + \sigma_l^{(k)^2}\right) + \frac{1}{J}\left(\frac{1}{J}\sum_{j=1}^{J} b_j^2\right)\,\psi.$$

Ici aussi, pour $J \to \infty$, il s'ensuit :

$$\mathrm{Var}\left(\bar{x}_l^{(k)}\right) \to c^2 \left(a_l^{(k)^2} + \sigma_l^{(k)^2}\right).$$

La matrice de variance-covariance (3.2) converge donc vers :

$$\mathbf{\Sigma} = c^2\, \mathbf{\Sigma}_{descr}.$$

Asymptotiquement, la matrice de variance-covariance des moyennes sur tous les juges est donc proportionnelle à la matrice de variance-covariance $\mathbf{\Sigma}_{descr}$ des descripteurs mesurés sans erreur. Le facteur $c^2$ n'influence pas l'analyse de la structure. Il est donc intéressant de baser l'analyse de la méthode CLV sur un modèle stipulant la matrice de variance-covariance $\mathbf{\Sigma}_{descr}$.

## 3.3 Profil libre

Considérons maintenant le cas du profil libre. Chaque juge choisit ses propres descripteurs. Un même descripteur n'est donc plus nécessairement évalué par tous les juges. De plus, il est possible que les juges ne donnent pas le même nom au même descripteur. Ici, il n'est donc ni possible ni souhaitable de baser l'analyse statistique sur la moyenne par descripteur. Nous considérons la note du $j^{ème}$ juge pour le descripteur $y_l^{(k)}$ :

$$x_{jl}^{(k)} = b_j a_l^{(k)} \xi^{(k)} + b_j z_l^{(k)} + \epsilon_{jl}^{(k)}.$$

Si le descripteur $y_l^{(k)}$ n'est évalué que par le $j^{ème}$ juge, on ne peut pas distinguer ses paramètres des paramètres du juge. Nous pouvons donc définir :

$$\tilde{\epsilon}_{jl}^{(k)} := b_j z_l^{(k)} + \epsilon_{jl}^{(k)}$$

et

$$\tilde{a}_{jl}^{(k)} := b_j a_l^{(k)}.$$

L'indice $j$ est important pour indiquer qu'il s'agit du $j^{ème}$ juge et du facteur d'échelle qui lui est associé. Cependant, il n'est pas nécessaire d'identifier ce paramètre, puisque, pour chaque descripteur $y_l^{(k)}$, il y a un seul paramètre $b_j$. Il est donc possible d'omettre l'indice $j$ et d'écrire $\tilde{\epsilon}_l^{(k)}$ à la place de $\tilde{\epsilon}_{jl}^{(k)}$ et $\tilde{a}_l^{(k)}$ à la place de $\tilde{a}_{jl}^{(k)}$ . Nous obtenons :

$$x_{jl}^{(k)} = \tilde{a}_l^{(k)} \xi^{(k)} + \tilde{\epsilon}_l^{(k)}.$$

La variance de $\tilde{\epsilon}_l^{(k)}$ est égale à :

$$\tilde{\psi}_l^{(k)} = b_j^2\, \sigma_l^{(k)^2} + \psi_j. \tag{3.3}$$

Avec :

$$\tilde{\mathbf{a}}^{(k)} = \begin{pmatrix} \tilde{a}_1^{(k)} \\ \tilde{a}_2^{(k)} \\ \vdots \\ \tilde{a}_{p^{(k)}}^{(k)} \end{pmatrix},$$

nous obtenons la matrice de variance-covariance :

$$\mathbf{\Sigma} = \begin{pmatrix} \tilde{\mathbf{a}}^{(1)}\tilde{\mathbf{a}}^{(1)\prime} & \phi^{(12)}\tilde{\mathbf{a}}^{(1)}\tilde{\mathbf{a}}^{(2)\prime} & \cdots & \phi^{(1K)}\tilde{\mathbf{a}}^{(1)}\tilde{\mathbf{a}}^{(K)\prime} \\ \phi^{(12)}\tilde{\mathbf{a}}^{(2)}\tilde{\mathbf{a}}^{(1)\prime} & \tilde{\mathbf{a}}^{(2)}\tilde{\mathbf{a}}^{(2)\prime} & \cdots & \phi^{(2K)}\tilde{\mathbf{a}}^{(2)}\tilde{\mathbf{a}}^{(K)\prime} \\ \cdots & \cdots & \cdots & \cdots \\ \phi^{(1K)}\tilde{\mathbf{a}}^{(K)}\tilde{\mathbf{a}}^{(1)\prime} & \cdots & \cdots & \tilde{\mathbf{a}}^{(K)}\tilde{\mathbf{a}}^{(K)\prime} \end{pmatrix} + \begin{pmatrix} \tilde{\psi}_1^{(1)} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \tilde{\psi}_{p^{(K)}}^{(K)} \end{pmatrix}.$$

Ici, les variances de l'erreur ne sont pas égales. Si nous voulons exiger qu'elles soient égales, nous devons exiger que les paramètres $\sigma_l^{(k)}$ et $\psi_j$ soient égaux, mais aussi que les facteurs d'échelle $b_j$ soient les mêmes pour tous les juges. Ceci résulte de la formule (3.3).

Pour le profil libre, le modèle à prendre en compte dans l'analyse de la méthode CLV est donc comparable à celui du profil conventionnel, à la différence près que les variances de l'erreur sont plus importantes.

## 3.4 Illustration

Dans ce paragraphe, le modèle décrit ci-dessus est illustré sur des données sensorielles issues d'une étude sur seize variétés de cafés. Il s'agit de l'analyse sensorielle par un des panels participant à l'étude européenne : *European sensory and consumer study* [7]. Ce panel était constitué de huit juges. Chaque juge a évalué chaque produit selon 23 descripteurs sur une échelle de 0 à 100. Pour illustrer le modèle développé ci-dessus, nous allons considérer des groupes de descripteurs. Ces groupes sont déterminés par une classification autour de composantes latentes comme décrit dans les chapitres suivants. Nous choisissons deux groupes issus de l'arbre hiérarchique, dont un groupe très homogène. Le premier groupe, disons le groupe $A$, comprend les descripteurs "goût doux-piquant", "goût brulé", "arrière goût amer" et "intensité du goût", le deuxième groupe, disons le groupe $B$, comprend les descripteurs "odeur chocolat", "odeur moisi", "odeur sucrée", "odeur chèvre" et "odeur caramel". Pour estimer les paramètres du profil conventionnel, nous considérons le tableau moyen sur tous les juges. Nous analysons les deux groupes de descripteurs séparément en nous basant sur le modèle :

$$y_l^{(A)} = \mu_l^{(A)} + s_l^{(A)}\left(a_l^{(A)}\xi^{(A)} + z_l^{(A)}\right), \qquad l = 1, \ldots, 4$$

et

$$y_l^{(B)} = \mu_l^{(B)} + s_l^{(B)}\left(a_l^{(B)}\xi^{(B)} + z_l^{(B)}\right), \qquad l = 1, \ldots, 5.$$

Dans un premier temps, la moyenne et la variance de chaque descripteur sont estimées. Ensuite, l'analyse en facteurs communs et spécifiques est effectuée sur la matrice de corrélation. Le groupe $A$ est très homogène. Les corrélations entre les descripteurs de ce groupe varient entre 0,97 et 0,99. L'estimation par la méthode du maximum de vraisemblance fournit les valeurs suivantes :

$$\hat{\mathbf{a}}^{(A)} = \begin{pmatrix} 0,990 \\ 0,994 \\ 0,997 \\ 0,981 \end{pmatrix},$$

ce qui correspond à une proportion de variance expliquée par le modèle de 0,980, 0,987, 0,993 et 0,963, et une proportion de variance de l'erreur de 0,020, 0,013, 0,007 et 0,037. A titre d'exemple, le modèle complet est donné pour le descripteur "goût doux-piquant", qui a une note moyenne de 51 et un écart-type de 13 :

$$y_{douxpiquant} = 51 + 13\left(0,994\ \xi^{(A)} + z_{douxpiquant}\right)$$

où $z_{douxpiquant}$ a une variance de 0,020.

Le groupe $B$ est moins homogène. En valeurs absolues, les corrélations entre les variables de ce groupe sont comprises entre 0,44 et 0,81. Nous obtenons :

$$\hat{\mathbf{a}}^{(B)} = \begin{pmatrix} 0,857 \\ -0,747 \\ 0,680 \\ -0,920 \\ 0,798 \end{pmatrix},$$

ce qui correspond à une variance expliquée par le modèle de 0,735, 0,558, 0,463, 0,846 et 0,637. Les signes négatifs devant la deuxième et la quatrième saturation indiquent que ce groupe comprend des variables opposées. Les descripteurs 1, 3 et 5 ("odeur chocolat", "odeur sucrée", "odeur caramel") sont opposées aux descripteurs 2 et 4 ("odeur moisi", "odeur chèvre"). Les variances de l'erreur sont 0,265, 0,442, 0,537, 0,154 et 0,363. Pour obtenir le modèle qui prend en compte le niveau et l'étendu, il faut de nouveau considérer la moyenne et l'écart-type. Par exemple, pour le descripteur "odeur moisi", nous obtenons le modèle :

$$y_{moisi} = 18 + 5\left(-0,747\ \xi^{(B)} + z_{moisi}\right)$$

où $z_{moisi}$ a une variance de 0,442. L'analyse n'est basée que sur 16 observations (les 16 variétés de café), ce qui ne suffit pas pour des estimations précises. Cependant, elle nous donne une idée sur le rapport entre variance expliquée par le modèle et variance de l'erreur. En profil conventionnel, la proportion de variance expliquée par le modèle peut atteindre 99% pour quelques descripteurs et ne pas dépasser

50% pour d'autres.

Comme décrit dans le paragraphe 3.3, les variances de l'erreur sont plus importantes dans le profil libre. Pour l'estimation de celles-ci, nous avons simulé un tableau de données sensorielles selon le profil libre en constituant un tableau ayant la structure (produits × descripteurs). Cependant, les notes d'un descripteur donné sont celles d'un juge choisi au hasard. Par exemple, les notes du descripteur "goût doux-piquant" sont les notes du juge 5 et non plus la moyenne des notes de tous les juges. Nous obtenons ainsi des données qui s'apparentent à un profil libre.

Pour les descripteurs du groupe $A$, les corrélations sont comprises entre 0,48 et 0,75, et, donc, inférieures aux valeurs obtenues dans les cas du profil conventionnel. L'analyse en facteurs communs et spécifiques selon la méthode du maximum de vraisemblance fournit les estimations suivantes :

$$\hat{\tilde{\mathbf{a}}}^{(A)} = \begin{pmatrix} 0,769 \\ 0,955 \\ 0,677 \\ 0,751 \end{pmatrix},$$

ce qui correspond à une variance expliquée par le modèle de 0,591, 0,911, 0,459 et 0,564, et une variance de l'erreur de 0,409, 0,089, 0,541 et 0,436.

Pour les descripteurs du groupe $B$, les résultats obtenus par le profil libre ne sont pas pertinents. La matrice de corrélation ne permet pas l'estimation d'un modèle à un facteur. L'algorithme (proc factor dans le logiciel sas) est interrompu dans la deuxième itération à cause d'une saturation supèrieure à 1 et donc d'une variance de l'erreur négative. Par ailleurs, les corrélations obtenues par le profil libre ne sont pas en accord avec les corrélations obtenues par le profil conventionnel. Par exemple, la corrélation entre "odeur moisi" et "odeur chocolat" est négative (-0,70) pour le profil conventionnel, tandis que la corrélation entre "odeur moisi", mesuré par le juge 7, et "odeur chocolat", mesuré par le juge 2, est positive (0,17). Nous pouvons conclure que l'information sur les descripteurs obtenue par un profil libre peut être très inexacte. Il est à conseiller de baser une analyse statistique concernant les dépendances entre les descripteurs sur des notes obtenues par le profil conventionnel.

# Chapitre 4

# Matrice de variance-covariance théorique

La méthode CLV proposée par Vigneau *et al.* [23] est adaptée à deux cas de figure : La classification selon le critère $Q$ est utilisée si une corrélation négative signifie une opposition entre variables. La classification selon le critère $T$ est utilisée si une corrélation importante en valeur absolue signifie proximité entre variables sans tenir compte du signe de la corrélation. Ici, nous analysons la classification avec le critère $T$, puisque elle est, entre autres, adaptée à la classification de descripteurs sensoriels. En effet, deux descripteurs avec une forte corrélation négative fournissent des informations qui se recoupent. Par exemple, les descripteurs "dur" et "mou" conduisent à une même caractérisation des produits. Après une présentation du critère $T$, nous allons analyser la classification avec ce critère en considérant le modèle factoriel.

## 4.1 La classification hiérarchique

### 4.1.1 Le critère $T$ et $\Delta T$

Etant donné un ensemble de variables aléatoires $\mathbf{x} = (x_1, \ldots, x_p)'$, ces variables sont découpées en $K$ groupes $G^{(1)}, \ldots, G^{(K)}$. A chaque groupe est associée une variable latente

$$c^{(k)} = \mathbf{d}^{(k)\prime}\mathbf{x}^{(k)} = \sum_{j \in G^{(k)}} d_j^{(k)} x_j^{(k)}$$

qui est une combinaison linéaire des variables du groupe $G^{(k)}$. Pour $K$ fixé, nous cherchons la partition des $p$ variables en $K$ groupes et les variables latentes qui maximisent le critère T qui est défini par :

$$T^{(K)} = \sum_{k=1}^{K} \sum_{j \in G^{(k)}} \mathrm{Cov}^2\left(x_j, c^{(k)}\right). \tag{4.1}$$

L'indice $(K)$ indique que $T$ dépend du nombre de groupes $K$. Il faut une contrainte de détermination sur $\mathbf{d}^{(k)}$ ou $c^{(k)}$. Ici, la contrainte

$$\mathrm{Var}\left(c^{(k)}\right) = 1 \tag{4.2}$$

est choisie. Pour une partition donnée, dans chaque groupe $G^{(k)}$, il faut déterminer le vecteur $\mathbf{d}^{(k)}$ qui maximise :

$$\sum_{j \in G^{(k)}} \mathrm{Cov}^2\left(x_j^{(k)}, c^{(k)}\right) = \sum_{j \in G^{(k)}} \left[\sum_{i \in G^{(k)}} d_i^{(k)} \mathrm{Cov}\left(x_j^{(k)}, x_i^{(k)}\right)\right]^2 = \mathbf{d}^{(k)\prime} {\boldsymbol{\Sigma}^{(k)}}^2 \mathbf{d}^{(k)} \tag{4.3}$$

sous la contrainte

$$\mathrm{Var}\left(\mathbf{d}^{(k)\prime}\mathbf{x}^{(k)}\right) = \mathbf{d}^{(k)\prime}\boldsymbol{\Sigma}^{(k)}\mathbf{d}^{(k)} = 1. \tag{4.4}$$

A l'aide de la fonction de Lagrange :

$$\mathrm{L}\left(\mathbf{d}^{(k)}, \lambda\right) = \mathbf{d}^{(k)\prime} {\boldsymbol{\Sigma}^{(k)}}^2 \mathbf{d}^{(k)} - \lambda\left(\mathbf{d}^{(k)\prime}\boldsymbol{\Sigma}^{(k)}\mathbf{d}^{(k)} - 1\right)$$

et du vecteur de ses dérivées partielles par rapport aux $d_i^{(k)}$

$$\frac{\partial \mathrm{L}}{\partial \mathbf{d}^{(k)}} = 2{\boldsymbol{\Sigma}^{(k)}}^2\mathbf{d}^{(k)} - 2\lambda\boldsymbol{\Sigma}^{(k)}\mathbf{d}^{(k)},$$

nous obtenons l'équation suivante :

$$\boldsymbol{\Sigma}^{(k)}\left(\boldsymbol{\Sigma}^{(k)} - \lambda\,\mathbf{I}\right)\,\mathbf{d}^{(k)} = \mathbf{0}. \tag{4.5}$$

Ainsi, les solutions du problème d'optimisation sont liées à l'analyse spectrale de $\boldsymbol{\Sigma}^{(k)}$. Soit $\mathbf{a}^{(k)}$ un vecteur propre normé de $\boldsymbol{\Sigma}^{(k)}$ associé à la valeur propre $\lambda^{(k)}$. Le vecteur $\mathbf{a}^{(k)}$ et la valeur $\lambda^{(k)}$ sont alors une solution de l'équation (4.5). Avec le choix de $\mathbf{d}^{(k)} = \frac{1}{\sqrt{\lambda^{(k)}}}\mathbf{a}^{(k)}$, la contrainte (4.4) est vérifiée : $\mathbf{d}^{(k)\prime}\boldsymbol{\Sigma}^{(k)}\mathbf{d}^{(k)} = \mathbf{d}^{(k)\prime}\lambda^{(k)}\mathbf{d}^{(k)} = \frac{1}{\sqrt{\lambda^{(k)}}}\mathbf{a}^{(k)\prime}\lambda^{(k)}\frac{1}{\sqrt{\lambda^{(k)}}}\mathbf{a}^{(k)} = 1$. Le terme à maximiser (4.3) s'écrit $\mathbf{d}^{(k)\prime}{\boldsymbol{\Sigma}^{(k)}}^2\mathbf{d}^{(k)} = \mathbf{d}^{(k)\prime}\boldsymbol{\Sigma}^{(k)}\lambda^{(k)}\mathbf{d}^{(k)} = \lambda^{(k)}$. Il est maximal si $\lambda^{(k)}$ est la plus grande valeur propre $\lambda_1^{(k)}$ de $\boldsymbol{\Sigma}^{(k)}$. $\mathbf{d}^{(k)}$ est le vecteur propre associé à $\lambda_1^{(k)}$ avec $\mathbf{d}^{(k)\prime}\mathbf{d}^{(k)} = \frac{1}{\lambda_1^{(k)}}$. Pour une partition fixée, le critère $T^{(K)}$ s'écrit donc :

$$T^{(K)} = \sum_{k=1}^{K} \mathbf{d}^{(k)\prime}{\boldsymbol{\Sigma}^{(k)}}^2\mathbf{d}^{(k)} = \sum_{k=1}^{K} \lambda_1^{(k)}. \tag{4.6}$$

Au début de l'algorithme, il y a $p$ groupes, chacun contenant une variable. Le critère $T^{(p)}$ est égal à la somme des variances des $p$ variables et, donc, égal à la trace de $\boldsymbol{\Sigma}$. A chaque étape de l'algorithme, deux groupes, disons les groupes $G^{(k)}$ et $G^{(l)}$, sont réunis. Le critère $T$ diminue de $\Delta T = \lambda_1^{(k)} + \lambda_1^{(l)} - \lambda_1^{(G^{(k)} \cup G^{(l)})}$. A la fin de la classification hiérarchique, le critère $T^{(1)}$ est égale à $\lambda_1$, la plus grande valeur propre de $\boldsymbol{\Sigma}$.

Dans la première étape de l'algorithme hiérarchique, deux variables $x_i$ et $x_j$ sont réunies dans un groupe, disons le groupe $A$. La matrice de variance-covariance du nouveau groupe est donnée par :

$$\mathbf{\Sigma}^{(A)} = \begin{pmatrix} \sigma_i^2 & \sigma_{ij} \\ \sigma_{ij} & \sigma_j^2 \end{pmatrix}.$$

Ses valeurs propres sont les solutions de l'équation :

$$(\sigma_i^2 - \lambda)(\sigma_j^2 - \lambda) - \sigma_{ij}^2 = 0.$$

Nous pouvons montrer que la plus grande valeur propre $\lambda_1^{(A)}$ est donnée par :

$$\begin{aligned} \lambda_1^{(A)} &= \frac{1}{2}\left(\sigma_i^2 + \sigma_j^2\right) + \sqrt{\frac{1}{4}\left(\sigma_i^2 + \sigma_j^2\right)^2 + \sigma_{ij}^2 - \sigma_i^2\sigma_j^2} \\ &= \frac{1}{2}\left(\sigma_i^2 + \sigma_j^2\right) + \sqrt{\sigma_{ij}^2 + \frac{1}{4}\left(\sigma_i^2 - \sigma_j^2\right)^2}. \end{aligned}$$

La diminution du critère $T$ vaut

$$\begin{aligned} \Delta T &= \sigma_i^2 + \sigma_j^2 - \lambda_1^{(A)} \\ &= \frac{1}{2}\left(\sigma_i^2 + \sigma_j^2\right) - \sqrt{\frac{1}{4}\left(\sigma_i^2 + \sigma_j^2\right)^2 + \sigma_{ij}^2 - \sigma_i^2\sigma_j^2} \qquad (4.7) \\ &= \frac{1}{2}\left(\sigma_i^2 + \sigma_j^2\right) - \sqrt{\sigma_{ij}^2 + \frac{1}{4}\left(\sigma_i^2 - \sigma_j^2\right)^2}. \qquad (4.8) \end{aligned}$$

En particulier, si les variances de toutes les variables sont égales à $\sigma^2$, la réunion des variables $x_i$ et $x_j$ dans un groupe conduit à une diminution du critère $T$ de $\Delta T = \sigma^2 - \sigma_{ij}$. Dans la première étape de l'algorithme, nous réunissons donc les deux variables avec la plus grande corrélation.

Pour des variances quelconques, la formule (4.7) montre bien que, dans la première étape de l'algorithme, le critère $\Delta T$ dépend de la différence entre $\sigma_{ij}^2$ et $\sigma_i^2\sigma_j^2$. Si $x_i$ et $x_j$ sont parfaitement corrélées ($\mathrm{Cor}(x_i, x_j) = 1$ ou $\mathrm{Cor}(x_i, x_j) = -1$), $\sigma_{ij}^2$ est égale à $\sigma_i^2\sigma_j^2$. $\Delta T$ est alors égale à zéro.

Considérons une étape ultérieure dans l'algorithme hiérarchique. Lorsque les groupes $G^{(k)}$ et $G^{(l)}$ sont réunis, la diminution du critère $T$ vaut :

$$\Delta T = \lambda_1^{(k)} + \lambda_1^{(l)} - \lambda_1^{(G^{(k)} \cup G^{(l)})}.$$

Si, dans chaque groupe, un modèle factoriel avec un facteur et des variances de l'erreur égales est vrai, la première valeur propre et, donc, le critère $\Delta T$, s'expriment en fonction des paramètres de ce modèle comme cela est développé par la suite.

