

**THESE présentée  
pour l'obtention  
du  
DIPLOME de DOCTEUR de 3<sup>e</sup> CYCLE  
à  
L'UNIVERSITE DE PARIS VI**

**spécialité: STATISTIQUE MATHEMATIQUE  
mention:**

**par M<sup>r</sup> Alain Marcel DEQUIER**

**Sujet de la thèse: Contribution à l'étude des Tables de  
Contingence entre Trois Caractères.**

**Soutenu le 15.10.1973 devant la Commission composée de :**

**M<sup>r</sup> DUGUE D. President**

**M<sup>r</sup> BENZECRI J.P. Examineur**

**M<sup>r</sup> BONITZER J. //**

### Remerciements

A Messieurs les Professeurs :

D. DUGUE, qui a bien voulu accepter de présider mon jury,  
J.P. BENZECRI et J. BONITZER qui ont orienté et soutenu mes recherches,

A l'ECOLE NATIONALE DES PONTS ET CHAUSSEES et à la FACULTE DES SCIENCES de PARIS, pour avoir rendu possible ce travail,

A tous ceux qui, à l'INSTITUT DE RECHERCHE DES TRANSPORTS, m'ont témoigné compréhension, aide et encouragements.

SOMMAIRE

|   | Page |
|---|------|
| Introduction  | 1    |
| <br>  |      |
| I - Décomposition de la fonction "écart"  | 3    |
| I1 - Problèmes étudiés et notations   |      |
| I2 - L'hypothèse d'indépendance et la variable $\chi^2$                                   |      |
| I3 - Analyse de variance sur la fonction écart  |      |
| I31 - Théorème de décomposition   |      |
| I32 - Cas binaire   |      |
| I33 - Cas ternaire  |      |
| I4 - Autres décompositions  |      |
| I41 - Modèle fondamental pour l'analyse<br>factorielle à trois modes (L.R. TUCKER)        |      |
| I42 - L'analyse des correspondances ternaires   |      |
| I43 - Décomposition canonique d'une table à<br>trois entrées (J.D. CARROLL et J.J. CHANG) |      |
| II - Variances partielles et sous-hypothèses  | 51   |
| III - Les tests du $\chi^2$ sur les tables de contingence                                 | 63   |
| III1 - Présentation   |      |
| III2 - Approche et résultats théoriques   |      |
| III3 - Application aux tables de contingence  |      |
| III4 - La non-intéraction   |      |
| III5 - Remarque sur l'addition des hypothèses   |      |
| IV - Théorie de l'information et tables de contingence                                    | 82   |
| IV1 - Définitions, propriétés et exemples   |      |
| IV2 - Application aux contingences entre trois<br>caractères                              |      |

→  
cb  
w  
BEW

classique  
(→ MV)

F-1  
χ<sup>2</sup>  
sol. 7.2.1.  
d. 2

V - Comparaison des trois approches 101

V1 - Equivalence au voisinage de l'indépendance

V2 - Deux exemples numériques

VI - Le cas particulier de l'indépendance conditionnelle 116

V11 - Un modèle de prévision pour une contingence binaire

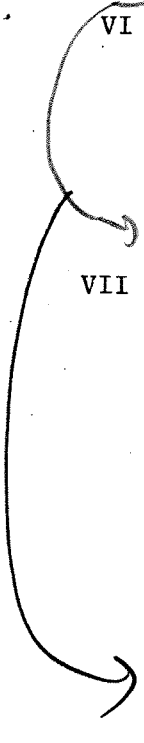
VI2 - Analyse factorielle conditionnelle

VI3 - Extension à plus de trois caractères

VII - Exemples de modèles de prévision et d'analyses conditionnelles 129

Bibliographie

on  
gr  
0



$$P_{ijk}^k = (P_{ik}^k / p_k) (P_{jk}^k / p_k) = P_i^k P_j^k$$

$$\hat{r}_{ij} = \sum_k (r_{ik} r_{jk}) / r_k$$

à comparer à  $\hat{r}_{ij}$

## INTRODUCTION

\* Les recensements statistiques modernes portent très souvent sur des variables qualitatives et mettent en relation des ensembles de caractères. Ainsi on préférera demander à une personne sa catégorie socio-professionnelle précise plutôt qu'une indication quantitative sur son revenu. La variable qualitative peut décrire la réalité avec plus de sens et de précision qu'une mesure et ceci dans beaucoup de domaines. De plus certaines grandeurs ne souffrent pas la qualification, l'utilisation d'une grandeur numérique pour représenter la visibilité des automobilistes n'a jamais pu rendre compte des différentes conditions atmosphériques existantes et encore moins de leurs effets sur la circulation.

\* L'introduction de variables qualitativement différentes dans les études statistiques a conduit à multiplier les cas à étudier pour décrire un certain phénomène dans toutes les circonstances possibles. La comparaison entre les caractères se traduit alors par des tests sur les résultats d'études séparées (une régression pour chaque modalité par exemple). Lorsque l'on ne peut pas conclure à une différence significative sur les résultats on en déduit que les caractères sont indiscernables ou ont un même effet sur les critères quantitatifs employés.

\* Les tests du  $\chi^2$  de Pearson et l'analyse factorielle appliquée aux tables de contingences permettent de faire directement des comparaisons globales entre les modalités de deux caractères. Soit sous forme d'un test concernant la validité d'une certaine hypothèse soit par une méthode factorielle permettant d'obtenir une visualisation des liaisons entre les modalités.

L'analyse factorielle conduit à une explication du résultat d'un test de Pearson : en effet lorsque l'hypothèse d'indépendance entre les ensembles est rejetée par le test, elle fait apparaître un ensemble de facteurs qui s'appuient sur les modalités des deux caractères mis en correspondances et qui expliquent la liaison entre les caractères étudiés.