### 4.1.2 Le critère $T$ et $\Delta T$ sous un modèle factoriel

Considérons la matrice de variance-covariance du groupe $G^{(k)}$ :

$$\mathbf{\Sigma}^{(k)} = \mathbf{b}^{(k)}\mathbf{b}^{(k)'} + \psi^{(k)}\mathbf{I}. \tag{4.9}$$

Comme décrit dans le paragraphe (2.3), la première valeur propre est égale à $\lambda_1^{(k)} = \mathbf{b}^{(k)'}\mathbf{b}^{(k)} + \psi^{(k)} = \sum_{j\in G^{(k)}} {b_j^{(k)}}^2 + \psi^{(k)}$. La première valeur propre d'un sous-groupe $G^{(k_1)}$ de $G^{(k)}$ est égale à $\lambda_1^{(k_1)} = \sum_{j\in G^{(k_1)}} {b_j^{(k)}}^2 + \psi^{(k)}$. La réunion de deux sous-groupes $G^{(k_1)}$ et $G^{(k_2)}$ disjoints se traduit par une diminution du critère $T$ de

$$\Delta T = \sum_{j\in G^{(k_1)}} {b_j^{(k)}}^2 + \psi^{(k)} + \sum_{j\in G^{(k_2)}} {b_j^{(k)}}^2 + \psi^{(k)} - \sum_{j\in (G^{(k_1)}\cup G^{(k_2)})} {b_j^{(k)}}^2 - \psi^{(k)} = \psi^{(k)}.$$

Ainsi, la diminution du critère $T$ est du même ordre que la variance de l'erreur.

Si les variances de l'erreur sont différentes pour les différentes variables, $\mathbf{\Sigma}^{(k)}$ est égale à :

$$\mathbf{\Sigma}^{(k)} = \mathbf{b}^{(k)}\mathbf{b}^{(k)'} + \begin{pmatrix} \psi_1^{(k)} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \psi_{p^{(k)}}^{(k)} \end{pmatrix}.$$

Si $\mathbf{v}_1 = (v_1, v_2, \ldots, v_{p_k})'$ désigne le premier vecteur propre de $\mathbf{\Sigma}^{(k)}$ avec $\mathbf{v}_1'\mathbf{v}_1 = 1$, nous avons :

$$\lambda_1^{(k)} = \mathbf{v}_1'\mathbf{\Sigma}^{(k)}\mathbf{v}_1 = \mathbf{v}_1'\mathbf{b}^{(k)}\mathbf{b}^{(k)'}\mathbf{v}_1 + \sum_{j\in G^{(k)}} \psi_j^{(k)} v_j^2.$$

La quantité $\mathbf{v}_1'\mathbf{b}^{(k)}\mathbf{b}^{(k)'}\mathbf{v}_1$ est inférieure à la première valeur propre de $\mathbf{b}^{(k)}\mathbf{b}^{(k)'}$ qui est égale à $\mathbf{b}^{(k)'}\mathbf{b}^{(k)}$ :

$$\mathbf{v}_1'\mathbf{b}^{(k)}\mathbf{b}^{(k)'}\mathbf{v}_1 < \mathbf{b}^{(k)'}\mathbf{b}^{(k)}.$$

De même, $\sum_{j\in G^{(k)}} \psi_j^{(k)} v_j^2$ est borné :

$$\sum_{j\in G^{(k)}} \psi_j^{(k)} v_j^2 < \sum_{j\in G^{(k)}} \max_i \left(\psi_i^{(k)}\right) v_j^2 = \max_i \left(\psi_i^{(k)}\right) \mathbf{v}_1'\mathbf{v}_1 = \max_i \left(\psi_i^{(k)}\right).$$

Il s'ensuit :

$$\lambda_1^{(k)} \leq \mathbf{b}^{(k)'}\mathbf{b}^{(k)} + \max_i \left(\psi_i^{(k)}\right).$$

De plus, comme $\mathbf{v}_1$ est le vecteur propre normé associé à la plus grande valeur propre, nous avons :

$$\begin{aligned} \mathbf{v}_1'\mathbf{\Sigma}^{(k)}\mathbf{v}_1 &\geq \frac{1}{\mathbf{b}^{(k)'}\mathbf{b}^{(k)}}\mathbf{b}^{(k)'}\mathbf{\Sigma}^{(k)}\mathbf{b}^{(k)} = \mathbf{b}^{(k)'}\mathbf{b}^{(k)} + \frac{1}{\mathbf{b}^{(k)'}\mathbf{b}^{(k)}} \sum_{j\in G^{(k)}} \psi_j^{(k)} {b_j^{(k)}}^2 \\ &\geq \mathbf{b}^{(k)'}\mathbf{b}^{(k)} + \frac{1}{\mathbf{b}^{(k)'}\mathbf{b}^{(k)}} \sum_{j\in G^{(k)}} \min_i \left(\psi_i^{(k)}\right) {b_j^{(k)}}^2 \\ &= \mathbf{b}^{(k)'}\mathbf{b}^{(k)} + \min_i \left(\psi_i^{(k)}\right). \end{aligned}$$

La première valeur propre $\lambda_1^{(k)}$ est donc encadrée par :

$$\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)} + \min_i \left(\psi_i^{(k)}\right) \leq \lambda_1^{(k)} \leq \mathbf{b}^{(k)\prime}\mathbf{b}^{(k)} + \max_i \left(\psi_i^{(k)}\right).$$

Pour la réunion de deux sous-groupes $G^{(k_1)}$ et $G^{(k_2)}$ du groupe $G^{(k)}$, le critère $\Delta T$ est donc encadré par :

$$\begin{aligned} \min_{i\in G^{(k_1)}} \left(\psi_i^{(k)}\right) + \min_{i\in G^{(k_2)}} \left(\psi_i^{(k)}\right) - \max_{i\in (G^{(k_1)}\cup G^{(k_2)})} \left(\psi_i^{(k)}\right) \;\; &\leq \;\; \Delta T \leq \\ \max_{i\in G^{(k_1)}} \left(\psi_i^{(k)}\right) + \max_{i\in G^{(k_2)}} \left(\psi_i^{(k)}\right) - \min_{i\in (G^{(k_1)}\cup G^{(k_2)})} \left(\psi_i^{(k)}\right). \end{aligned}$$

Après avoir analysé la réunion de variables d'un même groupe, examinons maintenant le cas de la réunion de deux groupes $G^{(k)}$ et $G^{(l)}$ séparés. Ici, nous supposons de nouveau des variances de l'erreur égales pour les variables d'un même groupe.

**Le cas orthogonal**

Considérons d'abord un cas extrême où les corrélations entre variables de différents groupes sont égales à zéro. L'ordre des variables peut être changé de manière à ce que la matrice de variance-covariance du groupe $G^{(k)} \cup G^{(l)}$ puisse s'écrire sous la forme :

$$\mathbf{\Sigma}^{(G^{(k)}\cup G^{(l)})} = \begin{pmatrix} \mathbf{\Sigma}^{(k)} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}^{(l)} \end{pmatrix}.$$

Pour trouver les valeurs propres de $\mathbf{\Sigma}^{(G^{(k)}\cup G^{(l)})}$, nous pouvons tirer profit des propriétés des matrices diagonales par blocs. L'ensemble des valeurs propres de $\mathbf{\Sigma}^{(G^{(k)}\cup G^{(l)})}$ est donné par toutes les valeurs propres de $\mathbf{\Sigma}^{(k)}$ et $\mathbf{\Sigma}^{(l)}$. La plus grande valeur propre $\lambda_1^{(G^{(k)}\cup G^{(k)})}$ de $\mathbf{\Sigma}^{(G^{(k)}\cup G^{(l)})}$ est donc égale à $\max(\lambda_1^{(k)}, \lambda_1^{(l)})$. La diminution du critère $T$ vaut

$$\begin{aligned} \Delta T &= \lambda_1^{(k)} + \lambda_1^{(l)} - \max(\lambda_1^{(k)}, \lambda_1^{(l)}) \\ &= \min(\lambda_1^{(k)}, \lambda_1^{(l)}). \end{aligned}$$

Si le modèle (4.9) est vrai pour chacun des deux groupes, $\Delta T$ est donc égale à

$$\Delta T = \min(\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)} + \psi^{(k)}, \mathbf{b}^{(l)\prime}\mathbf{b}^{(l)} + \psi^{(l)}).$$

Résumons les résultats obtenus jusqu'ici pour le modèle stipulant qu'il y a $K$ groupes de variables dont les variables latentes ne sont pas corrélées. La matrice de variance-covariance entre les $p$ variables manifestes est donnée par (à une permutation des lignes et colonnes près) :

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{b}^{(1)}\mathbf{b}^{(1)\prime} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{b}^{(K)}\mathbf{b}^{(K)\prime} \end{pmatrix} + \begin{pmatrix} \psi^{(1)}\mathbf{I}_{p^{(1)}} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \psi^{(K)}\mathbf{I}_{p^{(K)}} \end{pmatrix}. \qquad (4.10)$$

A la première étape de l'algorithme hiérarchique, chaque variable forme un groupe à elle seule. Le critère $T$ vaut alors :

$$T^{(p)} = \mathrm{tr}(\mathbf{\Sigma}) = \sum_{k=1}^{K} \left( \mathbf{b}^{(k)\prime} \mathbf{b}^{(k)} + p^{(k)} \psi^{(k)} \right).$$

Si $\psi^{(k)} < \mathbf{b}^{(l)\prime} \mathbf{b}^{(l)} + \psi^{(l)}$ pour tout $k, l = 1, \ldots, K$, l'algorithme hiérarchique forme correctement les $K$ groupes. Nous avons :

$$T^{(K)} = \sum_{k=1}^{K} \left( \mathbf{b}^{(k)\prime} \mathbf{b}^{(k)} + \psi^{(k)} \right).$$

Après formation des $K$ groupes, l'algorithme hiérarchique réunit, dans chaque étape, deux groupes séparés. Pour cela, le groupe avec la valeur minimale de $\mathbf{b}^{(k)\prime} \mathbf{b}^{(k)} + \psi^{(k)}$ est, dans un premier temps, agrégé avec n'importe quel autre groupe, et ainsi de suite, jusqu'à ce que toutes les variables soient réunies dans un seul groupe. Après cette dernière étape, le critère $T$ vaut :

$$T^{(1)} = \max_{k} \left( \mathbf{b}^{(k)\prime} \mathbf{b}^{(k)} + \psi^{(k)} \right).$$

Lors de la réunion de deux variables du groupe $G^{(k)}$, le critère $\Delta T$ est égale à $\psi^{(k)}$ et donc plus petit que la variance d'une variable $x_j^{(k)}$ du groupe qui est, elle, égale à ${b_j^{(k)}}^2 + \psi^{(k)}$. Lors de la réunion de deux groupes différents, $\Delta T$ est égale à $\min\left(\lambda_1^{(k)}, \lambda_1^{(l)}\right)$ et donc plus important que la variance d'une variable. Or, si une classification hiérarchique est effectuée sur des variables non corrélés (ce peut être le cas du bruit), le critère $\Delta T$ est égal à la variance d'une variable. Nous pouvons utiliser ce fait pour la détermination du nombre de groupes (voir le chapitre 7).

**Le cas oblique**

Lorsque les facteurs de deux groupes différents sont corrélés, leur matrice de variance-covariance est égale à :

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{b}^{(k)} \mathbf{b}^{(k)\prime} + \psi^{(k)} \mathbf{I}_{p^{(k)}} & \phi^{(kl)} \mathbf{b}^{(k)} \mathbf{b}^{(l)\prime} \\ \phi^{(kl)} \mathbf{b}^{(l)} \mathbf{b}^{(k)\prime} & \mathbf{b}^{(l)} \mathbf{b}^{(l)\prime} + \psi^{(l)} \mathbf{I}_{p^{(l)}} \end{pmatrix}$$

où $\phi^{(kl)}$ est la corrélation entre les facteurs des groupes $G^{(k)}$ et $G^{(l)}$. Pour déterminer la plus grande valeur propre, nous pouvons utiliser les résultats A.1 et A.4 (voir annexe A). Nous obtenons :

$$\begin{aligned} \lambda_1^{(G^{(k)} \cup G^{(l)})} &= \frac{1}{2} \left( \mathbf{b}^{(k)\prime} \mathbf{b}^{(k)} + \psi^{(k)} + \mathbf{b}^{(l)\prime} \mathbf{b}^{(l)} + \psi^{(l)} \right) \\ &+ \frac{1}{2} \sqrt{\left( \mathbf{b}^{(k)\prime} \mathbf{b}^{(k)} + \psi^{(k)} - \mathbf{b}^{(l)\prime} \mathbf{b}^{(l)} - \psi^{(l)} \right)^2 + 4 \, {\phi^{(kl)}}^2 \mathbf{b}^{(k)\prime} \mathbf{b}^{(k)} \mathbf{b}^{(l)\prime} \mathbf{b}^{(l)}}. \end{aligned}$$

Lorsque les groupes $G^{(k)}$ et $G^{(l)}$ sont réunis, le critère $\Delta T$ vaut

$$\begin{aligned}\Delta T &= \lambda_1^{(k)} + \lambda_1^{(l)} - \lambda_1^{(G^{(k)} \cup G^{(l)})} \\ &= \frac{1}{2}\left(\lambda_1^{(k)} + \lambda_1^{(l)} - \sqrt{\left(\lambda_1^{(k)} - \lambda_1^{(l)}\right)^2 + 4\,{\phi^{(kl)}}^2 \mathbf{b}^{(k)\prime}\mathbf{b}^{(k)}\mathbf{b}^{(l)\prime}\mathbf{b}^{(l)}}\right).\end{aligned}$$

Prenant en compte que

$$0 \leq {\phi^{(kl)}}^2 \leq 1,$$

nous avons :

$$\begin{aligned}\lambda_1^{(G^{(k)} \cup G^{(l)})} &\geq \frac{1}{2}\left(\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)} + \psi^{(k)} + \mathbf{b}^{(l)\prime}\mathbf{b}^{(l)} + \psi^{(l)}\right) \\ &\quad + \frac{1}{2}\sqrt{\left(\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)} + \psi^{(k)} - \mathbf{b}^{(l)\prime}\mathbf{b}^{(l)} - \psi^{(l)}\right)^2} \\ &= \max(\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)} + \psi^{(k)}, \mathbf{b}^{(l)\prime}\mathbf{b}^{(l)} + \psi^{(l)}) \\ &= \max(\lambda_1^{(k)}, \lambda_1^{(l)})\end{aligned}$$

et

$$\begin{aligned}\Delta T &\leq \lambda_1^{(k)} + \lambda_1^{(l)} - \max(\lambda_1^{(k)}, \lambda_1^{(l)}) \\ &= \min(\lambda_1^{(k)}, \lambda_1^{(l)}).\end{aligned}$$

De même,

$$\begin{aligned}\lambda_1^{(G^{(k)} \cup G^{(l)})} &\leq \frac{1}{2}\left(\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)} + \psi^{(k)} + \mathbf{b}^{(l)\prime}\mathbf{b}^{(l)} + \psi^{(l)}\right) \\ &\quad + \frac{1}{2}\sqrt{\left(\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)} + \psi^{(k)} - \mathbf{b}^{(l)\prime}\mathbf{b}^{(l)} - \psi^{(l)}\right)^2 + 4\,\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)}\mathbf{b}^{(l)\prime}\mathbf{b}^{(l)}}\end{aligned}$$

et

$$\begin{aligned}\Delta T &\geq \frac{1}{2}\left(\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)} + \psi^{(k)} + \mathbf{b}^{(l)\prime}\mathbf{b}^{(l)} + \psi^{(l)}\right) \\ &\quad - \frac{1}{2}\sqrt{\left(\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)} + \psi^{(k)} - \mathbf{b}^{(l)\prime}\mathbf{b}^{(l)} - \psi^{(l)}\right)^2 + 4\,\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)}\mathbf{b}^{(l)\prime}\mathbf{b}^{(l)}}.\end{aligned}$$

Ces formules se simplifient dans le cas $\psi^{(k)} = \psi^{(l)} = \psi$ :

$$\begin{aligned}\lambda_1^{(G^{(k)} \cup G^{(l)})} &\leq \frac{1}{2}\left(\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)} + \psi + \mathbf{b}^{(l)\prime}\mathbf{b}^{(l)} + \psi\right) \\ &\quad + \frac{1}{2}\sqrt{\left(\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)} + \psi - \mathbf{b}^{(l)\prime}\mathbf{b}^{(l)} - \psi\right)^2 + 4\,\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)}\mathbf{b}^{(l)\prime}\mathbf{b}^{(l)}} \\ &= \psi + \frac{1}{2}\left(\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)} + \mathbf{b}^{(l)\prime}\mathbf{b}^{(l)}\right) + \frac{1}{2}\sqrt{\left(\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)} + \mathbf{b}^{(l)\prime}\mathbf{b}^{(l)}\right)^2} \\ &= \psi + \mathbf{b}^{(k)\prime}\mathbf{b}^{(k)} + \mathbf{b}^{(l)\prime}\mathbf{b}^{(l)}\end{aligned}$$

et

$$\begin{aligned}\Delta T &\geq \psi + \frac{1}{2}\left(\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)} + \mathbf{b}^{(l)\prime}\mathbf{b}^{(l)}\right) - \frac{1}{2}\sqrt{\left(\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)} + \mathbf{b}^{(l)\prime}\mathbf{b}^{(l)}\right)^2} \\ &= \psi.\end{aligned}$$

Si $\psi^{(k)} = \psi^{(l)} = \psi$, nous avons donc

$$\psi \leq \Delta T \leq \min\left(\lambda_1^{(k)}, \lambda_1^{(l)}\right).$$

La borne inférieure est atteinte si $|\phi^{(kl)}| = 1$. Ceci correspond au cas où les deux "groupes" correspondent en réalité à un seule groupe. La borne supérieure est atteinte si $\phi^{(kl)} = 0$ et, donc, si les variables latentes des deux groupes ne sont pas corrélées. Aux extrèmes, nous retrouvons donc les résultats que nous avons déjà trouvé pour le cas de la réunion de deux sous-groupes d'un même groupe et pour le cas de la réunion de deux groupes orthogonaux.

## 4.2 L'algorithme de partitionnement

Il est d'usage de compléter la classification hiérarchique par une classification par agrégation autour des centres mobiles. La procédure CLV préconise également cette démarche (Vigneau *et al.* [23]). Il est à souligner que des procédures de classification basées sur le même critère que CLV ont été proposées (Braverman [3], Escoufier [6], Dhillon *et al.* [4]). Cependant, ces procédures ne sont basées que sur des algorithmes itératifs de type nuées dynamiques. De ce fait, elles dépendent de l'initialisation qui a un impact important sur les résultats, comme nous allons le démontrer dans ce paragraphe. Dans ce qui suit, il est important de faire la distinction entre les vrais groupes que nous désignons par $G^{(1)}, \ldots, G^{(K)}$ auxquels les variables appartiennent, et les groupes définis suite à l'exécution de l'algorithme. Pour cette raison, les groupes formés par l'algorithme sont désignés par $C^{(1)}, \ldots, C^{(K)}$. L'algorithme de partitionnement se compose des étapes suivantes :

1. Initialisation : détermination de $K$ centres (variables latentes) $c^{(1)}, \ldots, c^{(K)}$.
2. Réaffectation des variables aux classes sur la base du carré de leur covariance avec les centres.
3. Dans chaque classe $C^{(k)}$, calcul de la variable latente $c^{(k)}$ comme étant la première composante principale du groupe.
4. Répétition de 2 et 3 jusqu'à la convergence (aucune variable ne change de classe à l'étape 2).

Cet algorithme vise à maximiser le même critère que pour la classification hiérarchique (voir la formule (4.1)). Vigneau *et al.* [23] proposent l'initialisation par la coupure de l'arbre hiérarchique. Les $K$ centres initiaux sont alors les variables latentes $c^{(k)}$ $(k = 1, \ldots, K)$ des $K$ groupes issus de la classification hiérarchique. Une

autre possibilité consiste à choisir $K$ variables au hasard comme centres initiaux. Pour vérifier la contrainte (4.2), ces variables doivent être standardisées. Ensuite, la variable $x_j$ est affectée à la classe $C^{(k)}$ si :

$$\text{Cov}^2(x_j, c^{(k)}) = \max_{g=1,\ldots,K} \text{Cov}^2(x_j, c^{(g)}).$$

La mise à jour des variables latentes se fait par la formule

$$c^{(k)} = \mathbf{d}^{(k)\prime}\mathbf{x}^{(k)} = \sum_{i \in C^{(k)}} d_i^{(k)} x_i$$

où $\mathbf{d}^{(k)}$ est le premier vecteur propre de la matrice de variance-covariance de $\mathbf{x}^{(k)}$ (le vecteur des variables appartenant, à l'étape courante, à la classe $C^{(k)}$). Le vecteur propre $\mathbf{d}^{(k)}$ est normalisé comme décrit dans le paragraphe 4.1.1.

Considérons le modèle associé à la matrice de variance-covariance $\mathbf{\Sigma}$ définie par :

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}^{(1)} & \mathbf{\Sigma}^{(12)} & \ldots & \mathbf{\Sigma}^{(1K)} \\ \mathbf{\Sigma}^{(12)\prime} & \mathbf{\Sigma}^{(2)} & \ldots & \mathbf{\Sigma}^{(2K)} \\ \ldots & \ldots & \ldots & \ldots \\ \mathbf{\Sigma}^{(1K)\prime} & \ldots & \ldots & \mathbf{\Sigma}^{(K)} \end{pmatrix}$$

où

$$\mathbf{\Sigma}^{(k)} = \mathbf{b}^{(k)}\mathbf{b}^{(k)\prime} + \psi^{(k)}\mathbf{I}_{p^{(k)}}$$

et

$$\mathbf{\Sigma}^{(km)} = \phi^{(km)}\mathbf{b}^{(k)}\mathbf{b}^{(m)\prime}$$

avec $-1 \leq \phi^{(km)} \leq 1$. Dans un premier temps, nous nous intéressons au cas où la partition initiale est identique à la partition correcte des $K$ groupes.

### 4.2.1 La partition correcte comme partition initiale

Si chaque variable est correctement classée, la covariance au carré entre une variable et la variable latente devrait être maximale pour la classe dans laquelle la variable se trouve à l'étape courante. Si cela n'est pas le cas, le modèle n'est pas compatible avec le critère $T$. En fait, même pour des valeurs importantes de $\phi^{(km)}$, le critère $T$ a un maximum pour la partition correcte. La covariance au carré d'une variable $x_j^{(k)}$ avec la variable latente $c^{(k)}$ de son propre groupe est égale à :

$$\text{Cov}^2\left(x_j^{(k)}, c^{(k)}\right) = \left(\sum_{i=1}^{p^{(k)}} d_i^{(k)} \text{Cov}\left(x_j^{(k)}, x_i^{(k)}\right)\right)^2$$

$$
\begin{aligned}
&= \left( \sum_{i=1}^{p^{(k)}} \frac{1}{\sqrt{\lambda_1^{(k)}}} \frac{1}{\sqrt{\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)}}} \, b_i^{(k)} \mathrm{Cov}\left(x_j^{(k)}, x_i^{(k)}\right) \right)^2 \\
&= \frac{1}{(\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)} + \psi^{(k)})\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)}} \left( \sum_{i=1}^{p^{(k)}} b_i^{(k)} b_j^{(k)} b_i^{(k)} + b_j^{(k)} \psi^{(k)} \right)^2 \\
&= \frac{\left( b_j^{(k)} \mathbf{b}^{(k)\prime}\mathbf{b}^{(k)} + b_j^{(k)} \psi^{(k)} \right)^2}{(\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)} + \psi^{(k)})\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)}} = \frac{{b_j^{(k)}}^2 \left( \mathbf{b}^{(k)\prime}\mathbf{b}^{(k)} + \psi^{(k)} \right)}{\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)}} \\
&= \left( 1 + \frac{\psi^{(k)}}{\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)}} \right) {b_j^{(k)}}^2.
\end{aligned}
$$

La covariance au carré d'une variable $x_j^{(k)}$ avec la variable latente $c^{(m)}$ d'un autre groupe est égale à :

$$
\begin{aligned}
\mathrm{Cov}^2\left(x_j^{(k)}, c^{(m)}\right) &= \left( \sum_{i=1}^{p^{(m)}} d_i^{(m)} \mathrm{Cov}\left(x_j^{(k)}, x_i^{(m)}\right) \right)^2 \\
&= \frac{1}{(\mathbf{b}^{(m)\prime}\mathbf{b}^{(m)} + \psi_m)\mathbf{b}^{(m)\prime}\mathbf{b}^{(m)}} \left( \sum_{i=1}^{p^{(m)}} b_i^{(m)} \phi^{(km)} b_j^{(k)} b_i^{(m)} \right)^2 \\
&= \frac{\mathbf{b}^{(m)\prime}\mathbf{b}^{(m)}}{\mathbf{b}^{(m)\prime}\mathbf{b}^{(m)} + \psi^{(m)}} {\phi^{(km)}}^2 {b_j^{(k)}}^2
\end{aligned}
$$

La variable $x_j^{(k)}$ reste dans son groupe si

$$
\left( 1 + \frac{\psi^{(k)}}{\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)}} \right) {b_j^{(k)}}^2 > \frac{\mathbf{b}^{(m)\prime}\mathbf{b}^{(m)}}{\mathbf{b}^{(m)\prime}\mathbf{b}^{(m)} + \psi^{(m)}} {\phi^{(km)}}^2 {b_j^{(k)}}^2.
$$

Ceci est équivalent à

$$
\begin{aligned}
{\phi^{(km)}}^2 &< \left( 1 + \frac{\psi^{(k)}}{\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)}} \right) \frac{\mathbf{b}^{(m)\prime}\mathbf{b}^{(m)} + \psi^{(m)}}{\mathbf{b}^{(m)\prime}\mathbf{b}^{(m)}} \\
&= \left( 1 + \frac{\psi^{(k)}}{\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)}} \right) \left( 1 + \frac{\psi^{(m)}}{\mathbf{b}^{(m)\prime}\mathbf{b}^{(m)}} \right).
\end{aligned}
$$

Cette condition est toujours verifiée puisque $0 \leq {\phi^{(km)}}^2 \leq 1$, $\frac{\psi^{(m)}}{\mathbf{b}^{(m)\prime}\mathbf{b}^{(m)}} > 0$ et $\frac{\psi^{(k)}}{\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)}} > 0$.

Il est intéressant de constater que la condition est aussi vérifiée pour $\phi^{(km)} = 1$. Ceci signifie que si un groupe $G^{(k)}$ est partagé en deux classes $C^{(k_1)}$ et $C^{(k_2)}$, chaque variable de ce groupe reste dans la classe à laquelle elle est affectée à l'étape courante.

### 4.2.2 Une partition quelconque comme partition initiale

Pour analyser le comportement de l'algorithme de partitionnement basé sur une partition quelconque comme partition initiale, nous considérons des groupes non corrélés. En fonction des paramètres des vrais groupes $G^{(1)}, \ldots, G^{(K)}$, la matrice de variance-covariance d'une classe $C^{(k)}$ formée par l'algorithme, a la structure

$$\mathbf{\Sigma}^{(C^{(k)})} = \begin{pmatrix} \mathbf{b}^{(G^{(1)} \cap C^{(k)})} \mathbf{b}^{(G^{(1)} \cap C^{(k)})'} + \psi^{(1)} \mathbf{I} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{b}^{(G^{(K)} \cap C^{(k)})} \mathbf{b}^{(G^{(K)} \cap C^{(k)})'} + \psi^{(K)} \mathbf{I} \end{pmatrix}$$

où $\mathbf{b}^{(G^{(m)} \cap C^{(k)})}$ est un sous-vecteur du vecteur $\mathbf{b}^{(m)}$. Il contient les entrées qui correspondent à des variables du groupe $G^{(m)}$ appartenant, à l'étape courante, à la classe $C^{(k)}$. Les valeurs propres de $\mathbf{\Sigma}^{(C^{(k)})}$ sont égales aux valeurs propres des blocs diagonaux. La plus grande valeur propre est donc le maximum des plus grandes valeurs propres des blocs :

$$\lambda_1^{(C^{(k)})} = \max_l \sum_{i \in (G^{(l)} \cap C^{(k)})} \left( {b_i^{(l)}}^2 + \psi^{(l)} \right).$$

Supposons que le maximum soit atteint pour le groupe $G^{(m)}$. Le premier vecteur propre associé à $C^{(k)}$ (normalisé par $\mathbf{d}^{(C^{(k)})'} \mathbf{d}^{(C^{(k)})} = \frac{1}{\lambda_1^{(C^{(k)})}}$) est alors égal à

$$\mathbf{d}^{(C^{(k)})} = \frac{1}{\sqrt{\sum_{i (\in G^{(m)} \cap C^{(k)})} {b_i^{(m)}}^2 + \psi^{(m)}}} \frac{1}{\sqrt{\sum_{i \in (G^{(m)} \cap C^{(k)})} {b_i^{(m)}}^2}} \begin{pmatrix} \mathbf{0} \\ \mathbf{b}^{(G^{(m)} \cap C^{(k)})} \\ \mathbf{0} \end{pmatrix}.$$

La variable latente $c^{(k)}$ est donc une combinaison linéaire des variables de la classe $C^{(k)}$ qui appartiennent au groupe $G^{(m)}$. Les variables des autres groupes qui appartiennent à la classe $C^{(k)}$ ne sont pas prises en compte pour le calcul de $c^{(k)}$. Nous disons dans ce cas que le groupe $G^{(m)}$ domine la classe $C^{(k)}$. Ainsi, pour un groupe $G^{(l)}$ donné, il y a trois possibilités :

1. Il ne domine aucune classe. Dans ce cas, la covariance d'une variable de $G^{(l)}$ avec n'importe quelle variable latente $c^{(k)}$ est égale à zéro. La réaffectation des variables de $G^{(l)}$ aux classes $C^{(1)}, \ldots, C^{(K)}$ se fait d'une manière arbitraire.
2. Le groupe $G^{(l)}$ domine exactement une classe, disons la classe $C^{(r)}$. La covariance au carré d'une variable de $G^{(l)}$ avec la variable latente $c^{(r)}$ de $C^{(r)}$ est strictement positive tandis que les covariances au carré avec les variables latentes des autres groupes sont égales à zéro. Toutes les variables du groupe $G^{(l)}$ sont affectées au groupe $C^{(r)}$.
3. Le groupe $G^{(l)}$ domine deux classes ou plus, disons les classes $C^{(r_1)}, \ldots, C^{(r_T)}$. Une variable du groupe $G^{(l)}$ qui est dans une de ces classes, y reste. Ceci est

un résultat de la remarque à la fin du paragraphe 4.2.1. Pour une variable $x_j^{(l)}$ du groupe $G^{(l)}$ qui est dans une autre classe, nous avons :

$$\mathrm{Cov}^2\left(x_j^{(l)}, c^{(r_t)}\right) = \frac{\sum_{i\in(G^{(l)}\cap C^{(r_t)})} {b_i^{(l)}}^2}{\sum_{i\in(G^{(l)}\cap C^{(r_t)})} {b_i^{(l)}}^2 + \psi^{(l)}} \, {b_j^{(l)}}^2, \qquad t = 1, \ldots, T.$$

La covariance au carré avec les variables latentes des autres classes est égale à zéro. La variable $x_j^{(l)}$ est donc associée à la classe $C^{(r_s)}$ pour laquelle

$$\frac{\sum_{i\in(G^{(l)}\cap C^{(r_s)})} {b_i^{(l)}}^2}{\sum_{i\in(G^{(l)}\cap C^{(r_s)})} {b_i^{(l)}}^2 + \psi^{(l)}} = \max_t \left( \frac{\sum_{i\in(G^{(l)}\cap C^{(r_t)})} {b_i^{(l)}}^2}{\sum_{i\in(G^{(l)}\cap C^{(r_t)})} {b_i^{(l)}}^2 + \psi^{(l)}} \right).$$

La classe $C^{(r_s)}$ attire toutes les variables du groupe $G^{(l)}$ qui ne sont pas dans une des classes $C^{(r_1)}, \ldots, C^{(r_T)}$.

En conclusion, la partition correcte est atteinte dans une seule boucle de l'algorithme de partitionnement, si chaque groupe domine exactement une classe. Dans les autres cas, la partition correcte n'est pas trouvée. Les groupes qui ne dominent aucune classe se répartissent arbitrairement dans les classes. Les groupes qui dominent plusieurs classes se répartissent dans ces classes avec un plus fort poids sur une classe. Une bonne initialisation est donc primordiale. Si une initialisation aléatoire est choisie, il est fortement conseillé de répéter l'algorithme de partitionnement plusieurs fois, chaque fois avec une autre initialisation. Vigneau et al. [23] proposent l'initialisation par l'algorithme hiérarchique. Sous certaines conditions, cet algorithme fournit déjà la bonne partition (voir le paragraphe 4.1.2). L'algorithme de partitionnement n'est donc pas nécessaire, si la classification est basée sur la matrice de variance-covariance de la population. Le comportement de l'algorithme hiérarchique sur la matrice de variance-covariance empirique est analysé dans le chapitre suivant. Dans le chapitre 6, l'approche CLV (algorithme hiérarchique et partitionnement) est comparé avec d'autres méthodes à l'aide d'une étude de simulation.

# Chapitre 5

# Matrice de variance-covariance empirique

Les résultats enoncés dans le chapitre précédent concernent la matrice $\mathbf{\Sigma}$ de variance-covariance de la population. Dans ce chapitre, nous analysons le comportement de l'algorithme hiérarchique sur la matrice de variance-covariance empirique $S$ basée sur un échantillon de taille $n$. Les critères $T$ et $\Delta T$ calculés à partir de $S$ sont notées par $\hat{T}$ et $\Delta\hat{T}$. Dans un premier temps, nous calculons l'espérance mathématique de $\hat{T}$ et $\Delta\hat{T}$. Ensuite, nous comparons les formules obtenues avec les résultats de simulations. Une analyse complète du comportement de l'algorithme devrait tenir compte de la variance de $\Delta\hat{T}$ ainsi que les différentes étapes de la classification hiérarchique. Celles-ci dépendent de toutes les étapes précédentes. Il est évident qu'une telle analyse est très complexe. Comme nous allons le voir dans ce chapitre, même les formules pour l'espérance de $\Delta\hat{T}$ ne sont pas suffisamment exactes. Pour ces raisons, la performance de la classification hiérarchique est étudiée au moyen d'une étude de simulations présentée dans le chapitre 6.

## 5.1 Espérance mathématique du critère $\hat{T}$

Si la plus grande valeur propre $\lambda_1$ de la matrice de variance-covariance théorique est distincte des autres valeurs propres, l'espérance mathématique de la plus grande valeur propre $l_1$ de la matrice de variance-covariance empirique est égale à :

$$\mathrm{E}(l_1) = \lambda_1 + \frac{\lambda_1}{n-1}\sum_{i=2}^{p}\left(\frac{\lambda_i}{\lambda_1-\lambda_i}\right) + O\left(\frac{1}{(n-1)^2}\right) \tag{5.1}$$

(Lawley [12]). Cette approximation est appropriée si le nombre d'individus $n$ est suffisamment grand par rapport à l'inverse de $\lambda_1 - \lambda_2$.

Rappelons qu'avec un modèle factoriel, la matrice de variance-covariance associée

au groupe $G^{(k)}$ s'écrit :

$$\mathbf{\Sigma}^{(k)} = \mathbf{b}^{(k)}\mathbf{b}^{(k)'} + \psi^{(k)}\mathbf{I}.$$

Ses valeurs propres sont $\lambda_1^{(k)} = \mathbf{b}^{(k)'}\mathbf{b}^{(k)} + \psi^{(k)}$ et $\lambda_2^{(k)} = \ldots = \lambda_{p^{(k)}}^{(k)} = \psi^{(k)}$. La première valeur propre étant distincte des autres valeurs propres, la formule (5.1) s'applique. Nous obtenons :

$$\mathrm{E}\left(l_1^{(k)}\right) = \lambda_1^{(k)} + \frac{1}{n-1}\ \left(p^{(k)} - 1\right)\ \psi^{(k)}\left(1 + \frac{\psi^{(k)}}{\mathbf{b}^{(k)'}\mathbf{b}^{(k)}}\right) + O\left(\frac{1}{(n-1)^2}\right).$$

S'il y a $K$ groupes, l'espérance du critère $\hat{T}^{(K)}$ pour la partition correcte vaut donc :

$$\begin{aligned}
\mathrm{E}\left(\hat{T}^{(K)}\right) &= \sum_{k=1}^{K} \mathrm{E}\left(l_1^{(k)}\right) \\
&= T^{(K)} + \frac{1}{n-1}\sum_{k=1}^{K}\left[\left(p^{(k)} - 1\right)\psi^{(k)}\left(1 + \frac{\psi^{(k)}}{\mathbf{b}^{(k)'}\mathbf{b}^{(k)}}\right)\right] \\
&\quad + O\left(\frac{1}{(n-1)^2}\right).
\end{aligned}$$

$\hat{T}^{(K)}$ a un biais positif. Si nous interprétons $\hat{T}$ comme la variance expliquée par les variables latentes $c^{(k)}$ des groupes, il s'ensuit que nous surestimons la variance expliquée.

## 5.2 Espérance mathématique du critère $\Delta\hat{T}$

A partir des résultats précédents, nous pouvons déduire l'espérance mathématique de $\Delta\hat{T}$. Pour la réunion des groupes $G^{(k)}$ et $G^{(m)}$, nous obtenons

$$\begin{aligned}
\mathrm{E}(\Delta\hat{T}) &= \mathrm{E}\left(l_1^{(k)} + l_1^{(m)} - l_1^{(G^{(k)}\cup G^{(m)})}\right) \\
&= \lambda_1^{(k)} + \delta_{p^{(k)}}\frac{\lambda_1^{(k)}}{n-1}\sum_{i=2}^{p^{(k)}}\left(\frac{\lambda_i^{(k)}}{\lambda_1^{(k)} - \lambda_i^{(k)}}\right) \\
&\quad + \lambda_1^{(m)} + \delta_{p^{(m)}}\frac{\lambda_1^{(m)}}{n-1}\sum_{i=2}^{p^{(m)}}\left(\frac{\lambda_i^{(m)}}{\lambda_1^{(m)} - \lambda_i^{(m)}}\right) \\
&\quad - \lambda_1^{(G^{(k)}\cup G^{(m)})} - \frac{\lambda_1^{(G^{(k)}\cup G^{(m)})}}{n-1}\sum_{i=2}^{p^{(k)}+p^{(m)}}\left(\frac{\lambda_i^{(G^{(k)}\cup G^{(m)})}}{\lambda_1^{(G^{(k)}\cup G^{(m)})} - \lambda_i^{(G^{(k)}\cup G^{(m)})}}\right) \\
&\quad + O\left(\frac{1}{(n-1)^2}\right)
\end{aligned}$$

$$
\begin{aligned}
&= \Delta T \\
&\quad + \delta_{p^{(k)}} \frac{\lambda_1^{(k)}}{n-1} \sum_{i=2}^{p^{(k)}} \left( \frac{\lambda_i^{(k)}}{\lambda_1^{(k)} - \lambda_i^{(k)}} \right) + \delta_{p^{(m)}} \frac{\lambda_1^{(m)}}{n-1} \sum_{i=2}^{p^{(m)}} \left( \frac{\lambda_i^{(m)}}{\lambda_1^{(m)} - \lambda_i^{(m)}} \right) \\
&\quad - \frac{\lambda_1^{(G^{(k)} \cup G^{(m)})}}{n-1} \sum_{i=2}^{p^{(k)}+p^{(m)}} \left( \frac{\lambda_i^{(G^{(k)} \cup G^{(m)})}}{\lambda_1^{(G^{(k)} \cup G^{(m)})} - \lambda_i^{(G^{(k)} \cup G^{(m)})}} \right) + O\left( \frac{1}{(n-1)^2} \right)
\end{aligned}
$$

où $\delta_{p^{(k)}} = 1$ si $p^{(k)} \geq 2$, et 0, sinon, et $\delta_{p^{(m)}} = 1$ si $p^{(m)} \geq 2$, et 0, sinon.

Si deux sous-groupes $G^{(k_1)}$ et $G^{(k_2)}$ du groupe $G^{(k)}$ sont réunis, l'espérance de $\Delta\hat{T}$ est égale à

$$
\begin{aligned}
\mathrm{E}(\Delta\hat{T}) &= \Delta T + \frac{\psi^{(k)}}{n-1} \left[ \left(p^{(k_1)} - 1\right) \left( 1 + \frac{\psi^{(k)}}{\sum_{i \in G^{(k_1)}} {b_i^{(k)}}^2} \right) \right. \\
&\quad + \left(p^{(k_2)} - 1\right) \left( 1 + \frac{\psi^{(k)}}{\sum_{i \in G^{(k_2)}} {b_i^{(k)}}^2} \right) \\
&\quad \left. - \left(p^{(k_1)} + p^{(k_2)} - 1\right) \left( 1 + \frac{\psi^{(k)}}{\sum_{i \in G^{(k_1)}} {b_i^{(k)}}^2 + \sum_{i \in G^{(k_2)}} {b_i^{(k)}}^2} \right) \right] \\
&\quad + O\left( \frac{1}{(n-1)^2} \right). \qquad (5.2)
\end{aligned}
$$

Considérons maintenant la matrice de variance-covariance de deux groupes dont les variables latentes ne sont pas corrélées :

$$
\mathbf{\Sigma}^{(G^{(k)} \cup G^{(m)})} = \begin{pmatrix} \mathbf{\Sigma}^{(k)} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}^{(m)} \end{pmatrix} = \begin{pmatrix} \mathbf{b}^{(k)}\mathbf{b}^{(k)\prime} + \psi^{(k)}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{b}^{(m)}\mathbf{b}^{(m)\prime} + \psi^{(m)}\mathbf{I} \end{pmatrix}.
$$

Ses valeurs propres sont l'ensemble des valeurs propres de $\mathbf{\Sigma}^{(k)}$ et des valeurs propres de $\mathbf{\Sigma}^{(m)}$. Supposons, pour fixer les idées, que $\lambda_1^{(k)} > \lambda_1^{(m)}$. La plus grande valeur propre de $\mathbf{\Sigma}^{(G^{(k)} \cup G^{(m)})}$ est alors $\lambda_1^{(k)}$. L'approximation (5.1) pour l'espérance de $l_1^{(G^{(k)} \cup G^{(m)})}$ n'est valable que si $\lambda_1^{(k)} >> \lambda_1^{(m)}$. Si la différence entre $\lambda_1^{(k)}$ et $\lambda_1^{(m)}$ est suffisamment importante, nous avons :

$$
\begin{aligned}
\mathrm{E}\left( l_1^{(G^{(k)} \cup G^{(m)})} \right) &= \lambda_1^{(G^{(k)} \cup G^{(m)})} + \frac{\lambda_1^{(G^{(k)} \cup G^{(m)})}}{n-1} \sum_{i=2}^{p^{(k)}+p^{(m)}} \frac{\lambda_i^{(G^{(k)} \cup G^{(m)})}}{\lambda_1^{(G^{(k)} \cup G^{(m)})} - \lambda_i^{(G^{(k)} \cup G^{(m)})}} \\
&\quad + O\left( \frac{1}{(n-1)^2} \right)
\end{aligned}
$$

$$
\begin{aligned}
&= \lambda_1^{(k)} + \frac{\lambda_1^{(k)}}{n-1}\left(\frac{\lambda_1^{(m)}}{\lambda_1^{(k)}-\lambda_1^{(m)}} + (p^{(k)}-1)\frac{\psi^{(k)}}{\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)}}\right.\\
&\quad \left. + (p^{(m)}-1)\frac{\psi^{(m)}}{\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)}+\psi^{(k)}-\psi^{(m)}}\right).\\
&\quad + O\left(\frac{1}{(n-1)^2}\right).
\end{aligned}
$$

Ensuite, nous obtenons :

$$
\begin{aligned}
\mathrm{E}\left(\Delta\hat{T}\right) &= \lambda_1^{(m)} + \frac{p^{(k)}-1}{n-1}\,\psi^{(k)}\left(1+\frac{\psi^{(k)}}{\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)}}\right) + \frac{p_B-1}{n-1}\,\psi^{(m)}\left(1+\frac{\psi^{(m)}}{\mathbf{b}^{(m)\prime}\mathbf{b}^{(m)}}\right)\\
&\quad - \frac{\lambda_1^{(k)}\lambda_1^{(m)}}{(n-1)(\lambda_1^{(k)}-\lambda_1^{(m)})} - \frac{p^{(k)}-1}{n-1}\,\psi^{(k)}\left(1+\frac{\psi^{(k)}}{\mathbf{b}^{(A)\prime}\mathbf{b}^{(A)}}\right)\\
&\quad - \frac{(p^{(m)}-1)\lambda_1^{(k)}\psi^{(m)}}{(n-1)(\lambda_1^{(k)}-\psi^{(m)})} + O\left(\frac{1}{(n-1)^2}\right)\\
&= \Delta T + \frac{p^{(m)}-1}{n-1}\,\psi^{(m)}\left(1+\frac{\psi^{(m)}}{\mathbf{b}^{(m)\prime}\mathbf{b}^{(m)}} - \frac{\lambda_1^{(k)}}{\lambda_1^{(k)}-\psi^{(m)}}\right)\\
&\quad - \frac{\lambda_1^{(k)}\lambda_1^{(m)}}{(n-1)(\lambda_1^{(k)}-\lambda_1^{(m)})} + O\left(\frac{1}{(n-1)^2}\right) \qquad (5.3)
\end{aligned}
$$

Comme nous allons le voir dans le paragraphe 5.3, cette approximation n'est pas souvent valable à cause d'une trop petite différence entre $\lambda_1^{(k)}$ et $\lambda_1^{(m)}$.