\* Lorsque nous avons affaire à trois ensembles de modalités il est toujours possible d'utiliser un test du  $\chi^2$ , on peut aussi juger des indépendances totales, partielles et même conditionnelles. L'analyse factorielle propose alors une description entre un ensemble et la réunion ou le produit (au sens des associations possibles entre modalités) des deux autres. Les méthodes d'étude statistiques de variables quantitatives fournissent une solution pour l'étude globale des influences de groupes de facteurs sur une grandeur numérique. Elles procèdent par décomposition d'une variance totale en différents effets ou interactions entre les facteurs étudiés. Les  $\chi^2$  de dépendances introduits par K Pearson et utilisés par M. S. ROY (l'approche sous l'angle de la théorie des tests de M. S. ROY est résumée au chapitre III) apparaissent aussi comme la variance d'une fonction que nous appellerons "écart" ; l'objet de cette étude sera de décomposer cette grandeur pour mettre en évidence des effets et des interactions entre caractères.

\* Le problème des tableaux tridimensionnels a été abordé par la théorie de l'information, nous verrons le lien étroit qui existe entre tous les résultats. La question des tableaux de contingence conditionnels sera approfondie car elle permet de traiter les relations de causalité qui peuvent apparaître entre les ensembles étudiés.

## I - DECOMPOSITION DE LA FONCTION "ECART"

### II - Problèmes étudiés et notations.

On désire étudier des tableaux provenant de sondages aléatoires dans une population importante ou des tableaux de comptage portant sur des réalisations aléatoires.

On disposera toujours d'individus ou de phénomènes ayant au plus trois caractères,  $i$ ,  $j$  et  $k$ ; ces caractères pouvant prendre plusieurs modalités. Nous résumerons ceci par la notation  $\{i \in I\}$ ,  $\{j \in J\}$ ,  $\{k \in K\}$ .

Le nombre d'individus correspondant aux modalités  $(i, j, k)$  est  $n_{ijk}$ .

On définit :

$$n_{i.} = \sum_k n_{ijk}$$
$$n_{.jk} = \sum_i n_{ijk}$$
$$n_{.j.} = \sum_k n_{ijk}$$

Ce sont les tableaux de contingence marginaux.

Ainsi que :

$$n_i = \sum_{j,k} n_{ijk}$$
$$n_j = \sum_{i,k} n_{ijk}$$
$$n_k = \sum_{i,j} n_{ijk}$$

qui représentent les marges sur les trois ensembles (ou trois caractères)  $I$ ,  $J$  et  $K$ .

$$n = \sum_{ijk} n_{ijk} \text{ est l'effectif total}$$

De tous ces tableaux de comptage on passe aux fréquences observées par division par l'effectif total.

I2 - L'hypothèse d'indépendance et la variable  $\chi^2$

La non-influence entre les trois ensembles peut s'exprimer par l'hypothèse d'indépendance en probabilité entre les variables aléatoires  $X_I, Y_J, Z_K$  application de l'ensemble des phénomènes ou des individus intervenant dans les comptages, sur les ensembles I, J et K des modalités des trois caractères étudiés :

Proba ( $X_{I=i}; Y_{J=j}; Z_{K=k}$ ) = Proba ( $X_{I=i}$ ) x Proba ( $Y_{J=j}$ ) x Proba ( $Z_{K=k}$ )  
ce qu'on peut écrire plus simplement :

$$P_{ijk} = P_i \times P_j \times P_k$$

Cette hypothèse signifie qu'il n'y a aucune liaison entre les caractères et que la probabilité de tirer dans un sondage élémentaire un individu ayant les modalités i, j, k est simplement proportionnelle à l'importance réelle de ces modalités dans la population totale.

Pour juger de la vraisemblance de cette hypothèse on construit à partir de l'observation constituée par le tableau de comptage la variable numérique suivante :

$$\chi^2 = \sum_{i,j,k} \left( n_{ijk} - \frac{n_i \cdot n_j \cdot n_k}{n^2} \right)^2 / \frac{n_i \cdot n_j \cdot n_k}{n^2}$$



Cette variable suit une loi de probabilité connue lorsque l'hypothèse d'indépendance entre les ensembles est vérifiée, c'est-à-dire lorsque le tirage aléatoire des individus se fait bien selon le schéma probabiliste correspondant à cette hypothèse.

Dans tous les cas la variable numérique  $\chi^2$  apparaît comme la variance d'une certaine fonction d'écart par rapport à la densité de probabilité résultant de l'indépendance.

Faisons apparaître les fréquences observées :

$$f_{ijk} = \frac{n_{ijk}}{n} \quad \text{fréquence observée du groupe (i, j, k)}$$

$$f_{ijk}^* = \frac{n_i n_j n_k}{n^3} = \frac{n_i}{n} \times \frac{n_j}{n} \times \frac{n_k}{n} = f_i^* f_j^* f_k^*$$

fréquence théorique du groupe (i, j, k) déduite de l'hypothèse en estimant les probabilités marginales par des fréquences ( $n_i p_j p_k =$  espérance ( $n_{ijk}$ ) est estimée par  $n_i n_j n_k / n^3$ )

$$\chi^2 = n \sum_{ijk} (f_{ijk} - f_{ijk}^*)^2 / f_{ijk}^*$$

$$= n \sum_{ijk} \left( \frac{f_{ijk}}{f_{ijk}^*} - 1 \right)^2 \times f_{ijk}^* = \text{Var} \left( \frac{f_{ijk}}{f_{ijk}^*} \right)$$

En effet l'espérance du rapport des fréquences par rapport à la densité estimée  $f_{ijk}^*$  est bien 1.