Considérons maintenant le cas d'une matrice de variance-covariance de deux groupes corrélés :

$$
\mathbf{\Sigma}^{(G^{(k)}\cup G^{(m)})} = \begin{pmatrix} \mathbf{b}^{(k)}\mathbf{b}^{(k)\prime} + \psi^{(k)}\mathbf{I} & \phi_{km}\,\mathbf{b}^{(k)}\mathbf{b}^{(m)\prime} \\ \phi_{km}\,\mathbf{b}^{(m)}\mathbf{b}^{(k)\prime} & \mathbf{b}^{(m)}\mathbf{b}^{(m)\prime} + \psi^{(m)}\mathbf{I} \end{pmatrix}.
$$

Nous appliquons les résultats A.1 à A.4 de l'annexe et obtenons la plus grande valeur propre :

$$
\lambda_1^{(G^{(k)}\cup G^{(m)})} = \frac{1}{2}\left(\lambda_1^{(k)} + \lambda_1^{(m)} + \sqrt{\left(\lambda_1^{(k)}-\lambda_1^{(m)}\right)^2 + 4\,\phi_{km}^2\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)}\mathbf{b}^{(m)\prime}\mathbf{b}^{(m)}}\right)
$$

avec $\lambda_1^{(k)} = \mathbf{b}^{(k)\prime}\mathbf{b}^{(k)} + \psi^{(k)}$ et $\lambda_1^{(m)} = \mathbf{b}^{(m)\prime}\mathbf{b}^{(m)} + \psi^{(m)}$. Toujours d'après les résultats de l'annexe, nous pouvons affirmer que :

$$
\lambda_2^{(G^{(k)}\cup G^{(m)})} = \frac{1}{2}\left(\lambda_1^{(k)} + \lambda_1^{(m)} - \sqrt{\left(\lambda_1^{(k)}-\lambda_1^{(m)}\right)^2 + 4\,\phi_{km}^2\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)}\mathbf{b}^{(m)\prime}\mathbf{b}^{(m)}}\right)
$$

est une autre valeur propre de $\mathbf{\Sigma}^{(G^{(k)}\cup G^{(m)})}$ et que $\psi^{(k)}$ (respectivement $\psi^{(m)}$) en est une valeur propre de multiplicité $(p^{(k)}-1)$ (respectivement $(p^{(m)}-1)$). Ainsi, l'espérance mathématique de $l_1^{(G^{(k)}\cup G^{(m)})}$ vaut :

$$\begin{aligned}
\mathrm{E}\left(l_1^{(G^{(k)}\cup G^{(m)})}\right) &= \lambda_1^{(G^{(k)}\cup G^{(m)})} + \frac{\lambda_1^{(G^{(k)}\cup G^{(m)})}}{n-1} \sum_{i=2}^{p^{(k)}+p^{(m)}} \frac{\lambda_i^{(G^{(k)}\cup G^{(m)})}}{\lambda_1^{(G^{(k)}\cup G^{(m)})} - \lambda_i^{(G^{(k)}\cup G^{(m)})}} \\
&\quad + O\left(\frac{1}{(n-1)^2}\right) \\
&= \lambda_1^{(G^{(k)}\cup G^{(m)})} + \frac{\lambda_1^{(G^{(k)}\cup G^{(m)})}\lambda_2^{(G^{(k)}\cup G^{(m)})}}{(n-1)\left(\lambda_1^{(G^{(k)}\cup G^{(m)})} - \lambda_2^{(G^{(k)}\cup G^{(m)})}\right)} \\
&\quad + \frac{(p^{(k)}-1)\lambda_1^{(G^{(k)}\cup G^{(m)})}\psi^{(k)}}{(n-1)\left(\lambda_1^{(G^{(k)}\cup G^{(m)})} - \psi^{(k)}\right)} + \frac{(p^{(m)}-1)\lambda_1^{(G^{(k)}\cup G^{(m)})}\psi^{(m)}}{(n-1)\left(\lambda_1^{(G^{(k)}\cup G^{(m)})} - \psi^{(m)}\right)} \\
&\quad + O\left(\frac{1}{(n-1)^2}\right) \\
&= \lambda_1^{(G^{(k)}\cup G^{(m)})} + \frac{1}{n-1}\frac{\lambda_1^{(k)}\lambda_1^{(m)} - \phi_{km}^2\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)}\mathbf{b}^{(m)\prime}\mathbf{b}^{(m)}}{\sqrt{\left(\lambda_1^{(k)} - \lambda_1^{(m)}\right)^2 + 4\phi_{km}^2\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)}\mathbf{b}^{(m)\prime}\mathbf{b}^{(m)}}} \\
&\quad + \frac{(p^{(k)}-1)\lambda_1^{(G^{(k)}\cup G^{(m)})}\psi^{(k)}}{(n-1)\left(\lambda_1^{(G^{(k)}\cup G^{(m)})} - \psi^{(k)}\right)} + \frac{(p^{(m)}-1)\lambda_1^{(G^{(k)}\cup G^{(m)})}\psi^{(m)}}{(n-1)\left(\lambda_1^{(G^{(k)}\cup G^{(m)})} - \psi^{(m)}\right)} \\
&\quad + O\left(\frac{1}{(n-1)^2}\right)
\end{aligned}$$

Ensuite, nous obtenons :

$$\begin{aligned}
\mathrm{E}\left(\Delta\hat{T}\right) &= \lambda_1^{(k)} + \lambda_1^{(m)} - \lambda_1^{(G^{(k)}\cup G^{(m)})} \\
&\quad + \frac{p^{(k)}-1}{n-1}\psi^{(k)}\left(1 + \frac{\psi^{(k)}}{\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)}}\right) + \frac{p^{(m)}-1}{n-1}\psi^{(m)}\left(1 + \frac{\psi^{(m)}}{\mathbf{b}^{(m)\prime}\mathbf{b}^{(m)}}\right) \\
&\quad - \frac{1}{n-1}\frac{\lambda_1^{(k)}\lambda_1^{(m)} - {\phi^{(km)}}^2\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)}\mathbf{b}^{(m)\prime}\mathbf{b}^{(m)}}{\sqrt{\left(\lambda_1^{(k)} - \lambda_1^{(m)}\right)^2 + 4{\phi^{(km)}}^2\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)}\mathbf{b}^{(m)\prime}\mathbf{b}^{(m)}}} \\
&\quad - \frac{(p^{(k)}-1)\lambda_1^{(G^{(k)}\cup G^{(m)})}\psi^{(k)}}{(n-1)\left(\lambda_1^{(G^{(k)}\cup G^{(m)})} - \psi^{(k)}\right)} - \frac{(p^{(m)}-1)\lambda_1^{(G^{(k)}\cup G^{(m)})}\psi^{(m)}}{(n-1)\left(\lambda_1^{(G^{(k)}\cup G^{(m)})} - \psi^{(m)}\right)} \\
&\quad + O\left(\frac{1}{(n-1)^2}\right)
\end{aligned}$$

$$
\begin{aligned}
= \; & \Delta T \\
& + \frac{1}{n-1}\left( (p^{(k)}-1)\psi^{(k)}\left(1+\frac{\psi^{(k)}}{\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)}}-\frac{\lambda_1^{(G^{(k)}\cup G^{(m)})}}{\lambda_1^{(G^{(k)}\cup G^{(m)})}-\psi^{(k)}}\right)\right. \\
& + (p^{(m)}-1)\psi^{(m)}\left(1+\frac{\psi^{(m)}}{\mathbf{b}^{(m)\prime}\mathbf{b}^{(m)}}-\frac{\lambda_1^{(G^{(k)}\cup G^{(m)})}}{\lambda_1^{(G^{(k)}\cup G^{(m)})}-\psi^{(m)}}\right) \\
& \left. - \frac{\lambda_1^{(k)}\lambda_1^{(m)}-{\phi^{(km)}}^2\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)}\mathbf{b}^{(m)\prime}\mathbf{b}^{(m)}}{\sqrt{\left(\lambda_1^{(k)}-\lambda_1^{(m)}\right)^2+4{\phi^{(km)}}^2\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)}\mathbf{b}^{(m)\prime}\mathbf{b}^{(m)}}}\right) \\
& + O\left(\frac{1}{(n-1)^2}\right). \qquad (5.4)
\end{aligned}
$$

## 5.3 Simulations

Les formules pour l'espérance de $\Delta\hat{T}$ ne sont que des approximations. Leur qualité dépend du nombre d'individus $n$ et de la différence entre la première et la deuxième valeur propre des matrices concernées. Pour analyser la qualité de ces approximations, nous avons effectué une étude de simulation comparant les valeurs effectives et les approximations proposées. Pour cela, un groupe, $A$, constitué de deux variables et un groupe, $B$, constitué de quatre variables sont définis avec les paramètres suivantes :

$$\mathbf{b}^{(A)} = \begin{pmatrix} 0,9 \\ 0,8 \end{pmatrix}, \quad \psi^{(A)} = 0,2$$

et

$$\mathbf{b}^{(B)} = \begin{pmatrix} 0,9 \\ 0,9 \\ 0,8 \\ 0,7 \end{pmatrix}, \quad \psi^{(B)} = 0,1 \quad \text{et} \quad \phi^{(AB)} = 0.$$

Lors de la réunion de variables ou sous-groupes d'un même groupe, la formule (5.2) est à utiliser. Elle constitue une bonne approximation pour l'espérance de $\Delta T$ puisque la première valeur propre de la matrice de variance-covariance des variables d'un même groupe est beaucoup plus grande que les autres valeurs propres. Ceci est confirmé par les résultats des simulations portant sur 20000 ensembles de données simulés selon un loi normale (voir le tableau 5.1). Même pour $n = 10$ individus, les valeurs obtenues par simulations sont assez proches de celles obtenues par la formule (5.2). Pour indiquer la dispersion des valeurs obtenues, le tableau donne également les quantiles d'ordre 5% et 95%.

| | | | résultats des simulations | | |
|---|---|---|---|---|---|
| Groupes | $\Delta T$ | $\mathrm{E}(\Delta\hat{T})$ | moyenne | $q_{0,05}$ | $q_{0,95}$ |
| n=10 | | | | | |
| $\{x_1^{(A)}\} \cup \{x_2^{(A)}\}$ | 0,2 | 0,1747 | 0,1749 | 0,0592 | 0,3379 |
| $\{x_1^{(B)}\} \cup \{x_2^{(B)}\}$ | 0,1 | 0,0882 | 0,0876 | 0,0298 | 0,1699 |
| $\{x_1^{(B)}, x_2^{(B)}\} \cup \{x_3^{(B)}, x_4^{(B)}\}$ | 0,1 | 0,0893 | 0,0898 | 0,0305 | 0,1720 |
| n=50 | | | | | |
| $\{x_1^{(A)}\} \cup \{x_2^{(A)}\}$ | 0,2 | 0,1954 | 0,1954 | 0,1345 | 0,2663 |
| $\{x_1^{(B)}\} \cup \{x_2^{(B)}\}$ | 0,1 | 0,0978 | 0,0976 | 0,0676 | 0,1320 |
| $\{x_1^{(B)}, x_2^{(B)}\} \cup \{x_3^{(B)}, x_4^{(B)}\}$ | 0,1 | 0,0980 | 0,0980 | 0,0676 | 0,1328 |

TAB. 5.1 – Réunion de variables et sous-groupes d'un même groupe : Comparaison de la valeur pour $\mathrm{E}(\Delta\hat{T})$ obtenue par la formule 5.2 à celles obtenues pour 20000 ensembles de données simulées.

Lorsque deux groupes différents (ou sous-groupes de groupes différents) sont réunis, on considère la formule (5.3). Dans ce cas, la plus grande valeur propre du nouveau groupe peut être très proche de la deuxième valeur propre ce qui rend problématique l'approximation pour l'espérance de $l_1$ et de $\Delta\hat{T}$. Pour $n = 10$, la formule (5.3) fournit même des valeurs négatives lors de la réunion des variables $x_2^{(A)}$ et $x_1^{(B)}$ ainsi que lors de la réunion du groupe $A$ avec les variables $x_1^{(B)}$ et $x_2^{(B)}$ (voir le tableau 5.2). Dans les deux cas, la différence entre les deux premières valeurs propres de la matrice de variance-covariance du nouveau groupe est égale à 0,07. Le nombre d'individus nécessaire pour utiliser l'approximation est donc relativement important. Pour cette raison, le tableau 5.2 contient aussi les valeurs obtenues avec 200 individus. Pour l'approximation de $\mathrm{E}(\Delta\hat{T})$ correspondant à la réunion du groupe $A$ avec les variables $x_1^{(B)}$ et $x_2^{(B)}$, cette taille d'échantillon est encore insuffisante.

Le bon classement dans les premières étapes de l'algorithme est très important puisque la hiérarchie ne remet pas en cause les agrégations effectuées à des étapes antérieures. Pour cela, il est intéressant de comparer les valeurs de $\Delta\hat{T}$ obtenues pour la réunion de deux variables issues d'un même groupe d'une part, et issues de différents groupes, d'autre part. Si le minimum obtenu pour la réunion de variables de différents groupes est supérieur au maximum obtenu pour la réunion de variables du même groupe, il est légitime de considérer que les premières étapes de l'algorithme hiérarchique réunissent des variables d'un même groupe. Pour $n = 10$, ceci est le cas pour 77% des ensembles de données simulées, et pour $n = 50$, pour plus de 99% des ensembles simulés. Cependant, le nombre de 50 individus (50 produits) ainsi qu'une corrélation $\phi^{(AB)}$ de nulle ne sont pas réalistes pour le profil sensoriel.

| | | | résultats des simulations | | |
|---|---|---|---|---|---|
| Groupes | $\Delta T$ | $E(\Delta\hat{T})$ | moyenne | $q_{0,05}$ | $q_{0,95}$ |
| n=10 | | | | | |
| $\{x_2^{(A)}\} \cup \{x_1^{(B)}\}$ | 0,84 | -0,3733 | 0,5173 | 0,1853 | 0,9711 |
| $\{x_2^{(A)}\} \cup \{x_4^{(B)}\}$ | 0,59 | 0,3697 | 0,4072 | 0,1455 | 0,7686 |
| $A \cup B$ | 1,65 | 1,2160 | 1,2078 | 0,4355 | 2,2767 |
| $A \cup \{x_1^{(B)}, x_2^{(B)}\}$ | 1,65 | -2.8546 | 0,9996 | 0,3660 | 1,8757 |
| n=50 | | | | | |
| $\{x_2^{(A)}\} \cup \{x_1^{(B)}\}$ | 0,84 | 0,6117 | 0,7170 | 0,5115 | 0,9419 |
| $\{x_2^{(A)}\} \cup \{x_4^{(B)}\}$ | 0,59 | 0,5495 | 0,5442 | 0,3806 | 0,7231 |
| $A \cup B$ | 1,65 | 1,5703 | 1,5641 | 1,0900 | 2,1052 |
| $A \cup \{x_1^{(B)}, x_2^{(B)}\}$ | 1,65 | 0,8226 | 1,3848 | 0,9911 | 1,8239 |
| n=200 | | | | | |
| $\{x_2^{(A)}\} \cup \{x_1^{(B)}\}$ | 0,84 | 0,7851 | 0,7917 | 0,6781 | 0,9088 |
| $\{x_2^{(A)}\} \cup \{x_4^{(B)}\}$ | 0,59 | 0,5800 | 0,5791 | 0,4878 | 0,6763 |
| $A \cup B$ | 1,65 | 1,6304 | 1,6309 | 1,3702 | 1,9117 |
| $A \cup \{x_1^{(B)}, x_2^{(B)}\}$ | 1,65 | 1,4463 | 1,5329 | 1,3124 | 1,7557 |

TAB. 5.2 – Réunion de variables ou sous-groupes de différents groupes dont les variables latentes ne sont pas corrélées : Comparaison de la valeur pour $E(\Delta\hat{T})$ obtenue par la formule 5.3 à celles obtenues pour 20000 ensembles de données simulées.

| | | | résultats des simulations | | |
|---|---|---|---|---|---|
| Groupes | $\Delta T$ | $E(\Delta\hat{T})$ | moyenne | $q_{0,05}$ | $q_{0,95}$ |
| n=10 | | | | | |
| $\{x_2^{(A)}\} \cup \{x_1^{(B)}\}$ | 0,6562 | 0,4740 | 0,4744 | 0,1678 | 0,8929 |
| $\{x_2^{(A)}\} \cup \{x_4^{(B)}\}$ | 0,5056 | 0,3816 | 0,3795 | 0,1336 | 0,7077 |
| $A \cup B$ | 1,4021 | 1,1192 | 1,0877 | 0,3842 | 2,0628 |
| $A \cup \{x_1^{(B)}, x_2^{(B)}\}$ | 1,2239 | 0,9084 | 0,8991 | 0,3189 | 1,6961 |
| n=50 | | | | | |
| $\{x_2^{(A)}\} \cup \{x_1^{(B)}\}$ | 0,6562 | 0,6227 | 0,6200 | 0,4322 | 0,8358 |
| $\{x_2^{(A)}\} \cup \{x_4^{(B)}\}$ | 0,5056 | 0,4828 | 0,4815 | 0,3337 | 0,6503 |
| $A \cup B$ | 1,4021 | 1,3502 | 1,3497 | 0,9294 | 1,8340 |
| $A \cup \{x_1^{(B)}, x_2^{(B)}\}$ | 1,2239 | 1,1659 | 1,1628 | 0,8068 | 1,5718 |

TAB. 5.3 – Réunion de variables ou sous-groupes de différents groupes dans le cas d'une structure présentant une la corrélation de 0,3 entre les variables latentes : Comparaison de la valeur pour $E(\Delta\hat{T})$ obtenue par la formule 5.4 à celles obtenues pour 20000 ensembles de données simulées.

Si les variables latentes des groupes sont corrélées, on peut penser que la classification correcte est plus difficile à obtenir. Nous avons simulé des groupes de même nature que les groupes $A$ et $B$ décrits ci-dessus, mais avec une corrélation de 0,3 entre les variables latentes $\xi^{(A)}$ et $\xi^{(B)}$. Pour $n = 10$ individus, dans 72% des cas, toutes les valeurs de $\Delta\hat{T}$ obtenu pour la réunion de variables d'un même groupe sont inférieures aux valeurs obtenues pour la réunion de variables de différents groupes. Pour $n = 50$ individus, ceci est le cas pour plus de 99% des ensembles. Par ailleurs, l'approximation de $\mathrm{E}(\Delta\hat{T})$ par la formule (5.4) correspond bien aux valeurs obtenues par simulation (voir le tableau 5.3).

Dans ce chapitre nous n'avons considéré que le comportement du critère $\hat{T}$ et $\Delta\hat{T}$ dans une étape donnée. Cependant, la partition obtenue à une étape ne dépend pas seulement de la valeur de $\Delta\hat{T}$ dans cette étape, mais aussi de toutes les étapes précédentes. Dans le chapitre suivant, la qualité de la partition obtenue à l'issue de la classification est évaluée et comparée avec les performances d'autres méthodes.

# Chapitre 6

# Comparaison avec d'autres méthodes

## 6.1 Méthodes

La méthode CLV est une méthode spécialement conçue pour la classification de variables. L'objectif de ce chapitre est de comparer cette méthode avec d'autres méthodes qui répondent au même objectif et que nous pouvons classer en quatre catégories.

1. La plupart des méthodes de classification concerne la classification d'individus et sont basées sur une matrice de dissimilarités. Il est possible d'adapter cette démarche pour la classification de variables après avoir défini des dissimilarités entre variables. Si une corrélation $r$ importante en valeur absolue signifie qu'il y a proximité entre variables sans tenir compte du signe de la corrélation, une mesure de similarité est naturellement donnée par $r^2$ et une mesure de dissimilarité par $1 - r^2$. Pour la comparaison de méthodes, nous avons choisi cette dissimilarité et avons ensuite effectué un algorithme hiérarchique largement utilisé : la méthode de Ward. Une description de cette méthode se trouve, par exemple, dans le livre de Saporta [20].
2. Le logiciel SAS intègre une méthode conçue pour la classification de variables, la procédure Varclus. Il s'agit d'une classification hiérarchique descendante basée sur la matrice de corrélation ou la matrice de variance-covariance. Au début de l'algorithme, toutes les variables sont dans un même groupe. Dans la première étape, ce groupe est séparé en deux selon les résultats d'une analyse en composantes principales avec rotation. Après une phase de consolidation, un des deux groupes résultants est séparé en deux, et ainsi de suite. Les détails de la procédure sont décrits dans le SAS/STAT User's Guide [21].
3. Deux autres possibilités pour classer des variables reposent sur l'analyse en composantes principales (ACP), voir Jolliffe [10]. Certes, l'analyse en facteurs communs et spécifiques a inspiré la formulation du modèle décrit dans le chapitre 3. Cependant, il n'est pas possible d'utiliser cette méthode si le nombre d'individus est petit par rapport au nombre de variables. Pour cette raison,

nous préférons ici considérer l'ACP. Dans le modèle factoriel avec $K$ groupes orthogonaux, les $K$ vecteurs propres de la matrice de variance-covariance $\mathbf{\Sigma}$ correspondant aux $K$ plus grandes valeurs propres peuvent être utilisés pour former les groupes de variables. En effet, dans ce modèle, la matrice de variance-covariance $\mathbf{\Sigma}$ est une matrice diagonale par blocs. Toutes les valeurs propres des blocs sont aussi des valeurs propres de $\mathbf{\Sigma}$. Soit $\lambda_1^{(k)}$ la plus grande valeur propre du bloc correspondant au groupe $G^{(k)}$ $(k = 1, \ldots, K)$. Si toutes les variances de l'erreur sont inférieures aux valeurs $\lambda_1^{(k)}$ $(k = 1, \ldots, K)$, les valeurs propres $\lambda_1^{(1)}, \ldots, \lambda_1^{(K)}$ sont alors les $K$ plus grandes valeurs propres de $\mathbf{\Sigma}$. De plus, au vecteur propre $\mathbf{v}^{(k)}$ de $\mathbf{\Sigma}^{(k)}$ (matrice de variance-covariance du groupe $G^{(k)}$) correspond un vecteur propre de $\mathbf{\Sigma}$ constitué des composantes de $\mathbf{v}^{(k)}$ et de valeurs égales à zéro. De fait, une variable sera associée au groupe $G^{(k)}$ si le coefficient qui lui correspond est non nul. Lors de la classification basée sur un échantillon, la structure n'est pas aussi claire puisque, en général, il n'y a pas de coefficients égaux à zéro. Néanmoins, il est possible de former des groupes selon l'importance des coefficients sur les $K$ premiers vecteurs propres. Une variable est affectée au groupe $G^{(k)}$ si son coefficient sur le $k^{ième}$ vecteur propre est maximal en valeur absolue.