Cette variable numérique apparait donc comme la variance d'un rapport de deux distributions de fréquence, l'une correspondant à la fréquence observée et l'autre à la fréquence résultant d'une hypothèse probabiliste et d'une estimation. La variance est calculée à l'aide de cette seconde distribution de fréquence qui joue le rôle de référence par rapport à laquelle on calcule des écarts.

I3 - "Analyse de Variance" sur la fonction Ecart

I31 - Théorème de décomposition

Considérons des ensembles d'indices I, J, K comportant respectivement, card I, card J, card K éléments et des fonctions réelles définies sur ces ensembles  $f_I = \{f_i \mid i \in I\}$ . Ces fonctions forment l'espace  $R^I$  de dimension : Card I. (Les card I valeurs de  $f_I$  peuvent être choisies indépendamment les unes des autres et toute fonction de  $R^I$  se décompose sur la base des card I fonctions  $\delta_i = \begin{cases} 0 & \text{si } i' \neq i \\ 1 & \text{si } i' = i \end{cases}$  linéairement indépendantes.

A l'espace  $R^I$  nous associons une métrique de rang card I, définie par une matrice diagonale  $m_I$  dont toutes les valeurs sont strictement positive ( $m_i > 0, \forall i$ ).

Alors à l'ensemble produit  $(IJK) = I \times J \times K$  des triplets d'indices est associé l'espace de fonctions réelles  $f_{IJK} = \{f_{ijk} \mid i \in I, j \in J, k \in K\}$  et une métrique de rang Card I x Card J x card K que l'on notera  $m_I \otimes m_J \otimes m_K$  et dont les éléments sont les produits  $m_i \times m_j \times m_k$

Le théorème de décomposition des fonctions sur un ensemble produit IJK donne une base de cet espace de fonctions, construite à partir des bases de fonctions définies sur les différents ensembles d'indices.

Si les ensembles

$$\{\phi_{\alpha}^I \mid \alpha = 1, \text{card } I\} \quad \{\phi_{\beta}^J \mid \beta = 1, \text{card } J\} \quad \{\phi_{\gamma}^K \mid \gamma = 1, \text{card } K\}$$

constituent les bases orthonormées de  $R^I$ ,  $R^J$  et  $R^K$  alors les triplets  $\{\phi_{\alpha}^I \phi_{\beta}^J \phi_{\gamma}^K \mid \alpha \in (1, \text{card } I), \beta \in (1, \text{card } J), \gamma \in (1, \text{card } K)\}$  donnent une base orthonormée de  $R^{IJK}$  pour la métrique produit  $m_I \otimes m_J \otimes m_K$  déduite des métriques sur  $R^I$ ,  $R^J$  et  $R^K$ .

La forme de la métrique qui a été choisie dans  $R^{IJK}$  fait que la vérification de cette propriété est immédiate. On démontre en effet que des triplets sont orthogonaux si l'un de leurs indices inférieurs diffère.

$$\begin{aligned} \text{Produit scalaire } \langle \phi_{\alpha}^I \phi_{\beta}^J \phi_{\gamma}^K, \phi_{\alpha'}^I \phi_{\beta'}^J \phi_{\gamma'}^K \rangle &= \sum_{i,j,k} m_{i,j,k} \phi_{\alpha}^i \phi_{\beta}^j \phi_{\gamma}^k \phi_{\alpha'}^i \phi_{\beta'}^j \phi_{\gamma'}^k \\ &= \langle \phi_{\alpha}^I \phi_{\alpha'}^I \rangle \langle \phi_{\beta}^J \phi_{\beta'}^J \rangle \langle \phi_{\gamma}^K \phi_{\gamma'}^K \rangle = \delta_{\alpha}^{\alpha'} \delta_{\beta}^{\beta'} \delta_{\gamma}^{\gamma'} \end{aligned}$$

où  $\delta_{\alpha}^{\alpha'}$  désigne le symbole de kronecker.

Etudions à présent le cas particulier de la base utilisée en analyse de variance.

Pour chacun des ensembles le premier élément est une fonction constante égale à 1 que l'on notera  $1^I$  pour l'ensemble  $I$  par exemple. Cette convention donne une propriété particulière pour les fonctions de base suivante :  $\langle 1^I, \phi^I \rangle = 0 = \sum_i 1 \cdot x_i \cdot \phi_{m_i}^I = \sum_i \phi_{m_i}^I =$  valeur moyenne de  $\phi^I$  relativement à la pondération  $m_I$ .

Nous dirons que ces fonctions sont de moyenne nulle :

Les bases ainsi définies :

$$\{1^I, \phi_\alpha^I \mid \alpha \in (1, \text{card } I-1)\}$$

$$\{1^J, \phi_\beta^J \mid \beta \in (1, \text{card } J-1)\}$$

$$\{1^K, \phi_\gamma^K \mid \gamma \in (1, \text{card } K-1)\}$$

font apparaître les triplets suivants :

$$1^I 1^J 1^K, \text{ base des fonctions constantes sur } \mathbb{R}^{IJK}$$

$\{1^I 1^J \phi_\gamma^K \mid \gamma \in (1, \text{card } K-1)\}$  et les permutations en IJK : base des fonctions ne dépendant que d'un indice et de moyenne nulle relativement à cet indice.

$\{1^I \phi_\beta^J \phi_\gamma^K \mid \beta \in (1, \text{card } J-1), \gamma \in (1, \text{card } K-1)\}$  et les permutations : base des fonctions indépendantes d'un indice, et de moyenne nulle relativement à chacun des deux autres

$\{\phi_\alpha^I \phi_\beta^J \phi_\gamma^K \mid \alpha \in (1, \text{card } I-1), \beta \in (1, \text{card } J-1), \gamma \in (1, \text{card } K-1)\}$  : base de fonctions de moyenne nulle relativement aux trois indices.

Nous noterons dorénavant  $\mathbb{R}^{IJK}$  ou  $\mathbb{R}^I \otimes \mathbb{R}^J \otimes \mathbb{R}^K$  pour l'espace des fonctions  $f_{IJK}$  et  $\mathbb{R}^I \times \mathbb{R}^J \times \mathbb{R}^K$  pour les fonctions produits  $f_I f_J f_K$ .

La base utilisée en analyse de variance fait apparaître une certaine décomposition de  $\mathbb{R}^I \otimes \mathbb{R}^J \otimes \mathbb{R}^K$ . Posons  $\mathbb{R}^I = C^I \oplus V^I$ , l'espace des fonctions sur I est la somme directe de deux sous-espaces orthogonaux, celui des fonctions constantes (dont le vecteur de base est  $1^I$ ), et celui des fonctions de moyenne nulle pour les métriques  $m_I$  (engendré par les  $\phi_\alpha^I$ ).

$R^I \otimes R^J \otimes R^K$  associé à la métrique  $m_I \otimes m_J \otimes m_K$  se décompose de la façon suivante :

$$\begin{aligned} (C^I \otimes V^I) \otimes (C^J \otimes V^J) \otimes (C^K \otimes V^K) &= (C^I \otimes C^J \otimes C^K) \oplus (C^I \otimes C^J \otimes V^K) \\ &\oplus (C^I \otimes V^J \otimes C^K) \oplus (V^I \otimes C^J \otimes C^K) \oplus (C^I \otimes V^J \otimes V^K) \oplus (V^I \otimes C^J \otimes V^K) \\ &\oplus (V^I \otimes V^J \otimes C^K) \oplus (V^I \otimes V^J \otimes V^K) \end{aligned}$$

Somme directe de  $2^n$  sous-espaces orthogonaux pour n ensembles d'indices dont les bases se déduisent simplement des bases de  $R^I$ ,  $R^J$  et  $R^K$  et dont les dimensions respectives sont les suivantes :

1 ; (card K-1) ; (card J-1) ; (card I-1); (card K-1) x (card J-1) ; (card I-1) x (card K-1) ; (card J-1) x (card I-1); (card K-1)x (card I-1)x (card J-1).

La démonstration de ce théorème se généralise rapidement puisque ne s'agit que de faire des combinaisons de vecteurs de base . Et l'on arrive au résultat général suivant :

Toute fonction définie sur une famille d'indice  $F = \{I, J, K, L, M\}$  peut se décomposer en une somme de fonctions définies sur les parties de  $F$   $P(F)$ , les composantes seront donc au nombre de  $2^{\text{card}(F)}$  et la métrique utilisée peut être la métrique produit déduite des métriques introduites dans les espaces des fonctions à un indice.

Cette décomposition relative aux associations d'indices permet de faire apparaître des effets liés à un indice ou une modalité d'un caractère et des interactions liées aux associations de certaines modalités de deux ou plusieurs caractères.

- Décomposition et étude dimensionnelle.

La fonction écart qui sera décomposée est celle qui apparaît dans le  $\chi^2$  de contingence à deux dimensions lorsque les marges du tableau sont estimées

$$\chi^2 = n \sum_{ij} \left( \frac{\frac{n_{ij}}{n_i n_j}}{\frac{n_{ij}}{n_i n_j}} - 1 \right)^2 \times \frac{n_i n_j}{n^2} = n \text{Var} \left( \frac{\frac{n_{ij}}{n_i n_j}}{\frac{n_i n_j}{n_i n_j}} \right)$$

La métrique  $m_i$  &  $m_j$  est définie par la densité  $\frac{n_i n_j}{n^2}$  déduite de densité  $m_i = \frac{n_i}{n}$ ,  $m_j = \frac{n_j}{n}$ .

La décomposition selon la base de l'analyse de la variance donne

$$\left\{ \frac{n_{ij}}{n_i n_j} \right\} = \frac{f_{IJ}}{f_I \times f_J} = \underbrace{g_{IJ} = \langle g_{IJ}, 1^I 1^J \rangle}_{\text{moyenne générale}} \cdot 1^I 1^J + \underbrace{\langle g_{Ii} - m_i, 1^J \rangle}_{\text{effet en I}} \cdot 1^J + \underbrace{\langle g_{Jj} - m_j, 1^I \rangle}_{\text{effet en J}} \cdot 1^I + \underbrace{\phi_{IJ}}_{\text{résidu}}$$

Le résidu  $\phi_{IJ}$  est ici obtenu par différence.

Pour cette fonction écart la moyenne générale est égale à 1 et les effets sont nuls.

Le résidu est de moyenne nulle en i et en j, il appartient à un sous espace de dimension (card I-1) (Card J-1).

On sait qu'il peut se décomposer en produit de fonctions (ou facteurs) sur I et J.

$$\phi^{IJ} = \sum_{\alpha, \beta} \lambda_{\alpha\beta} \phi_{\alpha}^I \phi_{\beta}^J$$

$\{ \phi_{\alpha}^I \mid \alpha \in A \}$  est une base de  $V^I$   $A = \{1, \text{card } I-1\}$

$\{ \phi_{\beta}^J \mid \beta \in B \}$  est une base de  $V^J$   $B = \{1, \text{card } J-1\}$

Etudions le sous-ensemble engendré par le produit de deux fonctions de ce type soit :

$f_I^1 \times f_J^1$  un élément de l'ensemble  $R^I \times R^J$  inclus dans  $R^I \otimes R^J$

$f_I^1 = \sum_i \lambda_i \delta_i$ ,  $f_J^1 = \sum_j \lambda_j \delta_j$  en réutilisant la base initiale  $\delta_{\alpha} = \begin{cases} 1 & \text{sur } \alpha \\ 0 & \text{ailleurs} \end{cases}$

$$f_I^1 \times f_J^1 = \sum_{i,j} \lambda_i \lambda_j \delta_i \delta_j = \sum_{i,j} \lambda_i \lambda_j \delta_{ij} = \sum_{i,j} u_{ij} \delta_{ij} \in R^I \otimes R^J$$

Les coordonnées de cette fonction produit ont une forme très particulière  $u_{ij} = \lambda_i \lambda_j$ .