4. Les méthodes de rotation mentionnées dans le paragraphe 2.2 pour l'analyse en facteurs communs et spécifiques peuvent aussi être utilisées dans le cadre d'une analyse en composantes principales. La rotation est effectuée sur la matrice qui contient les $K$ premiers vecteurs propres. Elle permet d'obtenir une structure facilement interprétable. La formation de groupes se fait de la même manière que décrit pour l'ACP sans rotation. Dans la comparaison des méthodes, nous intégrons une rotation orthogonale qui est souvent utilisée dans la pratique : la rotation VARIMAX (Kaiser [11]).

Nous comparons, par la suite, la méthode CLV avec ces quatre autres méthodes : l'algorithme hiérarchique de Ward effectué sur la dissimilarité $1 - r^2$, la procédure VARCLUS du logiciel SAS, l'affectation des variables sur la base de l'ACP et l'ACP avec rotation varimax. Concernant la méthode CLV, nous considérons trois possibilités : l'algorithme hiérarchique, l'algorithme de partitionnement avec la partition obtenue à partir de la coupure de l'arbre hiérarchique comme partition initiale et l'algorithme de partitionnement avec une partition aléatoire comme partition initiale.

## 6.2 Simulations

### 6.2.1 Structure des données

Les méthodes décrites ci-dessus sont comparées à l'aide d'une étude de simulations. Pour cela, des ensembles de données avec deux à quatre groupes ont été simulés. Pour chaque groupe, un modèle factoriel à un facteur a été utilisé. Ainsi,

la $l^{ième}$ variable du groupe $G^{(k)}$ se compose de la valeur de la variable latente du groupe $G^{(k)}$ et d'une erreur :

$$x_l^{(k)} = b_l^{(k)}\xi^{(k)} + \sqrt{\psi_l^{(k)}}\epsilon_l^{(k)}, \qquad \xi^{(k)}, \epsilon_l^{(k)}\ i.i.d.\ \mathcal{N}(0,1).$$

Les saturations $b_l^{(k)}$ et les variances de l'erreur $\psi_l^{(k)}$ ont été choisies selons les résultats du paragraphe 3.4 où nous avons estimé ces paramètres pour des descripteurs sensoriels. Nous avons simulé des variables avec des saturations de 0,95 (variance expliquée par le modèle : 0,9025) et des variances de l'erreur de 0,09. Ceci correspond à de "bonnes conditions". Nous avons aussi simulé des variables avec des saturations de 0,7 (variance expliquée par le modèle : 0,49) et des variances de l'erreur de 0,49. Ceci correspond aux "conditions difficiles". Les corrélations entre les variables latentes sont fixées à zéro. Sur la base de ces paramètres, nous avons défini trois structures :

*Structure 1 :* Bonnes conditions pour toutes les variables. Même si cette structure n'est pas réaliste, nous l'avons simulé afin de pouvoir comparer les méthodes dans le cas d'une très forte structure où la petite taille des échantillons représente la seule difficulté.

*Structure 2 :* Conditions difficiles pour toutes les variables. Cette structure est réaliste pour des ensembles de données issus de profils sensoriels.

*Structure 3 :* Bonnes conditions pour un groupe (deux groupes pour les ensembles avec quatre groupes) et conditions difficiles pour les autres groupes. Selon les résultats du paragraphe 3.4, nous pouvons considérer cette structure comme étant réaliste.

Pour chaque structure, des ensembles de données avec 15 ou 30 individus, 10 ou 30 variables et 2 à 4 groupes ont été simulés (voir le tableau 6.1). Concernant la répartition des variables dans les groupes, nous avons simulé des ensembles de données avec des groupes équilibrés mais aussi des ensembles de données dont les groupes ont des tailles différentes (voir le tableau 6.2).

Pour chaque ensemble de données, des groupes de variables ont été formés par chacune des méthodes décrites dans le paragraphe 6.1. Le nombre de groupes est supposé connu. Pour les ensembles simulés avec $K$ groupes, nous avons toujours retenu les partitions en $K$ groupes ($K = 2, 3, 4$). Ayant simulé des variables avec des variances similaires, il est superflu de distinguer entre l'analyse basée sur la matrice de variance-covariance et l'analyse basée sur la matrice de corrélation. Ici, il a été choisi d'utiliser la matrice de variance-covariance sauf pour l'algorithme de Ward qui est effectué sur $1 - r^2$. Il est à noter que la procédure Varclus admet différentes options. A part la décision concernant l'utilisation de la matrice de variance-covariance

| | $p = 10$ variables | $p = 30$ variables |
|---|---|---|
| $n = 15$ individus | 10000 ensembles avec 2 groupes | 10000 ensembles avec 2 groupes<br>10000 ensembles avec 3 groupes<br>10000 ensembles avec 4 groupes |
| $n = 30$ individus | 10000 ensembles avec 2 groupes | 10000 ensembles avec 2 groupes<br>10000 ensembles avec 3 groupes<br>10000 ensembles avec 4 groupes |

TAB. 6.1 – Plan des simulations.

| | 2 groupes, 10 variables | 2 groupes, 30 variables | 3 groupes | 4 groupes |
|---|---|---|---|---|
| Nombre d'ensembles | 5000 | 5000 | 5000 | 5000 |
| Groupe 1 | 5 | 15 | 10 | 7 |
| Groupe 2 | 5 | 15 | 10 | 7 |
| Groupe 3 | - | - | 10 | 8 |
| Groupe 4 | - | - | - | 8 |
| Nombre d'ensembles | 5000 | 5000 | 5000 | 5000 |
| Groupe 1 | 3 | 10 | 4 | 4 |
| Groupe 2 | 7 | 20 | 13 | 7 |
| Groupe 3 | - | - | 13 | 7 |
| Groupe 4 | - | - | - | 12 |

TAB. 6.2 – Nombre de variables par groupe.

(et non celle des corrélations) et la décision concernant le nombre de groupes, les autres options ont été choisies par défaut.

### 6.2.2 Résultats

Les structures utilisées pour les simulations sont des structures très fortes avec des groupes orthogonaux et devraient permettre de classer toutes les variables correctement. Nous avons donc, pour chaque ensemble de données et pour chaque méthode, évalué si la partition correcte a été retrouvée. Le tableau 6.3 montre le pourcentage d'ensembles simulés pour lesquelles la partition correcte a été trouvée, en fonction de la structure, du nombre d'individus et de la méthode. Les abréviations utilisées dans le tableau sont définies comme suit :

| | |
|---|---|
| CLVH : | CLV, algorithme hiérarchique |
| CLVHP : | CLV, algorithme de partitionnement,<br>initialisation : coupure de l'arbre hiérarchique |

CLVP : CLV, algorithme de partitionnement, initialisation : partition aléatoire
WARD : l'algorithme hiérarchique de Ward, effectué sur la dissimilarité $1 - r^2$
VARCLUS : la procédure VARCLUS du logiciel SAS
ACP : l'analyse en composante principales
VARIMAX : l'ACP avec rotation varimax.

| Structure | 1 | | 2 | | 3 | |
|---|---|---|---|---|---|---|
| Nombre d'individus | 15 | 30 | 15 | 30 | 15 | 30 |
| CLVH | **98,69** | **99,89** | 24,61 | 76,21 | **61,66** | **93,58** |
| CLVHP | **97,97** | **99,85** | **31,91** | **86,18** | 57,45 | **95,11** |
| CLVP | 72,15 | 72,97 | 21,13 | 56,13 | 41,02 | 70,75 |
| WARD | **99,47** | **99,99** | 26,18 | 76,28 | **66,87** | **94,28** |
| VARCLUS | **97,58** | **99,79** | **33,25** | **87,90** | 52,39 | **93,65** |
| ACP | 43,84 | 62,63 | 14,36 | 35,38 | 23,37 | 46,29 |
| VARIMAX | **97,54** | **99,85** | **31,96** | **86,35** | 50,41 | **93,31** |

TAB. 6.3 – Pourcentage d'ensembles simulés pour lesquelles la partition correcte a été trouvée.

Il ressort du tableau 6.3 que les méthodes CLVH, CLVHP, WARD, VARCLUS et VARIMAX ont des performances comparables. Pour la structure 1 qui représente les conditions idéales, elles ont presque toujours retrouvé la partition correcte, même sur de petits échantillons de 15 individus. Cependant, il ne faut pas oublier qu'il s'agit d'une situation peu réaliste, surtout pour des données issues d'un profil sensoriel. La performance relativement faible de l'algorithme de partitionnement de la méthode CLV avec une partition aléatoire comme partition initiale ne surprend pas, puisque, dans le paragraphe 4.2, nous avons démontré l'importance de la partition initiale. Quant à l'ACP, les résultats des simulations montrent l'importance d'une rotation pour la définition de groupes.

La performance des cinq meilleures méthodes est satisfaisante concernant les échantillons avec 30 individus simulés selon la structure 3. Pour toutes ces méthodes, la partition correcte est retrouvée pour plus de 93% des ensembles simulés. Avec seulement 15 individus, ce pourcentage est beaucoup moins important. L'algorithme de Ward est la méthode la plus performante, suivie de la classification hiérarchique CLV sans partitionnement. Ici, il semble que l'algorithme de partitionnement n'améliore pas la partition, mais au contraire qu'il transfère des variables bien classées dans un autre groupe auquel elles n'appartiennent pas.

Cependant, dans les conditions les plus difficiles (la structure 2), où toutes les variables ont des variances de l'erreur aussi importantes que les variances expliquées par le modèle, l'algorithme de partitionnement semble améliorer les partitions issues de la coupure de l'arbre hiérarchique. Ici, les méthodes de classification avec une phase de consolidation (CLVHP et VARCLUS) et l'ACP avec rotation VARIMAX sont les méthodes les plus performantes. Cependant, il faut noter que ces méthodes trouvent la partition correcte dans moins de 90% des cas pour les ensembles avec 30 individus et dans moins de 35% des cas pour les ensembles avec 15 individus. Cette situation étant la plus réaliste pour le profil sensoriel, nous allons inspecter les résultats des cinq meilleures méthodes plus en détail.

Nous avons calculé le critère Rand ajusté (Hubert et Arabie [9]). Le maximum de ce critère est égal à 1. Le maximum est atteint si les deux partitions sont exactement les mêmes. Le critère est autour de zéro si les deux partitions sont indépendantes. Ses valeurs dépendent du nombre d'objets (ici, de variables). Pour cette raison, nous considérons les ensembles de données avec 10 variables séparément de ceux avec 30 variables.

| | 10 variables | | | |
|---|---|---|---|---|
| Méthode | miminum | $q_{0,05}$ | $q_{0,25}$ | médiane |
| CLVH | -0,13 | -0,06 | 0,59 | 1 |
| CLVHP | -0,13 | 0,05 | 0,60 | 1 |
| WARD | -0,13 | 0,06 | 0,60 | 1 |
| VARCLUS | -0,13 | 0,06 | 0,60 | 1 |
| VARIMAX | -0,13 | 0,06 | 0,60 | 1 |
| | 30 variables | | | |
| Méthode | miminum | $q_{0,05}$ | $q_{0,25}$ | médiane |
| CLVH | -0,01 | 0,31 | 0,55 | 0,74 |
| CLVHP | -0,06 | 0,37 | 0,63 | 0,81 |
| WARD | -0,07 | 0,32 | 0,56 | 0,74 |
| VARCLUS | -0,05 | 0,36 | 0,63 | 0,81 |
| VARIMAX | -0,06 | 0,38 | 0,63 | 0,81 |

TAB. 6.4 – Critère Rand ajusté, échantillons de 15 individus, structure 2.

Concernant les échantillons de 15 individus, nous pouvons constater que la médiane est égale à 1 (cela veut dire : la partition correcte a été formée dans plus de 50% des cas) pour les structures avec 10 variables (voir le tableau 6.4). Le quartile inférieur est égale à 0,60 (0,59 pour la méthode CLVH) ce qui correspond à des partitions où seulement une variable est mal classée. Cependant, dans plus de 5% des cas, les partitions obtenues par les différentes méthodes n'ont rien en commun avec la partition correcte puisque les quantiles d'ordre 5% sont proches

de zéro. Les résultats pour les ensembles avec 30 variables sont comparables à ceux avec 10 variables. Ici, les valeurs de Rand ajusté sont plus élévés pour les méthodes CLVHP, VARCLUS et VARIMAX que pour les méthodes CLVH et WARD. Cependant, les différences ne sont pas importantes. Les valeurs médiane de 0,81 et 0,74 correspondent à des partitions avec deux variables mal classées, les quartiles inférieurs de 0,63 et 0,55 correspondent à des partitions avec quatre à cinq variables mal classées. Le mauvais groupement de quatre à cinq variables sur trente semble être un taux d'erreur acceptable. Cependant, lors de la sélection de descripteurs, la sélection d'un descripteur par groupe peut conduire à une mauvaise sélection, si un (ou plusieurs) descripteurs choisis sont mal classés. De plus, dans au moins 5% des cas, la partition obtenue est très perturbée par rapport à la partition correcte. Ainsi, le tableau 6.5 montre le croisement entre deux partitions dont la valeur de Rand ajusté est égale à 0,39, c'est-à-dire une valeur plus importante que le quantile d'ordre 5%. Ici, il y a évidemment un lien entre les deux partitions. Cependant, si la partition 1 est considérée comme la partition correcte et la partition 2 comme la partition obtenue par la classification, la structure des groupes de descripteurs n'est pas bien reflétée par la partition obtenue.

| | partition 2 | | | | |
|---|---|---|---|---|---|
| partition 1 | 4 | 0 | 0 | 3 | 7 |
| | 0 | 4 | 3 | 0 | 7 |
| | 0 | 0 | 4 | 0 | 4 |
| | 0 | 3 | 0 | 9 | 12 |
| | 4 | 7 | 7 | 12 | 30 |

TAB. 6.5 – Croisement de deux partitions de quatre groupes dont la valeur du critère Rand ajusté est égale à 0,39.

Pour les données selon la structure 2 avec 30 individus, les cinq meilleures méthodes retrouvent dans plus de 75% la bonne partition (voir le tableau 6.3). Quant aux ensembles pour lesquels la partition obtenue diffère de la partition correcte, il n'y a, en général, pas beaucoup de variables mal classées (voir le tableau 6.6). Pour les ensembles avec 10 variables, le quantile d'ordre 5% de ces cinq méthodes est égal à 0,60 ce qui correspond à une seule variable mal classée. Concernant les ensembles avec 30 variables, ce quantile est égal à 0,87 pour les méthodes CLVHP, VARCLUS et VARIMAX. Cela correspond à une seule variable mal classée. Il est égal à 0,80 pour les méthodes CLVH et WARD. Cela correspond à deux variables mal classées. Cependant, pour quelques-uns des ensembles, la partition obtenue semble être indépendante de la partition correcte comme le montrent les petites valeurs minimales du critère Rand ajusté.

| | 10 variables | | | |
|---|---|---|---|---|
| Méthode | miminum | $q_{0,05}$ | $q_{0,25}$ | médiane |
| CLVH | -0,13 | 0,60 | 1 | 1 |
| CLVHP | -0,13 | 0,60 | 1 | 1 |
| WARD | -0,12 | 0,60 | 1 | 1 |
| VARCLUS | -0,13 | 0,60 | 1 | 1 |
| VARIMAX | -0,12 | 0,60 | 1 | 1 |
| | 30 variables | | | |
| Méthode | miminum | $q_{0,05}$ | $q_{0,25}$ | médiane |
| CLVH | 0,13 | 0,80 | 0,92 | 1 |
| CLVHP | 0,13 | 0,87 | 1 | 1 |
| WARD | -0,01 | 0,80 | 0,91 | 1 |
| VARCLUS | 0,18 | 0,87 | 1 | 1 |
| VARIMAX | 0,34 | 0,87 | 1 | 1 |

TAB. 6.6 – Critère Rand ajusté, échantillons de 30 individus, structure 2.

### 6.2.3 Conclusion

Nous avons comparé trois variantes de la méthode CLV avec plusieurs méthodes pour la classification de variables. Les performances de cinq de ces méthodes se sont avérées équivalentes. Ces méthodes sont l'algorithme hiérarchique de la méthode CLV, l'algorithme de partitionnement de la méthode CLV, initialisé par la coupure de l'arbre hiérarchique, l'algorithme de Ward sur la dissimilarité $(1 - r^2)$, la procédure Varclus du logiciel SAS et l'ACP avec rotation varimax. Même dans les conditions définies avec des variances de l'erreur importantes, ces méthodes ont de bonnes performances pour des échantillons avec 30 individus. De même, pour des échantillons avec 15 individus, elles peuvent avoir des performances acceptables. Cependant, il faut être conscient que la partition obtenue par la classification de variables n'est pas nécessairement identique à la partition correcte. Il peut arriver que des variables dont la corrélation est égale à zéro (des variables appartenant à deux groupes orthogonaux) soient groupées ensemble. Il faut s'attendre à encore plus d'erreurs de groupement s'il y a des corrélations non-nulles entre les variables latentes des différents groupes.

Le fait que l'étude de simulations ait montré que CLV a une performance comparable à des méthodes connues, à savoir Varclus et l'algorithme de Ward sur la dissimilarité $(1 - r^2)$, ne minimise pas l'intérêt de cette démarche. En effet, la méthode CLV est, comparativement à Varclus, conceptuellement plus facile car elle vise à optimiser un critère bien identifié, et elle suit un schéma classique avec algorithme hiérarchique suivi d'un partitionnement, alors que dans l'algorithme de VARCLUS, une étape de consolidation a lieu à chaque étape de la hiérarchie. De plus, la mé-

thode CLV est flexible en ce sens qu'elle permet de tenir compte, le cas échéant, de données externes ce qui n'est pas le cas de Varclus et de l'algorithme de Ward sur la dissimilarité $(1 - r^2)$.

# Chapitre 7

# Détermination du nombre de groupes

Dans le chapitre précédent, nous avons supposé que le nombre de groupes est connu. En réalité, ceci n'est pas le cas. Dans la procédure CLV, il est préconisé de déterminer le nombre de classes par un examen visuel du graphique indiquant l'évolution du critère d'agrégation (voir Vigneau et Qannari [22]). Cependant, il serait souhaitable de concevoir une démarche basée sur une procédure automatique. Dans le paragraphe 7.1.1, une telle procédure est proposée en considérant des simulations basées sur des permutations. Une autre méthode qui a été proposée par Sahmer *et al.* [18] pour la détermination du nombre de groupes dans le cadre de la segmentation de consommateurs s'adapte facilement à la classification avec le critère $T$. Il s'agit d'une combinaison d'un test dit *"cluster tendency test"* et d'une succession de tests dits *"cluster validity tests"*. Cette méthode est brièvement décrite dans le paragraphe 7.1.2. Les deux méthodes sont comparées par simulations dans le paragraphe 7.2.

## 7.1 Méthodes

### 7.1.1 Procédure de permutations

Le principe général de la procédure est basée sur la comparaison de la valeur de $\Delta\hat{T}$ à chaque étape de l'algorithme hiérarchique avec les valeurs correspondantes obtenues à partir de données simulées pour lesquelles les variables ne sont pas corrélées. Dans un premier temps, nous décrivons la démarche et, dans un deuxième temps, nous la justifions en considérant le modèle factoriel.

#### Méthode

Nous procédons comme suit :

1. Classification hiérarchique des colonnes des données observées $\mathbf{X}$. $\Delta\hat{T}^{(i)}$ est la valeur du critère d'agrégation qui correspond à l'étape à l'issue de laquelle il y a $i$ groupes.

2. Répétition $B$ fois (par exemple, 100 fois) de la procédure suivante :
   (a) Permutation aléatoire des valeurs de chaque colonne de $\mathbf{X}$, indépendamment des autres colonnes. $\mathbf{X}_{PERM}$ désigne la matrice qui en résulte. Les variances des variables de $\mathbf{X}_{PERM}$ sont égales à la variance des variables de $\mathbf{X}$. Cependant, les corrélations empiriques de $\mathbf{X}_{PERM}$ correspondent aux corrélations empiriques de variables non corrélées.
   (b) Classification hiérarchique des colonnes de $\mathbf{X}_{PERM}$. $\Delta T^{(i)}_{PERM}$ représente la valeur du critère d'agrégation qui correspond à l'étape à l'issue de laquelle il y a $i$ groupes.
3. Pour chaque étape $i$ de la classification hiérarchique, calcul du quantile d'ordre 5% $q^{(i)}_{0.05}$ des $B$ valeurs de $\Delta T^{(i)}_{PERM}$.
4. La décision quant au nombre de groupes se fait selon les règles suivantes :
   (a) A la première étape de la classification, c'est à dire l'étape où, pour la première fois, deux variables sont réunies dans un groupe, si la valeur de $\Delta\hat{T}^{(p-1)}$ est supérieure à la valeur de $q^{(p-1)}_{0.05}$, nous décidons que les données observées correspondent à des réalisations de variables non corrélées. Il n'y a pas de groupes.
   (b) A la dernière étape de la classification, c'est-à-dire l'étape à l'issue de laquelle il n'y a plus qu'un groupe, si $\Delta\hat{T}^{(1)}$ est inférieur à $q^{(1)}_{0.05}$, nous décidons qu'il y a un seul groupe de variables, sinon nous passons au point suivant.
   (c) Si $\Delta\hat{T}^{(2)}$ est inférieur à $q^{(2)}_{0.05}$ à l'avant dernière étape (passage de trois à deux groupes), nous décidons qu'il y a deux groupes de variables, sinon la procédure se poursuit jusqu'à ce que $\Delta\hat{T}^{(i)}$ soit inférieur à $q^{(i)}_{0.05}$. Si cette condition est remplie pour $i = K$, alors nous décidons qu'il y a $K$ groupes.

**Justification**

Rappelons que dans le cadre du modèle factoriel, la matrice de variance-covariance du groupe $G^{(k)}$ s'écrit sous la forme : $\mathbf{\Sigma}^{(k)} = \mathbf{b}^{(k)}\mathbf{b}^{(k)\prime} + \psi^{(k)}\mathbf{I}$. Au paragraph 4.1.2, nous avons vu que, lorsqu'il existe une structure de groupes, la réunion de deux variables ou sous-groupes du groupe $G^{(k)}$ se traduit par une diminution du critère $T$ égale à $\psi^{(k)}$. Dans ce cas, la variation $\Delta T$ est moins importante que la variance de chacune des variables du groupe $G^{(k)}$. Si deux groupes $G^{(k)}$ et $G^{(l)}$ non corrélés sont réunis, le critère $\Delta T$ est égal à $\min\left(\mathbf{b}^{(k)\prime}\mathbf{b}^{(k)} + \psi^{(k)}, \mathbf{b}^{(l)\prime}\mathbf{b}^{(l)} + \psi^{(l)}\right)$. Sous des conditions peu contraignantes, ce minimum est plus important que la variance de chacune des variables des deux groupes. Or, lors de la classification d'un ensemble de variables non corrélées, à chaque étape, le critère $\Delta T$ est égal à la variance d'une variable. Pour déterminer le nombre de groupes, nous pouvons donc

comparer l'évolution du critère $\Delta T$ avec l'évolution du critère $\Delta T_0$ lors de la classification de variables non corrélées mais de mêmes variances que les variables à classer.