On s'aperçoit que (card I + card J) valeurs numériques arbitraires suffisent à définir toutes les coordonnées  $u_{ij}$ .

Les relations qui lient les coordonnées  $u_{ij}$  ne sont pas linéaires et le couple  $f_I^1 \times f_J^1$  n'engendre pas un sous espace de  $R^I \otimes R^J$ . Nous pouvons donc seulement nous intéresser au nombre de valeurs arbitraires (ou paramètres) nécessaire pour définir une décomposition d'un élément de  $R^I \otimes R^J$  en somme d'éléments de  $R^I \times R^J$ . On admettra qu'il y a recouvrement des deux ensembles lorsqu'ils font intervenir un même nombre de paramètres.

Recherchons une décomposition sans répétition de facteur :

$$f^{IJ} = \sum_{\alpha \in A \text{ ou } B} \lambda_{\alpha} f_{\alpha}^I f_{\alpha}^J$$

qui aurait pour intérêt de ne jamais associer une même fonction de I à plusieurs fonction de J et ainsi de donner une orthogonalité entre les associations son seul ment dans  $R^I \otimes R^J$  mais aussi au niveau des facteurs qui les composent.

Considérons donc un second couple  $f_I^2 \times f_J^2$  tel que  $\langle f_I^2 ; f_I^1 \rangle = \langle f_J^2 ; f_J^1 \rangle$  nous dirons que les couples sont complètement orthogonaux, les deux conditions constituent des relations linéaires entre les composantes de  $f_I^2$  et de  $f_J^2$ , card I + card J valeurs numériques suffisent à définir le second couple. Pour le même couple complètement orthogonal aux précédents il en faudra card I + card J - 2(m-1).

Avec une décomposition sans répétition et en supposant dorénavant card I  $\geq$  card J on peut utiliser seulement card J couples, et donc introduire :

$$m = \text{card J} \\ \sum_{m=1}^{\text{Card I} + \text{Card J} - 2(m-1)} = \text{card J} (\text{card I} + \text{card J} - (\text{card$$

$$1)) = \text{card J} (\text{card I} + 1).$$

C'est-à-dire un nombre toujours supérieur à celui des valeurs numériques nécessaire à définir un élément  $f^{IJ}$  de  $R^I \otimes R^J$ . Le raisonnement s'étend aisément à l'espace des facteurs de moyennes nulles. La décomposition sans répétition du résidu  $\phi^{IJ}$  semble donc possible.

Pour l'explicitier on part de l'hypothèse :

$$\frac{f_{ij}}{f_i f_j} = 1 + \sum_{\alpha \in A} \lambda_{\alpha} \phi_{\alpha}^i \phi_{\alpha}^j \quad \text{avec} \quad \begin{cases} \sum_i \phi_{\alpha} f_i = 0 & \phi_{\epsilon}^I v^I \\ \sum_j \phi_{\alpha} f_j = 0 & \forall \alpha \quad \phi_{\epsilon}^J v^J \end{cases}$$

et les conditions de normalisation suivantes :

$$\sum_i \phi_{\alpha}^i \phi_{\alpha}^i f_i = \delta_{\alpha}^{\alpha'} \quad \sum_i \phi_{\alpha} \phi_{\alpha} f_i = \delta_{\alpha}^{\alpha'}$$

pour obtenir une solution unique



En utilisant les propriétés d'orthogonalité des facteurs  $\phi$  on a la simplification suivante :

$$\sum_i \frac{f_{ij}}{f_i f_j} \phi_{\alpha'}^i f_i = \sum_i \phi_{\alpha'}^i f_i + \sum_{\substack{\alpha \in A \\ i \in I}} \lambda_{\alpha} \phi_{\alpha}^i \phi_{\alpha'}^i \phi_{\alpha'}^i f_i = \lambda_{\alpha'} \phi_{\alpha'}^i$$

d'où

$$(1) \quad \sum_i \frac{f_{ii}}{f_i} \phi_{\alpha}^i = \lambda_{\alpha} \phi_{\alpha}^i$$

formule dite de transition qui permet de passer du facteur sur I au facteur sur J par l'intermédiaire d'une répartition de fréquence  $f^j = \left\{ \frac{f_{ij}}{f_j}, i \in I \right\}$

En tenant compte que l'on a aussi :

(2)  $\sum_j \frac{f_{ij}}{f_i} \phi_{\alpha}^j = \lambda_{\alpha} \phi_{\alpha}^i$  on obtient une procédure simple pour calculer les facteurs

$$\phi_{\alpha}^I \text{ et } \phi_{\alpha}^J; (1) \text{ et } (2) \Rightarrow \sum_j \frac{f_{ij}}{f_i} \sum_j \frac{f_{i'j}}{f_j} \phi_{\alpha}^{i'} = \lambda_{\alpha}^2 \phi_{\alpha}^i$$