En pratique, nous ne connaissons pas les vraies valeurs de $\Delta T$. Nous les estimons par les valeurs du critère $\Delta \hat{T}$ lors de la classification selon la matrice de variance-covariance empirique. Une première idée pourrait être de comparer ces valeurs avec les variances empiriques ordonnées des variables. Mais cela se heurte au fait que, lors de la classification selon la matrice de variance-covariance empirique de variables non corrélées, le critère $\Delta \hat{T}$ n'est pas égal aux variances empiriques. La procédure de permutations sert à estimer les valeurs de $\Delta \hat{T}$ lors de la classification d'un ensemble de variables non corrélées.

### Illustration par un exemple

Pour illustrer la démarche, nous considérons la classification selon la matrice de variance-covariance :

$$\mathbf{\Sigma} = \begin{pmatrix} 0,9925 & 0,9025 & 0,9025 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0,9025 & 0,9925 & 0,9025 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0,9025 & 0,9025 & 0,9925 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,9800 & 0,4900 & 0,4900 & 0,4900 & 0,4900 & 0,4900 & 0,4900 \\ 0 & 0 & 0 & 0,4900 & 0,9800 & 0,4900 & 0,4900 & 0,4900 & 0,4900 & 0,4900 \\ 0 & 0 & 0 & 0,4900 & 0,4900 & 0,9800 & 0,4900 & 0,4900 & 0,4900 & 0,4900 \\ 0 & 0 & 0 & 0,4900 & 0,4900 & 0,4900 & 0,9800 & 0,4900 & 0,4900 & 0,4900 \\ 0 & 0 & 0 & 0,4900 & 0,4900 & 0,4900 & 0,4900 & 0,9800 & 0,4900 & 0,4900 \\ 0 & 0 & 0 & 0,4900 & 0,4900 & 0,4900 & 0,4900 & 0,4900 & 0,9800 & 0,4900 \\ 0 & 0 & 0 & 0,4900 & 0,4900 & 0,4900 & 0,4900 & 0,4900 & 0,4900 & 0,9800 \end{pmatrix} \tag{7.1}$$

Il est clair que cette matrice correspond à deux groupes de variables. Le groupe $G_1$ comprend trois variables avec des saturations de 0,95 et des variances de l'erreur de 0,09. Le groupe $G_2$ comprend sept variables avec des saturations de 0,7 et des variances de l'erreur de 0,49. La corrélation entre les variables latentes des deux groupes est égale à zéro. Il s'agit d'une des structures utilisées pour les simulations dans le paragraph 6.2. La figure 7.1 montre l'évolution de $\Delta T$ et $\Delta T_0$ théorique. Lors de la classification hiérarchique de dix variables non corrélées avec les mêmes variances que les dix variables étudiées, le critère $\Delta T_0$ est d'abord égal à 0,98. Lors des trois dernières étapes (qui correspondent à la réunion des sept variables de variance minimum), il est égal à 0,9925. Comparons maintenant les valeurs de $\Delta T$ avec ces valeurs. Tout d'abord, les variables du groupe $G_1$ sont réunies et le critère $\Delta T$ est égale à 0,09 ; puis les variables du groupe $G_2$ sont réunies et $\Delta T$ est égal à 0,49. Dans tous ces cas, le critère est inférieur à $\Delta T_0$. Quand les deux groupes sont réunis, le critère $\Delta T$ est égal à 2,7975 et est donc supérieur à $\Delta T_0$.

Supposons maintenant qu'il y ait une corrélation de $\phi^{(12)}$ entre les deux variables latentes de $G^{(1)}$ et $G^{(2)}$. La valeur de $\Delta T$ lors de la réunion de deux variables du même groupe est la même que pour des groupes avec des variables latentes non corrélées. Cependant, la valeur de $\Delta T$ lors de la réunion des deux groupes est moins importante que pour la réunion de deux groupes dont les variables latentes ne sont

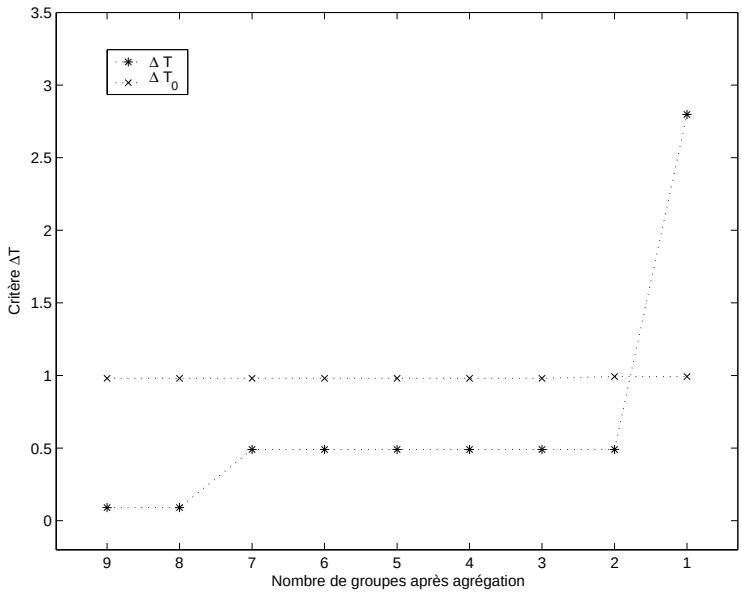


FIG. 7.1 –
Comparaison de $\Delta T$ correspondant à la classification de variables réparties dans deux groupes avec les valeurs de $\Delta T_0$ correspondant à la classification de variables non corrélées.

pas corrélées. Elle est égale à (voir le paragraphe 4.1.2) :

$$\Delta T_{oblique} = \frac{1}{2}\left(\lambda_1^{(1)} + \lambda_1^{(2)} - \sqrt{\left(\lambda_1^{(1)} - \lambda_1^{(2)}\right)^2 + 4\,\phi^{(12)2}\mathbf{b}^{(1)\prime}\mathbf{b}^{(1)}\mathbf{b}^{(2)\prime}\mathbf{b}^{(2)}}\right). \quad (7.2)$$

Avec des corrélations $\phi^{(12)}$ modérées, la valeur de $\Delta T_{oblique}$ reste néanmoins plus importante que 0,9925, la valeur maximale de $\Delta T_0$. En effet, il est aisé de montrer que si la valeur absolue de $\phi^{(12)}$ est inférieure à 0,7543, cette condition est verifiée. Ainsi, si on considère que des groupes dont la corrélation entre les variables latentes est supérieure à 0,7543 ne sont pas des groupes distincts, la méthode proposée permet d'identifier la partition en deux groupes, au moins pour la classification selon la matrice de variance-covariance de la population.

Considérons maintenant un échantillon de 30 individus distribués selon une loi normale multivariée d'espérance nulle et dont la matrice de variance-covariance est celle donnée en (7.1). Le graphique des résultats de la procédure décrite ci-dessus est donné dans la figure 7.2. La décision qui s'impose alors est de considérer qu'il existe deux groupes. Pour cet ensemble de données, le nombre de groupes est donc correctement déterminé. Une évaluation de la procédure sur la base d'une étude de simulations dans laquelle plusieurs types de structures de données sont considérés sera présentée dans le paragraphe 7.2. Lors de cette étude de simulations, la dé-

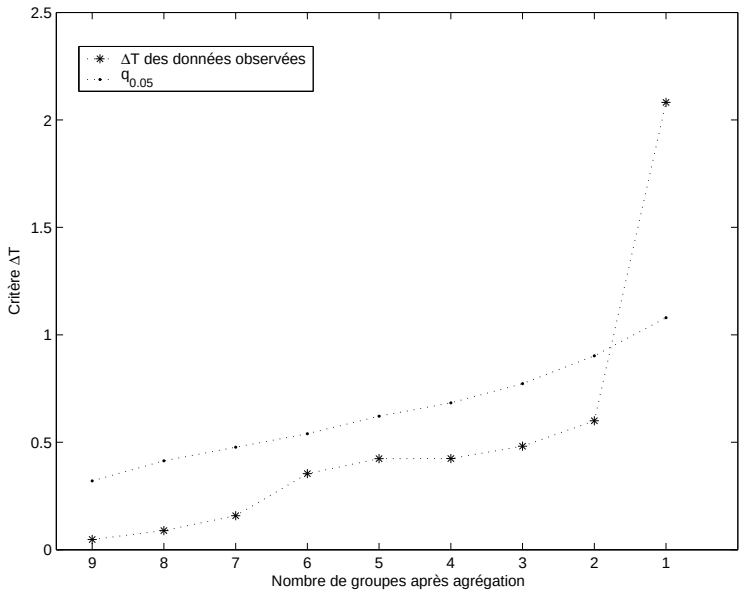


FIG. 7.2 –
Classification selon un échantillon de 30 individus : comparaison des valeurs de $\Delta\hat{T}$ avec les valeurs obtenues par la procédure de permutations.

marche par permutations sera comparée à une procédure que nous avons discutée pour une autre option de la méthode CLV (Sahmer *et al.* [18]) et que nous adaptons à la classification basée sur le critère $T$ dans le paragraphe suivant.

### 7.1.2 *Cluster tendency* et *cluster validity tests*

Cette procédure consiste en deux étapes. Dans un premier temps, un *cluster tendency test* est utilisé pour vérifier l'existence de groupes différents. Si la décision d'existence de groupes est prise, la classification hiérarchique est ensuite effectuée, et le nombre de groupes est déterminé par des *cluster validity tests*. Dans le cas de la classification selon le critère $T$, le *cluster tendency test* doit être une procédure qui aide à prendre une décision entre trois possibilités : l'unidimensionnalité des données, des données non structurées (par exemple, du bruit) et une structure autre que l'unidimensionnalité. Dans un travail antérieur (Sahmer *et al.* [16]), plusieurs méthodes pour la détermination de l'unidimensionnalité ont été comparées. Il s'est avéré que la méthode la plus efficace est un test de permutation avec la règle de Kaiser Guttman. La procédure se déroule comme suit :

1. Calcul de la matrice de corrélation empirique et ses deux plus grandes valeurs propres, $l_1$ et $l_2$.
2. Répétition $B$ fois (par exemple, 1000 fois) des étapes suivantes :

   (a) Permutation aléatoire des valeurs de chaque colonne de la matrice des données observées $\mathbf{X}$, indépendamment des autres colonnes. $\mathbf{X}_{PERM}$ désigne la matrice qui en résulte. Les corrélations empiriques de $\mathbf{X}_{PERM}$ correspondent aux corrélations empiriques de variables non corrélées.

   (b) Calcul de la matrice de corrélation empirique de $\mathbf{X}_{PERM}$ et de ses deux plus grandes valeurs propres, $l_1^*$ et $l_2^*$.

3. Evaluation de la proportion $p_1$ des valeurs de $l_1^*$ qui sont supérieures à $l_1$.
   - Si $p_1$ est supérieure à $\alpha$, par exemple $\alpha = 5\%$, nous considérons que les variables ne sont pas corrélées.
   - Sinon nous considérons qu'il y a une structure dans les données. Evaluation, ensuite, de la proportion $p_2$ des valeurs de $l_2^*$ qui sont inférieures à $l_2$. Si $p_2$ est inférieure à $\alpha$, nous supposons que les données sont unidimensionnelles, sinon nous en déduisons que la structure est autre que l'unidimensionnalité.

Si la procédure aboutit à la décision qu'une structure autre que l'unidimensionnalité existe, l'algorithme hiérarchique est executé. A chacune de ses étapes, des *cluster validity tests* sont effectués en commençant par l'avant-dernière étape de l'algorithme (passage de trois groupes à deux groupes). Ces *cluster validity tests* ont été décrits dans Sahmer *et al.* [18] pour une autre option de la méthode CLV. Ils s'utilisent d'une manière analogue pour la classification avec le critère $T$.

La procédure des *cluster validity tests* est décrite comme suit :

1. On pose $K = 2$.
2. Soient $A$ et $B$ les deux groupes qui sont réunis quand l'algorithme hiérarchique passe de $K+1$ groupes à $K$ groupes. On définie le critère $D = \frac{\Delta T}{\lambda^{(A)}+\lambda^{(B)}} = 1 - \frac{\lambda^{(A\cup B)}}{\lambda^{(A)}+\lambda^{(B)}}$.
3. Répétition $B$ fois (par exemple, 1000 fois) de la procédure suivante :
   (a) Choix aléatoire de $p^{(A)} + p^{(B)}$ variables parmi les $p$ variables initiales et répartition de ces variables sur deux groupes $A^*$ et $B^*$, de $p^{(A)}$ et de $p^{(B)}$ variables chacun.

   (b) Calcul du critère $D^*$ pour les groupes $A^*$ et $B^*$ selon la même expression qu'à l'étape 2.
4. Calcul de la proportion $q$ de valeurs de $D^*$ qui sont plus importantes que la valeur $D$.
5. Si $q \geq \alpha$ (par exemple, $\alpha = 5\%$), on prend la décision qu'il existe $K$ groupes. Si $q < \alpha$, on pose K=K+1 et on recommence à l'étape 2.

## 7.2 Comparaison par simulations

Pour évaluer la performance des deux méthodes, une étude de simulations a été effectuée. Pour cela, les mêmes structures que dans le paragraphe 6.2 pour la

comparaison de la méthode CLV avec d'autres méthodes ont été simulés. En plus des ensembles avec deux, trois et quatre groupes, nous avons simulés des données avec des variables non corrélées et des données constituées d'un seul groupe. Les variables non corrélés sont distribués *i.i.d.* $\mathcal{N}(0,1)$. Pour les données constituées d'un seul groupe, nous avons simulés les mêmes structures que pour les données avec plusieurs groupes :

*Structure 1 :* Bonnes conditions pour toutes les variables.
*Structure 2 :* Conditions difficiles pour toutes les variables.
*Structure 3 :* Bonnes conditions pour quelques variables, conditions difficiles pour d'autres variables.

Les bonnes conditions correspondent à des saturations de 0,95 et des variances de l'erreur de 0,09 et les conditions difficiles correspondent à des saturations de 0,7 et des variance de l'erreur de 0,49. Nous obtenons ainsi le plan des simulations dans le tableau 7.1.

| *Variables non corrélées :* | | |
|---|---|---|
| | $p = 10$ variables | $p = 30$ variables |
| $n = 15$ | 1000 ensembles | 1000 ensembles |
| $n = 30$ | 1000 ensembles | 1000 ensembles |
| *Pour chacune des structures 1, 2 et 3 :* | | |
| | $p = 10$ variables | $p = 30$ variables |
| $n = 15$ individus | 1000 ensembles avec 1 groupe<br>1000 ensembles avec 2 groupes | 1000 ensembles avec 1 groupe<br>1000 ensembles avec 2 groupes<br>1000 ensembles avec 3 groupes<br>1000 ensembles avec 4 groupes |
| $n = 30$ individus | 1000 ensembles avec 1 groupe<br>1000 ensembles avec 2 groupes | 1000 ensembles avec 1 groupe<br>1000 ensembles avec 2 groupes<br>1000 ensembles avec 3 groupes<br>1000 ensembles avec 4 groupes |

TAB. 7.1 – Plan des simulations.

Pour chaque ensemble de données, nous avons déterminé le nombre de groupes par la procédure de permutations et par la méthode qui combine un *cluster tendency test* et des *cluster validity tests*. Ensuite, nous avons évalué la proportion des simulations pour lesquelles le nombre correct de groupes a été déterminé. Pour les données avec des variables non corrélées, les deux méthodes prennent la décision correcte dans 95% des cas, indépendamment du nombre de variables et d'individus.

Les résultats concernant les données avec un ou plusieurs groupes sont résumés dans le tableau 7.2. Pour des petits échantillons (15 individus), la procédure de permutations est la meilleure méthode pour les structures 1 et 3, tandis que la procédure qui combine un *cluster tendency test* et des *cluster validity tests* est la meilleur méthode pour la structure 2 où toutes les variables ont des variances de l'erreur importantes. En présence de 30 individus, la procédure des permutations est la meilleure méthode pour toutes les structures.

| | $n = 15$ individus | |
|---|---|---|
| | Procédure de permutations | *Cluster tendency* et *cluster validity tests* |
| structure 1 | **95,5** | 84,3 |
| structure 2 | 45,4 | **68,3** |
| structure 3 | **85,4** | 73,4 |
| | $n = 30$ individus | |
| structure 1 | **99,7** | 90,6 |
| structure 2 | **92,0** | 82,3 |
| structure 3 | **97,9** | 83,9 |

TAB. 7.2 – Pourcentage de décisions correctes.

En raison de l'importance de la structure 2 qui est la structure la plus réaliste pour des données issues d'un profil sensoriel, nous inspectons les résultats pour cette structure plus en détail (voir le tableau 7.3). Pour les ensembles avec 15 individus, la procédure basée sur un *cluster tendency test* et des *cluster validity tests* est très performante quand il y a un seul groupe. Sa performance diminue en présence de plusieurs groupes. Pour des ensembles avec quatre groupes, la décision correcte n'est prise que dans 10% des cas. Dans cette situation, la procédure des permutations est un peu meilleure. Cependant, elle détermine correctement l'existence de quatre groupes dans 20% des cas seulement. Il est évident qu'un échantillon de 15 individus est trop petit pour déterminer le nombre de groupes d'une manière fiable. Concernant les échantillons avec 30 individus, la procédure basée sur des *cluster tendency* et *cluster validity tests* est la meilleure méthode pour les ensembles avec un ou deux groupes. Cependant, cette méthode ne réussit pas à détecter la présence de quatre groupes. La procédure de permutations est un peu plus faible concernant les ensembles simulés avec un ou deux groupes, mais réussit mieux la détection de trois ou quatre groupes, où le pourcentage de décisions correctes est encore supérieur à 85%.

En conclusion, la procédure de permutations est à préférer si la taille de l'échantillon est suffisamment importante (par exemple, 30 individus). Si, pour de

contraintes pratiques, l'échantillon est petit (par exemple, 15 individus), la méthode qui combine un *cluster tendency test* et des *cluster validity tests* peut être une alternative. Mais dans ce cas, il faut être conscient du taux important de fausses décisions.

| | | $n = 15$ individus | |
|---|---|---|---|
| Nombre de variables | Nombre de groupes simulés | Procédure de permutations | *Cluster tendency* et *cluster validity tests* |
| 10 | 1 | 65,6 | **92,1** |
| 30 | 1 | 70,5 | **98,5** |
| 10 | 2 | 33,7 | **74,7** |
| 30 | 2 | 53,7 | **78,0** |
| 30 | 3 | 28,8 | **55,7** |
| 30 | 4 | **20,2** | 10,6 |
| | | $n = 30$ individus | |
| 10 | 1 | **96,7** | **98,8** |
| 30 | 1 | **98,8** | **99,8** |
| 10 | 2 | 86,2 | **93,0** |
| 30 | 2 | **97,3** | **97,7** |
| 30 | 3 | **87,0** | 71,1 |
| 30 | 4 | **85,9** | 33,5 |

TAB. 7.3 – Structure 2 : pourcentage de décisions correctes.

# Chapitre 8

# Illustration des méthodes : étude de cas

Pour illustrer les méthodes, nous utilisons le même ensemble de données que dans le paragraphe 3.4. Ici, nous considérons l'ensemble des 23 descripteurs. Nous effectuons une classification des descripteurs sur la base du profil moyen. L'étude comprend 16 produits (variétés de café). Les simulations ayant démontré qu'une classification sur la base de 15 individus seulement ne permet pas de retrouver souvent la partition correcte, même quand les groupes sont orthogonaux, il faut être conscient des limites de l'analyse ci-dessous. La variance empirique des descripteurs varie entre 6,7 et 187,5. Il est donc judicieux de baser l'analyse sur la matrice de corrélation (et non sur la matrice de variance-covariance). Nous allons d'abord déterminer le nombre de groupes par les deux méthodes décrites dans le chapitre 7. Ensuite, nous allons comparer la partition obtenue par la méthode CLV avec les partitions obtenues par les autres méthodes qui se sont avérées performantes dans les simulations du chapitre 6.

La procédure de permutation indique l'existence de trois groupes (voir la figure 8.1). La même décision est prise par la procédure qui combine un *cluster tendency test* avec des *cluster validity tests*. La coupure de l'arbre hiérarchique en trois groupes fournit les groupes suivantes : le groupe 1 avec 14 descripteurs, le groupe 2 avec cinq descripteurs concernant l'odeur et le groupe 3 avec quatre descripteurs concernant le goût (voir la figure 8.2). Si une réduction du nombre de descripteurs est souhaitée, il faut donc prendre en compte aussi bien des descripteurs concernant le goût que des descripteurs concernant l'odeur. Par exemple, le descripteur "odeur chocolat" n'est pas dans le même groupe que le descripteur "goût chocolat". L'information de ces deux descripteurs n'est pas redondante. L'algorithme de partitionnement ne change pas la partition obtenue à partir de la coupure de l'arbre hiérarchique. La valeur du critère $T$ pour cette partition est égale à 15,95.

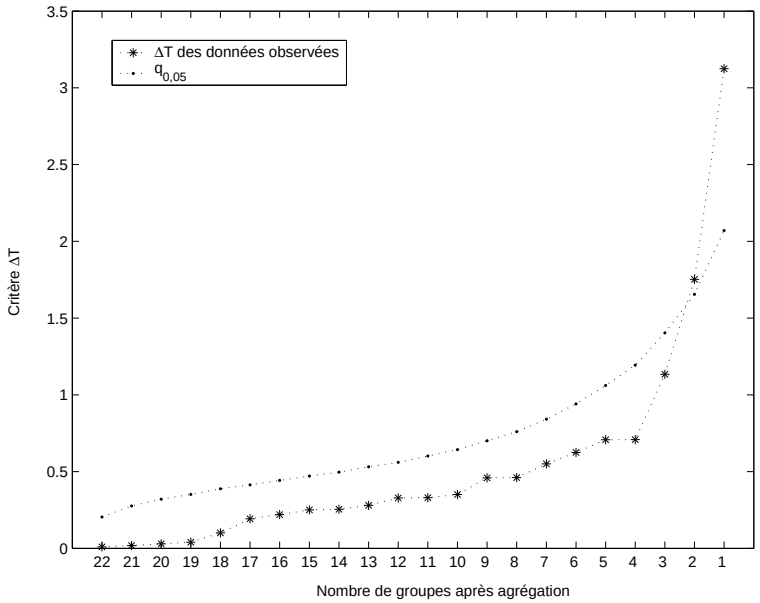


FIG. 8.1 –
Classification des données "café" : comparaison des valeurs de $\Delta T$ des données observées avec les valeurs obtenues par la procédure de permutation.

L'arbre hiérarchique de l'algorithme de Ward effectué sur la distance $1 - r^2$ (voir la figure 8.3) réunit les variables dans à peu près le même ordre que l'arbre hiérarchique de la méthode CLV. Il n'y a que des petites différences. La coupure de l'arbre en trois groupes mène à la même partition.

La partition en trois groupes obtenue par la procédure Varclus diffère légèrement de cette partition. Deux descripteurs du groupe 1 sont placés dans les deux autres groupes. Il s'agit du descripteur "odeur intensité" qui est mis dans le groupe 3, et du descripteur "odeur goudron" qui est associé au groupe 2. La valeur du critère $T$ pour cette partition est égale à 16,03 et donc très légèrement supérieure à celle obtenue par la méthode CLV. La détermination de groupes à l'aide d'une ACP avec rotation Varimax conduit à la même partition que la procédure Varclus.

La figure 8.4 montre le graphique des coefficients des variables sur les axes. Pour une meilleure comparaison, les coefficients obtenus par la méthode CLV sont normés de la même manière que ceux obtenus par l'ACP avec rotation Varimax. Pour l'analyse basée sur la matrice de corrélation, ces coefficients sont, en même temps, les corrélations avec l'axe. Après la rotation Varimax, le descripteur "odeur intensité" a des coefficients presque égaux sur les trois axes. Le descripteur "odeur goudron"

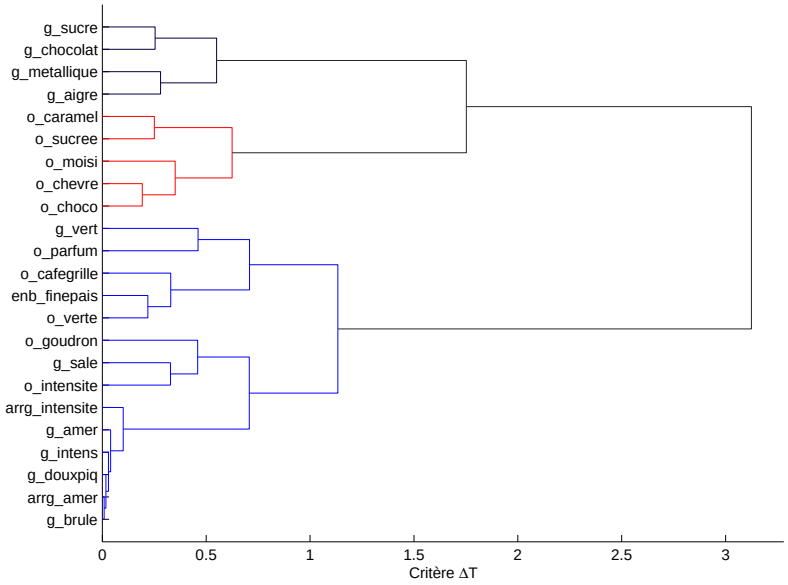


FIG. 8.2 –
Classification des 23 descripteurs sensoriels : arbre hiérarchique obtenu par la méthode CLV.

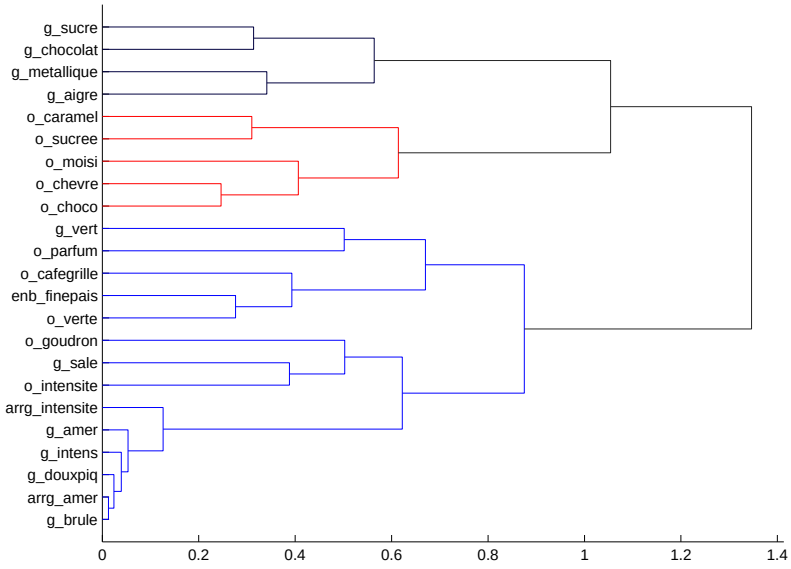


FIG. 8.3 –
Classification des 23 descripteurs sensoriels : arbre hiérarchique obtenu par l'algorithme de Ward effectué sur la distance $1 - r^2$.

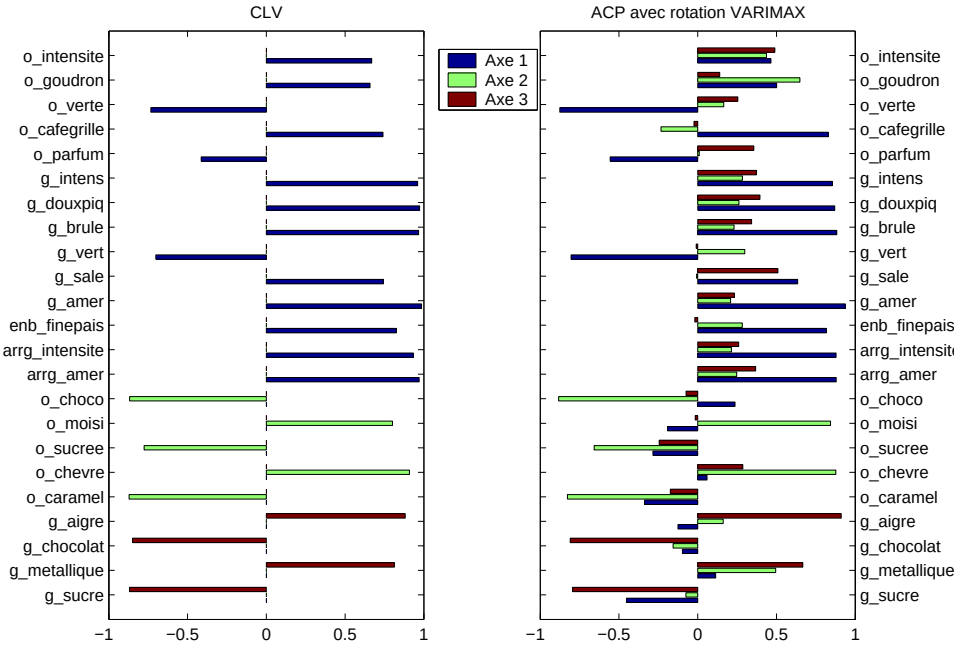


FIG. 8.4 –
Les coefficients des variables sur les axes.

a, quant à lui, des coefficients importants sur les axes 1 et 2. Ces deux descripteurs pourraient donc aussi être groupés dans le groupe 1, comme cela est le cas pour la classification avec la méthode CLV. Si nous acceptons une classification floue, le descripteur "odeur intensité" peut appartenir aux trois groupes et le descripteur "odeur goudron" peut appartenir aux groupes 1 et 2. Il y a d'autres descripteurs pour lesquelles la classification selon les résultats de la rotation Varimax n'est pas évidente, par exemple les descripteurs "goût salé", "goût metallique" et "goût sucré". Si nous considérons les corrélations des variables observées avec les variables latentes définies par la méthode CLV, nous constatons également que ces descripteurs sont proches de plusieurs groupes (voir le tableau 8.1). En conclusion, les résultats de l'ACP avec rotation Varimax et de la méthode CLV sont comparables. Nous pouvons remarquer que chercher à faire une affectation définitive à un groupe simplifie la lecture des résultats mais ne traduit pas forcément bien la situation réelle. Interpréter les coefficients des variables aux axes, comme avec l'ACP avec rotation, ou les coefficients de corrélation des variables observées aux variables latentes (calculé à posteriori, comme dans le tableau 8.1) donne une information supplémentaire.

| Descripteur | groupe 1 | groupe 2 | groupe 3 |
|---|---|---|---|
| o_intensite | **0.67** | 0.50 | 0.58 |
| o_goudron | **0.66** | 0.62 | 0.36 |
| o_verte | **-0.73** | 0.17 | 0.13 |
| o_cafegrille | **0.74** | -0.20 | 0.05 |
| o_parfum | **-0.41** | 0.00 | 0.17 |
| g_intens | **0.96** | 0.39 | 0.54 |
| g_douxpiq | **0.97** | 0.38 | 0.57 |
| g_brule | **0.97** | 0.34 | 0.52 |
| g_vert | **-0.70** | 0.22 | -0.08 |
| g_sale | **0.74** | 0.11 | 0.50 |
| g_amer | **0.99** | 0.29 | 0.42 |
| enb_finepais | **0.83** | 0.31 | 0.16 |
| arrg_intensite | **0.93** | 0.31 | 0.45 |
| arrg_amer | **0.97** | 0.36 | 0.54 |
| o_choco | 0.04 | **-0.87** | -0.28 |
| o_moisi | -0.01 | **0.80** | 0.14 |
| o_sucree | -0.43 | **-0.78** | -0.46 |
| o_chevre | 0.30 | **0.91** | 0.48 |
| o_caramel | -0.51 | **-0.87** | -0.46 |
| g_aigre | 0.14 | 0.30 | **0.88** |
| g_chocolat | -0.31 | -0.35 | **-0.85** |
| g_metallique | 0.37 | 0.57 | **0.81** |
| g_sucre | -0.62 | -0.28 | **-0.87** |

TAB. 8.1 – Corrélations des variables observées avec les variables latentes des trois groupes.

# Chapitre 9

# Conclusion et perspectives

Dans ce travail, nous avons étudié les propriétés de la méthode de classification de variables autour de composantes latentes (CLV). Cette méthode se compose d'un algorithme hiérarchique et d'un algorithme de partitionnement. Nous avons d'abord formulé un modèle statistique qui est particulièrement adapté aux données issues d'un profil sensoriel, aussi bien le profil conventionnel que le profil libre. Sur la base de ce modèle, nous avons, dans un premier temps, analysé la classification hiérarchique basée sur la matrice de variance-covariance théorique. Nous avons exprimé le critère de classification en fonction des paramètres du modèle. Nous avons pu constater que, sous des conditions peu contraignantes, l'algorithme hiérarchique forme correctement les groupes de variables. Par contre, cela n'est pas toujours le cas de l'algorithme de partitionnement dont les résultats dépendent de la partition initiale, même pour la classification selon la matrice de variance-covariance théorique et avec des groupes bien séparés.

Le comportement de la méthode CLV lors de la classification sur la base d'un échantillon a été analysé à l'aide d'une étude de simulations. Cette étude a également permis de comparer la méthode CLV à d'autres méthodes. Il s'est avéré que la performance de CLV est comparable à celle de la procédure VARCLUS du logiciel SAS, de l'algorithme hiérarchique de WARD, effectué sur la dissimilarité $(1 - r^2)$ où $r$ est le coefficient de corrélation et d'un groupement des variables basé sur les résultats d'une analyse en composantes principales avec rotation Varimax. Cela signifie que la méthode CLV est compétitive avec des méthodes connues. Cependant, il faut souligner qu'elle offre l'avantage de pouvoir prendre en compte des données externes. Il faut également noter que toutes ces méthodes ne sont pas très performantes en présence de petits échantillons et d'une grande variance de l'erreur.

Nous avons ensuite proposé et comparé par simulations deux procédures automatiques pour la détermination du nombre de groupes. La méthode la plus performante est une procédure de permutations. Néanmoins, pour obtenir des

résultats satisfaisants, il ne faut pas utiliser des échantillons de petite taille, comme cela peut être le cas dans des études de profils sensoriels.

Toutes les simulations ont été basées sur des structures avec des groupes bien séparés où les corrélations entre les variables de différents groupes sont nulles. Cela a permis d'étudier l'impact de la variance de l'erreur et de la taille des échantillons. Il serait intéressant de compléter les simulations en prenant en compte des groupes obliques, c'est à dire des corrélations non-nulles entre variables de différents groupes. Il est probable que, dans ce cas, la taille de l'échantillon doit être encore plus importante que celle qui a été constaté dans ce travail.

Un avantage de la méthode CLV est, comme cela est souligné ci-dessus, de permettre de prendre en compte des données externes. Pour cette option, il serait donc intéressant d'effectuer une analyse statistique comparable à celle effectuée dans cette thèse.

Il convient d'approfondir l'étude de la méthode CLV dans le cas où il y a des données manquantes. Un premier pas dans cette perspective a été déjà entrepris (Sahmer *et al.* [19] et Sahmer *et al.* [17]). Cependant, il faut souligner que des investigations supplémentaires ont orienté la recherche vers une méthode d'imputations qui s'appuie sur une méthode proposée par Grung et Manne [8] dans le cadre de l'analyse en composantes principales. Une méthode alternative a été proposée par Lorga da Silva *et al.* [13]. De même, une comparaison de ces différentes méthodes pourrait être intéressante.

# Bibliographie


[1] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. John Wiley and Sons, New Jersey, third edition, 2003.

[2] P. M. Bentler, Y. Kano. On the equivalence of factors and components. *Multivariate Behavioral Research*, 25 : 67–74, 1990.

[3] E. M. Braverman. Methods for the extremal grouping of parameters and the problem of determining essential factors. *Automation and Remote Control*, 1 : 108–116, 1970.

[4] I. S. Dhillon, E. M. Marcotte, U. Roshan. Diametrical clustering for identifying anti-correlated gene clusters. *Bioinformatics*, 19(13) : 1612–1619, 2003.

[5] P. Dickes. L'analyse factorielle linéaire et ses deux logiques d'application. *Psychologie française*, 41 : 9–22, 1996.

[6] Y. Escoufier. Beyond correspondence analysis. In H. H. Bock, editor, *Classification and related methods of data analysis*, pages 505–514. Elsevier Science Publishers, Amsterdam, 1988.

[7] ESN. A european sensory and consumer study : A case study on coffee. Published by the European Sensory Network, 1996.

[8] B. Grung, R. Manne. Missing values in principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 24 : 25–139, 1998.

[9] L. Hubert, P. Arabie. Comparing partitions. *Journal of Classification*, 2 : 193–218, 1985.

[10] I. T. Jolliffe. *Principal Component Analysis*. Springer-Verlag, New York, second edition, 2002.

[11] H. F. Kaiser. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3) : 187–200, 1958.

[12] D. N. Lawley. Tests of significance for the latent roots of covariance and correlation matrices. *Biometrika*, 43 : 128–136, 1956.

[13] A. Lorga Da Silva, G. Saporta, H. Bacelar-Nicolau. Missing data and imputation methods in partition of variables. In D. Banks, L. House, F. R. McMorris, P. Arabie, W. Gaul, editors, *Classification, Clustering, and Data Mining Applications*, volume 25 of *Studies in Classification, Data Analysis, and Knowledge Organization*, pages 631–637. Springer Verlag, Heidelberg-Berlin, 2004.

[14] A. E. Maxwell. Factor analysis. In S. Kotz, N. L. Johnson, editors, *Encyclopedia of statistical sciences*, volume 3, pages 2–8. John Wiley and Sons, New York, 1983.

[15] D. F. Morrison. *Multivariate statistical methods*. Thomson, Australia, United Kingdom, fourth edition, 2005.

[16] K. Sahmer, M. Hanafi, E. M. Qannari. Assessing unidimensionality within PLS path modeling framework. In M. Spiliopoulou, R. Kruse, A. Nürnberger, C. Borgelt, W. Gaul, editors, *From Data and Information Analysis to Knowledge Engineering*, volume 30 of *Studies in Classification, Data Analysis, and Knowledge Organization*, pages 222–229. Springer Verlag, Heidelberg-Berlin, 2006.

[17] K. Sahmer, E. Vigneau, E. M. Qannari. Classification de variables autour de composantes latentes en présence de valeurs manquantes. In *XXXVIèmes Journées de Statistique, Montpellier*, 2004. http ://www.sfds.asso.fr.

[18] K. Sahmer, E. Vigneau, E. M. Qannari. A cluster approach to analyze preference data : Choice of the number of clusters. *Food Quality and Preference*, 17(3-4) : 257–265, 2006.

[19] K. Sahmer, E. Vigneau, E. M. Qannari, J. Kunert. Clustering of variables with missing data : Application to preference studies. In C. Weihs, W. Gaul, editors, *Classification - The Ubiquitous Challenge*, volume 28 of *Studies in Classification, Data Analysis, and Knowledge Organization*, pages 208–215. Springer Verlag, Heidelberg-Berlin, 2005.

[20] G. Saporta. *Probalités, analyse des données, et statistique*. Editions Technip, Paris, 1990.

[21] SAS/STAT. User's guide. http ://support.sas.com/onlinedoc/913/docMainpage.jsp, SAS Institute Inc. : Cary, North Carolina, 2005.

[22] E. Vigneau, E. M. Qannari. Clustering of variables around latent components. *Communications in Statistics - Simulation and Computation*, 32(4) : 1131–1150, 2003.

[23] E. Vigneau, E. M. Qannari, K. Sahmer, D. Ladiray. Classification de variables autour de composantes latentes. *Revue de Statistique Appliquée*, LIV(1) : 27–45, 2006.

# Annexe A

# Valeurs propres d'une matrice partitionnée

Nous considérons une matrice partitionnée avec la structure :

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & \theta\ \mathbf{v}_1\mathbf{v}_2' \\ \theta\ \mathbf{v}_2\mathbf{v}_1' & \mathbf{A}_2 \end{pmatrix},$$

où $\mathbf{A}_1$ et $\mathbf{A}_2$ sont des matrices symétriques dont la plus grande valeur propre est distincte des autres valeurs propres. $\mathbf{v}_i$ est le premier vecteur propre standardisé de $\mathbf{A}_i$ avec $\mathbf{v}_i'\mathbf{v}_i = 1$ $(i = 1, 2)$. $\theta$ est un réel non nul.

**Résultat A.1** *Une valeur propre de* $\mathbf{A}$ *est égale à :*

$$\mu_1 = \frac{1}{2}\left(\lambda_1 + \lambda_2 + \sqrt{(\lambda_1 - \lambda_2)^2 + 4\,\theta^2}\right)$$

*où* $\lambda_i$ *est la plus grande valeur propre de* $\mathbf{A}_i$ *(*$i = 1, 2$*). Le vecteur propre standardisé associé à* $\mu_1$ *est égal à :*

$$\mathbf{u}_1 = \frac{1}{\sqrt{c_1^2 + \theta^2}} \begin{pmatrix} c_1\,\mathbf{v}_1 \\ \theta\ \mathbf{v}_2 \end{pmatrix}$$

*où*

$$c_1 = \frac{1}{2}\left(\lambda_1 - \lambda_2 + \sqrt{(\lambda_1 - \lambda_2)^2 + 4\,\theta^2}\right).$$

**Démonstration :** $\mu_1$ est une valeur propre de $A$ associée à $\mathbf{u}_1$ si :

$$\mathbf{A}\mathbf{u}_1 = \mu_1\mathbf{u}_1,$$

et donc si :

$$\mathbf{A}\begin{pmatrix} c_1\,\mathbf{v}_1 \\ \theta\ \mathbf{v}_2 \end{pmatrix} = \mu_1\begin{pmatrix} c_1\,\mathbf{v}_1 \\ \theta\,\mathbf{v}_2 \end{pmatrix}.$$

Ceci est équivalent à :

$$\begin{array}{ll} \text{I.} & c_1\,\mathbf{A}_1\mathbf{v}_1 + \theta^2\,\mathbf{v}_1\mathbf{v}_2'\mathbf{v}_2 = \mu_1\,c_1\,\mathbf{v}_1 \\ \text{II.} & c_1\,\theta\,\mathbf{v}_2\mathbf{v}_1'\mathbf{v}_1 + \theta\mathbf{A}_2\mathbf{v}_2 = \mu_1\,\theta\,\mathbf{v}_2 \end{array}$$

I. est vrai car :

$$\begin{aligned} & c_1\,\mathbf{A}_1\mathbf{v}_1 + \theta^2\,\mathbf{v}_1\mathbf{v}_2'\mathbf{v}_2 \\ &= \left[\lambda_1\,\frac{1}{2}\left(\lambda_1 - \lambda_2 + \sqrt{(\lambda_1-\lambda_2)^2 + 4\,\theta^2}\right) + \theta^2\right]\mathbf{v}_1 \\ &= \left[\frac{1}{2}\left(\lambda_1^2 - \lambda_1\lambda_2 + \lambda_1\sqrt{(\lambda_1-\lambda_2)^2 + 4\,\theta^2}\right) + \theta^2\right]\mathbf{v}_1 \\ &= \frac{1}{4}\Big[\lambda_1^2 - \lambda_2^2 + (\lambda_1+\lambda_2)\sqrt{(\lambda_1-\lambda_2)^2 + 4\,\theta^2} \\ &\quad + (\lambda_1-\lambda_2)\sqrt{(\lambda_1-\lambda_2)^2 + 4\,\theta^2} + (\lambda_1-\lambda_2)^2 + 4\theta^2\Big]\mathbf{v}_1 \\ &= \frac{1}{4}\left(\lambda_1 + \lambda_2 + \sqrt{(\lambda_1-\lambda_2)^2 + 4\,\theta^2}\right)\left(\lambda_1 - \lambda_2 + \sqrt{(\lambda_1-\lambda_2)^2 + 4\,\theta^2}\right)\mathbf{v}_1 \\ &= \mu_1\,c_1\,\mathbf{v}_1. \end{aligned}$$

II. est vrai car :

$$\begin{aligned} c_1\,\theta\,\mathbf{v}_2\mathbf{v}_1'\mathbf{v}_1 + \theta\mathbf{A}_2\mathbf{v}_2 &= \left[\frac{1}{2}\left(\lambda_1 - \lambda_2 + \sqrt{(\lambda_1-\lambda_2)^2 + 4\,\theta^2}\right) + \lambda_2\right]\theta\,\mathbf{v}_2 \\ &= \left[\frac{1}{2}\left(\lambda_1 + \lambda_2 + \sqrt{(\lambda_1-\lambda_2)^2 + 4\,\theta^2}\right)\right]\theta\,\mathbf{v}_2 = \mu_1\,\theta\,\mathbf{v}_2. \end{aligned}$$

**Résultat A.2** *Une autre valeur propre de* $\mathbf{A}$ *est égale à :*

$$\mu_2 = \frac{1}{2}\left(\lambda_1 + \lambda_2 - \sqrt{(\lambda_1-\lambda_2)^2 + 4\,\theta^2}\right).$$

*Le vecteur propre standardisé associé à* $\mu_2$ *est égal à :*

$$\mathbf{u}_2 = \frac{1}{\sqrt{c_2^2 + \theta^2}}\begin{pmatrix} c_2\,\mathbf{v}_1 \\ \theta\;\mathbf{v}_2 \end{pmatrix}$$