$\phi_{\alpha}^I$  apparaît comme le facteur propre de l'opérateur

$$\sigma_I^I = \left\{ \sum_j \frac{f_{ij} f_{i'j}}{f_i f_j} \mid i \text{ et } i' \in I \right\}$$

associé à la valeur propre  $\lambda_{\alpha}^2$

Les couples  $\{\phi_{\alpha}^I \phi_{\alpha}^J \mid \alpha \in A\}$  constituent des sous-ensembles orthogonaux  $R^I \times R^J$  dans  $R^I \otimes R^J$  le module de la projection de la fonction écart sur un de ces sous-ensembles est donné par l'expression suivante :

$$\left| \left\langle \frac{f_{11}}{f_1 f_1} - 1, \phi_{\alpha}^I \phi_{\alpha}^J \right\rangle \right| = \left| \sum_{\substack{i \in I \\ j \in J}} \left( \sum_{\alpha' \in A} \lambda_{\alpha'} \phi_{\alpha'}^i \phi_{\alpha'}^j \right) \phi_{\alpha}^i \phi_{\alpha}^j f_i f_j \right|$$

$$= |\lambda_{\alpha}|$$

Les facteurs déduits de l'opération  $\sigma_I^I$  par ordre décroissant de ses valeurs propres fournissent donc les couples  $\phi^I \phi^J$  qui expliquent la fonction d'écart au sens des moindres carrés dans  $R^I \otimes R^J$  muni de la métrique  $f_I \otimes f_J$ .

Remarquons que lorsqu'un premier couple est connu le calcul porte sur l'écart résiduel  $(\frac{f_{ij}}{f_i f_j} - 1 - \lambda_1 \phi_1^i \phi_1^j)$  qui amène à considérer :

$\sigma_i^{i'} = \sigma_{i'}^{i'} - \lambda_1 \phi_1^{i'} \phi_1^i$  l'opérateur  $\sigma_i^I = \delta_i^I - \lambda_1 \phi_1^i \phi_1^I$  qui n'est autre que l'opérateur  $\sigma_I^I$  dépourvu de sa plus forte valeur propre.

Le problème de la recherche des facteurs  $\phi^I$  et  $\phi^J$  aurait pu être posé directement en terme de moindre carré par exemple de la façon suivante :

$$\left\{ \begin{array}{l} \text{maximiser} \quad \sum_{i,j} \left( \frac{f_{ij}}{f_i f_j} - 1 \right) \phi_1^i \phi_1^j f_i f_j = \langle g_{ij}, \phi_1^i \phi_1^j \rangle = A \\ \text{avec} \quad B = \sum_i \phi_1^{i2} f_i = 1 \quad C = \sum_j \phi_1^{j2} f_j = 1 \end{array} \right.$$

$$\text{Soit } L = A - \lambda_1 (B - 1) - \lambda_2 (C - 1) \quad \text{maximum}$$

$$\frac{\partial L}{\partial \phi_1^i} = \phi_1^i \sum_j \left( \frac{f_{ij}}{f_i f_j} - 1 \right) \phi_1^j f_i f_j - 2 \lambda_1 \phi_1^i f_i = 0$$

$$\Rightarrow \sum_j \frac{f_{ij}}{f_j} \phi_1^j = 2 \lambda_1 \phi_1^i \quad \text{en tenant compte de} \quad \sum_i \phi_1^i f_i = 0$$

On retrouve donc la formule de transition qui conduit à rechercher les directions propres d'un opérateur (ici les valeurs propres sont égales à  $4 \lambda_1 \lambda_2$ ).

Dans les problèmes d'analyse de tableaux de fréquence il n'est pas nécessaire de fixer arbitrairement certaines propriétés de la base de décomposition, ainsi les conditions de moyenne nulle sur certains facteurs découlent des conditions de normalisation, car le premier couple de facteurs est nécessairement constitué par les fonctions constantes  $1_{I,J}$ .

$$\text{En effet } \langle g_{IJ}, \phi_\alpha^I, \phi_\alpha^J \rangle = \sum_{ij} \underbrace{\phi_\alpha^i \phi_\alpha^j}_{\text{facteur de norme 1 sur } f_j \text{ et } f_I} \times \underbrace{f_{ij}}_{\text{densité de fréquence}} = \lambda_\alpha$$

$\phi_\alpha^I$  et  $\phi_\alpha^J$  sont aussi de norme 1 sur  $f_{ij}$

L'expression ci-dessus apparaît donc comme une corrélation entre deux fonctions sur  $I \times J$  par rapport à la densité  $f_{IJ}$ , on a donc nécessairement  $\lambda_\alpha^2 \leq 1$ .

Or le couple  $1_{I,J}$  donne une corrélation égale à 1, il constitue bien la variété la plus proche de la fonction qui nous intéresse dans  $R^I \otimes R^J$  avant toute décomposition  $\{f^{ij}/f_i f_j \mid i \in I, j \in J\}$ .

Rappelons quelques propriétés de la décomposition utilisée en analyse des correspondances :

$$\begin{aligned} \chi^2 &= n \sum_{ij} \left( \frac{f_{ij}}{f_i f_j} - 1 \right)^2 f_i f_j = n \sum_{ij} \left( \sum_\alpha \lambda_\alpha \phi_\alpha^i \phi_\alpha^j \right) \left( \sum_{\alpha'} \lambda_{\alpha'} \phi_{\alpha'}^i \phi_{\alpha'}^j \right) f_i f_j \\ &= n \sum_{\alpha \alpha'} \sum_{ij} \lambda_\alpha \lambda_{\alpha'} (\phi_\alpha^i \phi_{\alpha'}^j f_i) (\phi_{\alpha'}^i \phi_\alpha^j f_j) \\ &= n \sum_{\alpha \alpha'} \lambda_\alpha \lambda_{\alpha'} \delta_{\alpha'}^\alpha \delta_{\alpha'}^\alpha \end{aligned}$$

$$\chi^2 = n \sum_{\substack{\alpha \\ \in A}} \lambda_{\alpha}^2 \text{ qui ne peut dépasser } n(\min(\text{card } I, \text{card } J) - 1).$$

La détermination des valeurs propres, peut se faire par réduction de  $\sigma_I^I = \left\{ \sum_j \frac{f_{ij} f_{ij'}}{f_i f_j} \mid i \text{ et } i' \in I \right\}$  ou de  $\sigma_J^J = \left\{ \sum_i \frac{f_{ij} f_{ij'}}{f_i f_j} \mid j, j' \in J \right\}$ , dont on vérifie que les directions propres relatives à une même valeur propre se déduisent de la formule de transition, et que deux facteurs relatifs à deux valeurs propres différentes sont orthogonaux dans la métrique  $f_I$  ou  $f_J$  selon qu'on étudie  $\sigma_I^I$  ou  $\sigma_J^J$ .

Les résultats de cette décomposition factorielle prennent toute leur ampleur lorsque l'on considère l'opérateur  $\sigma_J^J$  comme une matrice d'inertie d'un nuage de points  $\left\{ \frac{f_{ij}}{f_i}, j \in J \right\}$  dans un espace  $R^J$  muni de la métrique  $\frac{1}{f_j}$  dite métrique du  $\chi^2$ . Les facteurs calculés donnent alors des classements sur les  $\text{card } I$  points que l'on a décrits dans  $R^J$  grâce à des profils de fréquence. De plus ces facteurs sont non corrélés pour la densité  $f_I$  des poids des points  $i$ , ils estiment donc des causes de dispersion indépendantes. L'absence de répétition d'un même facteur simplifie considérablement l'interprétation des associations entre facteurs sur  $I$  et  $J$ , la formule de transition qui lie de façon biunivoque les facteurs d'un même couple permet de présenter simultanément les résultats obtenus par les deux approches possibles à l'aide d'un nuage de points ( $\text{card } I$  points dans  $R^J$  ou  $\text{card } J$  points des  $R^I$ ).

- Décomposition sur la base de l'analyse de la variance. Nous appuons le théorème de décomposition au rapport des deux densités de fréquence suivantes :  $f_{ijk}$  fréquence observée

$f_i f_j f_k = \frac{n_i}{n} \times \frac{n_j}{n} \times \frac{n_k}{n}$  fréquence estimant la probabilité de l'association (i,j,k) dans l'hypothèse d'indépendance. Soit  $g_{ijk} = f_{ijk}/f_i f_j f_k$

La moyenne générale est :  $\langle g_{ijk}, 1^I 1^J 1^K \rangle =$  projection sur l'ensemble des fonctions constantes.

$\sum_{ijk} \frac{f_{ijk}}{f_i f_j f_k} \cdot \frac{f_i f_j f_k}{f_i f_j f_k} = 1$  ; en utilisant la densité correspondant à notre hypothèse

Effet sur l'ensemble I =  $g_I$ , projection orthogonale de  $g_{IJK}$  sur l'ensemble des fonctions de I. On aura donc  $(g_{IJK} - g_I) \perp R^I$  donc  $(g_{IJK} - g_I) \perp 1^J 1^K \phi^I \forall \phi^I$  une condition nécessaire et suffisante pour obtenir ce résultat c'est d'avoir  $\langle g_{IJK}, 1^J 1^K \rangle_{f_I f_J f_K} = g_I$  en effet

$$\langle g_{IJK} - g_I, 1^J 1^K \phi^I \rangle_{f_I f_J f_K} = \langle \langle g_{IJK}, 1^J 1^K \rangle_{f_J f_K}, \phi^I \rangle_{f_I} - \langle g_I, \phi^I \rangle_{f_I} = 0$$

L'effet sur un ensemble est donc bien la moyenne de la fonction relativement aux autres ensembles.

$$\text{Effet en I} = \sum_{kj} \left( \frac{f_{ijk}}{f_i f_j f_k} - 1 \right) f_j f_k = \sum_{jk} \frac{f_{ijk}}{f_i} - 1 = 0; \text{ même résultat}$$

nul sur J et K ceci résulte de l'utilisation de la densité de référence comme métrique.

L'effet sur I x J est la fonction indépendante de K la plus proche de  $(g_{IJK} - 1)$ , on réutilise la célèbre propriété de la moyenne quant à la minimisation des écarts quadratiques d'où :

$$\sum_k \left( \frac{f_{ijk}}{f_i f_j f_k} - 1 \right) f_k = \frac{f_{ij.}}{f_i f_j} - 1$$

L'interaction entre les trois ensembles I, J, K est obtenu par différence puisque l'on sait qu'elle correspond au sous espace orthogonal qui complète l'espace des fonctions de trois indices.

$$\begin{aligned} \text{Intéraction} &= \frac{f_{ijk}}{f_i f_j f_k} - 1 - \left( \frac{f_{ij.}}{f_i f_j} - 1 \right) - \left( \frac{f_{.jk}}{f_j f_k} - 1 \right) - \left( \frac{f_{i.k}}{f_i f_k} - 1 \right) \\ &= \frac{f_{ijk} - f_{ij.} f_k - f_{.jk} f_i - f_{i.k} f_j + 2 f_i f_j f_k}{f_i f_j f_k} \end{aligned}$$

A partir de cette décomposition en sous-espaces orthogonaux de  $R^I \otimes R^J \otimes R^K$  on peut être tenté de rechercher des bases de ces sous-espaces sous forme de triplets de fonctions à un indice pour arriver à obtenir une explication de l'écart à l'indépendance aussi simple que dans le cas de l'analyse de correspondance.