*où*

$$c_2 = \frac{1}{2}\left(\lambda_1 - \lambda_2 - \sqrt{(\lambda_1-\lambda_2)^2 + 4\,\theta^2}\right).$$

**Démonstration :** $\mu_2$ est une valeur propre de $A$ associée à $\mathbf{u}_2$ si :

$$\mathbf{A}\mathbf{u}_2 = \mu_2 \mathbf{u}_2,$$

et donc si :

$$\mathbf{A} \begin{pmatrix} c_2\,\mathbf{v}_1 \\ \theta\,\mathbf{v}_2 \end{pmatrix} = \mu_2 \begin{pmatrix} c_2\,\mathbf{v}_1 \\ \theta\,\mathbf{v}_2 \end{pmatrix}.$$

Ceci est équivalent à :

$$\begin{aligned} &\text{I.} \quad c_2\,\mathbf{A}_1\mathbf{v}_1 + \theta^2\,\mathbf{v}_1\mathbf{v}_2'\mathbf{v}_2 = \mu_2\,c_2\,\mathbf{v}_1 \\ &\text{II.} \quad c_2\,\theta\,\mathbf{v}_2\mathbf{v}_1'\mathbf{v}_1 + \theta\mathbf{A}_2\mathbf{v}_2 = \mu_2\,\theta\,\mathbf{v}_2. \end{aligned}$$

I. est vérifié car :

$$\begin{aligned} &c_2\,\mathbf{A}_1\mathbf{v}_1 + \theta^2\,\mathbf{v}_1\mathbf{v}_2'\mathbf{v}_2 \\ &= \left[\lambda_1 \frac{1}{2}\left(\lambda_1 - \lambda_2 - \sqrt{(\lambda_1-\lambda_2)^2 + 4\,\theta^2}\right) + \theta^2\right]\mathbf{v}_1 \\ &= \left[\frac{1}{2}\left(\lambda_1^2 - \lambda_1\lambda_2 - \lambda_1\sqrt{(\lambda_1-\lambda_2)^2 + 4\,\theta^2}\right) + \theta^2\right]\mathbf{v}_1 \\ &= \frac{1}{4}\Big[\lambda_1^2 - \lambda_2^2 - (\lambda_1+\lambda_2)\sqrt{(\lambda_1-\lambda_2)^2 + 4\,\theta^2} \\ &\quad -(\lambda_1-\lambda_2)\sqrt{(\lambda_1-\lambda_2)^2 + 4\,\theta^2} + (\lambda_1-\lambda_2)^2 + 4\theta^2\Big]\mathbf{v}_1 \\ &= \frac{1}{4}\left(\lambda_1 + \lambda_2 - \sqrt{(\lambda_1-\lambda_2)^2 + 4\,\theta^2}\right)\left(\lambda_1 - \lambda_2 - \sqrt{(\lambda_1-\lambda_2)^2 + 4\,\theta^2}\right)\mathbf{v}_1 \\ &= \mu_2\,c_2\,\mathbf{v}_1. \end{aligned}$$

II. est vérifié car :

$$\begin{aligned} &c_2\,\theta\,\mathbf{v}_2\mathbf{v}_1'\mathbf{v}_1 + \theta\mathbf{A}_2\mathbf{v}_2 = \left[\frac{1}{2}\left(\lambda_1 - \lambda_2 - \sqrt{(\lambda_1-\lambda_2)^2 + 4\,\theta^2}\right) + \lambda_2\right]\theta\,\mathbf{v}_2 \\ &= \left[\frac{1}{2}\left(\lambda_1 + \lambda_2 - \sqrt{(\lambda_1-\lambda_2)^2 + 4\,\theta^2}\right)\right]\theta\,\mathbf{v}_2 = \mu_2\,\theta\,\mathbf{v}_2. \end{aligned}$$

**Résultat A.3** *Soit $p_1$ le nombre de colonnes de $\mathbf{A}_1$ et $p_2$ le nombre de colonnes de $\mathbf{A}_2$. Soit $\mu_3, \ldots, \mu_{p_1+1}$ les plus faibles valeurs propres de $\mathbf{A}_1$, et $\mu_{p_1+2}, \ldots, \mu_{p_1+p_2}$, les plus faibles valeurs propres de $\mathbf{A}_2$. Alors $\mu_3, \ldots, \mu_{p_1+2}, \ldots, \mu_{p_1+p_2}$ sont aussi des valeurs propres de $\mathbf{A}$.*

**Démonstration :** Il suffit de démontrer que :

$$\mathbf{A} = \sum_{k=1}^{p_1+p_2} \mu_k \mathbf{u}_k \mathbf{u}_k' \tag{A.1}$$

avec un choix approprié de $\mathbf{u}_3, \ldots, \mathbf{u}_{p_1+p_2}$, et où $\mu_1$, $\mu_2$, $\mathbf{u}_1$ et $\mathbf{u}_2$ sont comme définis dans les résultats A.1 et A.2. Nous allons démontrer que :

$$\mathbf{A} - \mu_1 \mathbf{u}_1 \mathbf{u}_1' - \mu_2 \mathbf{u}_2 \mathbf{u}_2' = \begin{pmatrix} \mathbf{A}_1 - \lambda_1 \mathbf{v}_1 \mathbf{v}_1' & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 - \lambda_2 \mathbf{v}_2 \mathbf{v}_2' \end{pmatrix} \tag{A.2}$$

La matrice de droite est une matrice diagonale par blocs. Ses valeurs propres correspondent donc aux valeurs propres des blocs. Or, $\mathbf{v}_1$ (respectivement $\mathbf{v}_2$) est le premier vecteur propre de $\mathbf{A}_1$ (respectivement $\mathbf{A}_2$) associé à la plus grande valeur propre $\lambda_1$ (respectivement $\lambda_2$). La formule (A.1) est donc vérifiée, si (A.2) est vérifiée. Pour $k = 3, \ldots, p_1 + 1$, les vecteurs $\mathbf{u}_k$ se composent alors d'un vecteur propre de $\mathbf{A}_1$ et d'un vecteur de zéros. Pour $k = p_1 + 2, \ldots, p_1 + p_2$, les $\mathbf{u}_k$ se composent d'un vecteur de zéros et d'un vecteur propre de $\mathbf{A}_2$. Pour la démonstration de (A.2), il faut démontrer que :

$$\text{I.} \quad \mathbf{A}_1 - \mu_1 \frac{c_1^2}{c_1^2 + \theta^2} \mathbf{v}_1 \mathbf{v}_1' - \mu_2 \frac{c_2^2}{c_2^2 + \theta^2} \mathbf{v}_1 \mathbf{v}_1' = \mathbf{A}_1 - \lambda_1 \mathbf{v}_1 \mathbf{v}_1'$$

$$\text{II.} \quad \mathbf{A}_2 - \mu_1 \frac{\theta^2}{c_1^2 + \theta^2} \mathbf{v}_2 \mathbf{v}_2' - \mu_2 \frac{\theta^2}{c_2^2 + \theta^2} \mathbf{v}_2 \mathbf{v}_2' = \mathbf{A}_2 - \lambda_2 \mathbf{v}_2 \mathbf{v}_2'$$

$$\text{III.} \quad \theta\, \mathbf{v}_1 \mathbf{v}_2' - \mu_1 \frac{c_1 \theta}{c_1^2 + \theta^2} \mathbf{v}_1 \mathbf{v}_2' - \mu_2 \frac{c_2 \theta}{c_2^2 + \theta^2} \mathbf{v}_1 \mathbf{v}_2' = \mathbf{0}$$

et donc que :

$$\text{I.} \quad \mu_1 \frac{c_1^2}{c_1^2 + \theta^2} + \mu_2 \frac{c_2^2}{c_2^2 + \theta^2} = \lambda_1$$

$$\text{II.} \quad \mu_1 \frac{1}{c_1^2 + \theta^2} + \mu_2 \frac{1}{c_2^2 + \theta^2} = \frac{\lambda_2}{\theta^2}$$

$$\text{III.} \quad \frac{\mu_1 c_1}{c_1^2 + \theta^2} + \frac{\mu_2 c_2}{c_2^2 + \theta^2} = 1$$

Pour la démonstration, il est utile de définir :

$$\begin{aligned} a &:= \lambda_1 + \lambda_2 \\ b &:= \lambda_1 - \lambda_2 \\ s &:= \sqrt{b^2 + 4\theta^2}. \end{aligned}$$

On peut maintenant écrire :

$$\mu_1 = \frac{1}{2}(a + s)$$

$$\begin{aligned}
\mu_2 &= \frac{1}{2}(a-s)\\
c_1 &= \frac{1}{2}(b+s)\\
c_2 &= \frac{1}{2}(b-s).
\end{aligned}$$

Ensuite, nous avons besoin de quelques expressions :

$$\begin{aligned}
c_1^2 &= \frac{1}{4}\left(b^2+2bs+b^2+4\theta^2\right) = \frac{1}{2}b^2+\theta^2+\frac{1}{2}bs\\
c_2^2 &= \frac{1}{4}\left(b^2-2bs+b^2+4\theta^2\right) = \frac{1}{2}b^2+\theta^2-\frac{1}{2}bs\\
c_1^2c_2^2 &= \frac{1}{4}b^4+b^2\theta^2+\theta^4-\frac{1}{4}b^2\left(b^2+4\theta^2\right) = \theta^4\\
c_1^2+c_2^2 &= b^2+2\theta^2\\
\mu_1+\mu_2 &= a\\
\mu_1c_1^2 &= \frac{1}{2}(a+s)\left(\frac{1}{2}b^2+\theta^2+\frac{1}{2}bs\right)\\
&= \frac{1}{4}ab^2+\frac{1}{2}a\theta^2+\frac{1}{4}abs+\frac{1}{4}b^2s+\frac{1}{2}s\theta^2+\frac{1}{4}b^3+b\theta^2\\
\mu_2c_2^2 &= \frac{1}{2}(a-s)\left(\frac{1}{2}b^2+\theta^2-\frac{1}{2}bs\right)\\
&= \frac{1}{4}ab^2+\frac{1}{2}a\theta^2-\frac{1}{4}abs-\frac{1}{4}b^2s-\frac{1}{2}s\theta^2+\frac{1}{4}b^3+b\theta^2\\
\mu_1c_1^2+\mu_2c_2^2 &= \frac{1}{2}ab^2+a\theta^2+\frac{1}{2}b^3+2b\theta^2\\
\mu_1c_2^2 &= \frac{1}{2}(a+s)\left(\frac{1}{2}b^2+\theta^2-\frac{1}{2}bs\right)\\
&= \frac{1}{4}ab^2+\frac{1}{2}a\theta^2-\frac{1}{4}abs+\frac{1}{4}b^2s+\frac{1}{2}s\theta^2-\frac{1}{4}b^3-b\theta^2\\
\mu_2c_1^2 &= \frac{1}{2}(a-s)\left(\frac{1}{2}b^2+\theta^2+\frac{1}{2}bs\right)\\
&= \frac{1}{4}ab^2+\frac{1}{2}a\theta^2+\frac{1}{4}abs-\frac{1}{4}b^2s-\frac{1}{2}s\theta^2-\frac{1}{4}b^3-b\theta^2\\
\mu_1c_2^2+\mu_2c_1^2 &= \frac{1}{2}ab^2+a\theta^2-\frac{1}{2}b^3-2b\theta^2\\
\mu_1c_1 &= \frac{1}{4}ab+\frac{1}{4}as+\frac{1}{4}bs+\frac{1}{4}b^2+\theta^2\\
\mu_1c_1c_2^2 &= \left(\frac{1}{4}ab+\frac{1}{4}as+\frac{1}{4}bs+\frac{1}{4}b^2+\theta^2\right)\left(\frac{1}{2}b^2+\theta^2-\frac{1}{2}bs\right)\\
&= \frac{1}{8}ab^3+\frac{1}{4}ab\theta^2-\frac{1}{8}ab^2s+\frac{1}{8}ab^2s+\frac{1}{4}as\theta^2-\frac{1}{8}ab(b^2+4\theta^2)
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{8}b^3s+\frac{1}{4}bs\theta^2-\frac{1}{8}b^2(b^2+4\theta^2)\\
&+\frac{1}{8}b^4+\frac{1}{4}b^2\theta^2-\frac{1}{8}b^3s+\frac{1}{2}b^2\theta^2+\theta^4-\frac{1}{2}bs\theta^2\\
&= -\frac{1}{4}ab\theta^2+\frac{1}{4}as\theta^2-\frac{1}{4}bs\theta^2+\frac{1}{4}b^2\theta^2+\theta^4\\
&= \left(-\frac{1}{4}ab+\frac{1}{4}as-\frac{1}{4}bs+\frac{1}{4}b^2+\theta^2\right)\theta^2\\
\mu_1c_1(c_2^2+\theta^2) &= \left(\frac{1}{2}as+\frac{1}{2}b^2+2\theta^2\right)\theta^2\\
\mu_2c_2 &= \frac{1}{4}ab-\frac{1}{4}as-\frac{1}{4}bs+\frac{1}{4}b^2+\theta^2\\
\mu_2c_2c_1^2 &= \left(\frac{1}{4}ab-\frac{1}{4}as-\frac{1}{4}bs+\frac{1}{4}b^2+\theta^2\right)\left(\frac{1}{2}b^2+\theta^2+\frac{1}{2}bs\right)\\
&= \frac{1}{8}ab^3+\frac{1}{4}ab\theta^2+\frac{1}{8}ab^2s-\frac{1}{8}ab^2s-\frac{1}{4}as\theta^2-\frac{1}{8}ab(b^2+4\theta^2)\\
&\quad -\frac{1}{8}b^3s-\frac{1}{4}bs\theta^2-\frac{1}{8}b^2(b^2+4\theta^2)\\
&\quad +\frac{1}{8}b^4+\frac{1}{4}b^2\theta^2+\frac{1}{8}b^3s+\frac{1}{2}b^2\theta^2+\theta^4+\frac{1}{2}bs\theta^2\\
&= -\frac{1}{4}ab\theta^2-\frac{1}{4}as\theta^2+\frac{1}{4}bs\theta^2+\frac{1}{4}b^2\theta^2+\theta^4\\
&= \left(-\frac{1}{4}ab-\frac{1}{4}as+\frac{1}{4}bs+\frac{1}{4}b^2+\theta^2\right)\theta^2\\
\mu_2c_2(c_1^2+\theta^2) &= \left(-\frac{1}{2}as+\frac{1}{2}b^2+2\theta^2\right)\theta^2
\end{aligned}
$$

Démonstration de I. :

$$
\begin{aligned}
\mu_1\frac{c_1^2}{c_1^2+\theta^2}+\mu_2\frac{c_2^2}{c_2^2+\theta^2} &= \frac{\mu_1c_1^2c_2^2+\mu_1c_1^2\theta^2+\mu_2c_1^2c_2^2+\mu_2c_2^2\theta^2}{c_1^2c_2^2+c_1^2\theta^2+c_2^2\theta^2+\theta^4}\\
&= \frac{(\mu_1+\mu_2)\,c_1^2c_2^2+(\mu_1c_1^2+\mu_2c_2^2)\,\theta^2}{c_1^2c_2^2+(c_1^2+c_2^2)\,\theta^2+\theta^4}\\
&= \frac{a\theta^4+\frac{1}{2}ab^2\theta^2+a\theta^4+\frac{1}{2}b^3\theta^2+2b\theta^4}{\theta^4+b^2\theta^2+2\theta^4+\theta^4}\\
&= \frac{2(a+b)\theta^2+\frac{1}{2}(a+b)b^2}{b^2+4\theta^2}\\
&= \frac{(a+b)\left(2\theta^2+\frac{1}{2}b^2\right)}{b^2+4\theta^2}\\
&= \frac{2\lambda_1\left(2\theta^2+\frac{1}{2}b^2\right)}{b^2+4\theta^2}\\
&= \lambda_1.
\end{aligned}
$$

Démonstration de II. :

$$\begin{aligned}
\mu_1\frac{1}{c_1^2+\theta^2}+\mu_2\frac{1}{c_2^2+\theta^2} &= \frac{\mu_1c_2^2+\mu_1\theta^2+\mu_2c_1^2+\mu_2\theta^2}{c_1^2c_2^2+c_1^2\theta^2+c_2^2\theta^2+\theta^4}\\
&= \frac{\mu_1c_2^2+\mu_2c_1^2+(\mu_1+\mu_2)\theta^2}{c_1^2c_2^2+\left(c_1^2+c_2^2\right)\theta^2+\theta^4}\\
&= \frac{\frac{1}{2}ab^2+a\theta^2-\frac{1}{2}b^3-2b\theta^2+a\theta^2}{\theta^4+\left(b^2+2\theta^2\right)\theta^2+\theta^4}\\
&= \frac{\frac{1}{2}b^2(a-b)+2\theta^2(a-b)}{\theta^2\left(b^2+4\theta^2\right)}\\
&= \frac{(a-b)\left(\frac{1}{2}b^2+2\theta^2\right)}{\theta^2\left(b^2+4\theta^2\right)}\\
&= \frac{2\lambda_2\left(\frac{1}{2}b^2+2\theta^2\right)}{\theta^2\left(b^2+4\theta^2\right)}\\
&= \frac{\lambda_2}{\theta^2}.
\end{aligned}$$

Démonstration de III. :

$$\begin{aligned}
\frac{\mu_1c_1}{c_1^2+\theta^2}+\frac{\mu_2c_2}{c_2^2+\theta^2} &= \frac{\mu_1c_1\left(c_2^2+\theta^2\right)+\mu_2c_2\left(c_1^2+\theta^2\right)}{c_1^2c_2^2+c_1^2\theta^2+c_2^2\theta^2+\theta^4}\\
&= \frac{\left(\frac{1}{2}as+\frac{1}{2}b^2+2\theta^2\right)\theta^2+\left(-\frac{1}{2}as+\frac{1}{2}b^2+2\theta^2\right)\theta^2}{\theta^4+\left(b^2+2\theta^2\right)\theta^2+\theta^4}\\
&= \frac{\left(b^2+4\theta^2\right)\theta^2}{\theta^2\left(b^2+4\theta^2\right)}\\
&= 1.
\end{aligned}$$

**Résultat A.4** *La plus grande valeur propre de* $\mathbf{A}$ *est égale à* $\mu_1$.

**Démonstration :** Selon les résultats A.1 à A.3, les valeurs propres de $\mathbf{A}$ sont égales à $\mu_1, \mu_2, \ldots, \mu_{p_1+p_2}$. D'abord il est évident que $\mu_1$ est plus grande que $\mu_2$ car

$$\begin{aligned}
\mu_1 &= \frac{1}{2}\left(\lambda_1+\lambda_2+\sqrt{\left(\lambda_1-\lambda_2\right)^2+4\theta^2}\right)\\
&> \frac{1}{2}\left(\lambda_1+\lambda_2-\sqrt{\left(\lambda_1-\lambda_2\right)^2+4\theta^2}\right)=\mu_2.
\end{aligned}$$

Ensuite,

$$\mu_1 > \frac{1}{2}\left(\lambda_1+\lambda_2+\sqrt{\left(\lambda_1-\lambda_2\right)^2}\right)$$

$$\begin{aligned} &= \frac{1}{2}\left(\lambda_1 + \lambda_2 + \max(\lambda_1, \lambda_2) - \min(\lambda_1, \lambda_2)\right) \\ &= \max(\lambda_1, \lambda_2). \end{aligned}$$

Or, $\lambda_1$ est la plus grande valeur propre de $\mathbf{A}_1$, et $\mu_3, \ldots, \mu_{p_1+1}$ sont les autres valeurs propres de $\mathbf{A}_1$. Nous avons donc

$$\mu_1 > \lambda_1 > \mu_i, \qquad i = 3, \ldots, p_1 + 1.$$

Avec une argumentation anologue sur les valeurs propres de $\mathbf{A}_2$, nous obtenons :

$$\mu_1 > \lambda_2 > \mu_i, \qquad i = p_1 + 2, \ldots, p_1 + p_2.$$

La valeur propre $\mu_1$ est donc la plus grande valeur propre de $\mathbf{A}$.

# Abstract

In this work, the properties of the method of clustering of variables around latent components (CLV) are investigated. A statistical model is postulated. This model is especially appropriate for sensory profiling data. It sheds more light on the method CLV. The clustering criterion can be expressed in terms of the parameters of the model. It is shown that, under weak conditions, the hierarchical algorithm of CLV finds the correct partition while the partitioning algorithm depends on the partition used as a starting point. Furthermore, the performance of CLV on the basis of a sample is investigated by means of a simulation study. It is shown that this performance is comparable to the performance of known methods such as the procedure Varclus of the software SAS. Finally, two methods for determining the number of groups are proposed and compared.



**Keywords :** Clustering of variables, principal component analysis, factor analysis, sensory analysis

# Résumé

Dans ce travail, les propriétés de la méthode de classification de variables autour de composantes latentes (CLV) sont étudiées. Un modèle statistique pour cette méthode est formulé. Ce modèle est particulièrement adapté aux données issues d'un profil sensoriel. Il permet de jeter un nouvel éclairage sur la méthode CLV. Le critère de classification s'écrit en fonction des paramètres du modèle. Il est démontré que, sous des conditions peu contraignantes, l'algorithme hiérarchique retrouve correctement les groupes de variables tandis que l'algorithme de partitionnement dépend de l'initialisation. Le comportement de la méthode CLV lors de la classification sur la base d'un échantillon est analysé à l'aide d'une étude de simulations. Il s'avère que la performance de CLV est comparable à celle de méthodes connues telles que la méthode Varclus du logiciel SAS. Finalement, deux procédures automatiques pour la détermination du nombre de groupes sont proposées et comparées.



**Mots clés :** Classification de variables, analyse en composantes principales, analyse en facteurs communs et spécifiques, analyse sensorielle



# Zusammenfassung

In der vorliegenden Arbeit werden die Eigenschaften der Methode CLV zum Clustern von Variablen untersucht. Ein statistisches Modell für diese Methode, das für sensorische Profildaten angemessen ist, wird formuliert. Das Clusterkriterium kann mithilfe der Parameter des Modells ausgedrückt werden. Es wird gezeigt, dass der hierarchische Algorithmus der Methode CLV unter schwachen Bedingungen die richtige Gruppierung der Variablen findet, während die Ergebnisse des partitionierenden Algorithmus von der Initialisierung abhängen. Die Leistungsfähigkeit der Methode CLV beim Clustern basierend auf einer Stichprobe wird mittels einer Simulationsstudie untersucht. Es zeigt sich, dass die Leistungsfähigkeit der Methode CLV mit der von bekannteren Methoden wie zum Beispiel der Prozedur Varclus des Programmpaketes SAS vergleichbar ist. Schließlich werden zwei Verfahren vorgestellt und verglichen, mit deren Hilfe eine automatische Bestimmung der Gruppenanzahl möglich ist.



**Schlagwörter :** Clustern von Variablen, Hauptkomponentenanalyse, Faktorenanalyse, sensorische Analyse