Le problème posé est donc de passer à une écriture semblable.

$$\frac{f_{ijk}}{f_i f_j f_k} = 1 + \sum_{\alpha \in A} \lambda_{\alpha} \phi_{\alpha}^i \phi_{\alpha}^j \phi_{\alpha}^k$$

De par l'existence même des effets sur les couples de deux caractères on est sûr que les facteurs  $\phi$  ne pourront pas tous être de moyenne nulle, cette décomposition sera possible à condition de répéter les facteurs constants  $1^I, 1^J$  ou  $1^K$  qui apparaissent déjà dans la moyenne générale.

La solution des projections orthogonales. L'interprétation qui a été faite de l'analyse des correspondances incite à rechercher une décomposition factorielle par une méthode du type des moindres carrés, c'est-à-dire de projections orthogonales. Ainsi la valeur explicative des facteurs qui apparaîtront les premiers sera toujours garantie. Nous avons vu que les facteurs constants sur un caractère devaient obligatoirement se répéter, on peut prévoir qu'ils ne seront pas les seuls à présenter cet inconvénient.

En effet un triplet  $\phi_m^I \phi_m^J \phi_m^K$  complètement orthogonal aux  $(m-1)$  précédents est défini par  $(\text{card } I + \text{card } J + \text{card } K - 3(m-1))$  en admettant qu'ils peuvent être égaux aux facteurs constants.

Nous disposerons en fin de calcul de  $\text{card } A$  triplets complètement orthogonaux, il faudra que nous ayons la relation suivante :

$\text{card } A (\text{card } I + \text{card } J + \text{card } K - \frac{3}{2}(\text{card } A - 1)) \geq (\text{card } I \times \text{card } J \times \text{card } K)$ . En admettant que  $\text{card } K \leq \text{card } J \leq \text{card } I$ , l'absence de répétition signifie  $\text{card } A \leq \text{card } K$ , alors il est évident que la relation entre les dimensions de  $R^I \otimes R^J \otimes R^K$  et l'espace engendré par les  $\text{card } A$  éléments de  $R^I \times R^J \times R^K$  ne peut pas être satisfaite.

On pourrait imaginer une approche du problème où l'on s'imposerait de minimiser le nombre de facteurs utilisés. Alors on devrait obtenir un effectif  $\phi_i$ ;  $\text{card} A^I \leq \min(\text{card } I, \text{card } J \times \text{card } K)$ , nombre de facteurs permettant une décomposition de  $R^{IJK}$  dans  $R^I \otimes R^{JK}$ , (cf. l'étude de L.R. TUCKER).

Considérons le cas le plus général de l'approximation de  $g_{IJK} =$

$$\left\{ \frac{f_{ijk}}{f_i f_j f_k} \mid i \in I, j \in J, k \in K \right\} \in R^I \otimes R^J \otimes R^K$$

par des éléments de  $R^I \times R^J \times R^K$

$G_{IJK}$  peut être considéré comme un vecteur dont on connaît les coordonnées dans la base  $\delta^i \otimes \delta^j \otimes \delta^k$  avec  $\delta^v = \begin{cases} 1 & \text{pour l'indice } v \\ 0 & \text{ailleurs} \end{cases}$ , un élément de

$R^I$  peut s'écrire  $e_p = \{\phi_p^i \delta^i \mid i \in I\}$  considérons  $\text{card } I$  vecteurs de  $R^I$  indicés par  $p$ ,  $\text{card } J$  vecteurs de  $R^J$  indices par  $m$  et  $\text{card } K$  vecteurs de  $R^K$  indicés par  $n$ . Effectuons trois changements de base simultanés, les vecteurs  $e_p, e_m, e_n$  étant les nouveaux vecteurs de base de  $R^I, R^J$  et  $R^K$  on peut donc construire une nouvelle base de  $R^I \otimes R^J \otimes R^K$ , où  $G_{IJK}$  aura pour coordonnées les valeurs

$$\left\{ \lambda_{pmn} \mid p \in L, m \in M, n \in N \right\}$$

soit

$$g_{IJK} = \sum_{i,j,k} \frac{f_{ijk}}{f_i f_j f_k} \delta^i \otimes \delta^j \otimes \delta^k = \sum_{p,m,n} \lambda_{pmn} e_p \otimes e_m \otimes e_n = g_{LMN}$$

Pour effectuer un véritable changement de base nous imposons aux vecteurs de la nouvelle base d'être normés.

$$\text{d'où } \sum_i \langle \phi_p^i, \delta^i \rangle_{f_i}^2 = \sum_i (\phi_p^i)^2 f_i = 1 \quad \text{de même}$$

en  $(j, m)$  et  $(k, n)$ .



On recherche donc le vecteur  $\lambda_{p m n} e_p \otimes e_m \otimes e_n$  le plus proche de  $G_{LMN}$ , au sens de la métrique définie dans  $R^I \otimes R^J \otimes R^K$  par  $f_I \otimes f_J \otimes f_K$ . La solution correspond au minimum de l'expression suivante :

$$\sum_{i,j,k} \left( \frac{f_{ijk}}{f_i f_j f_k} - \lambda_{p m n} \phi_p^i \phi_m^j \phi_n^k \right)^2 f_i f_j f_k$$

avec les contraintes de norme sur les  $\phi$ .

En effet si on ramène tous les vecteurs à la base originale on a :

$$\sum_{i,j,k} \frac{f_{ijk}}{f_i f_j f_k} \delta^i \otimes \delta^j \otimes \delta^k = \sum_{\substack{i,j,k \\ p,m,n}} \lambda_{p m n} \phi_p^i \phi_m^j \phi_n^k \delta^i \otimes \delta^j \otimes \delta^k$$

La différence entre  $g_{LJK}$  et sa première coordonnée sur la nouvelle base est bien :

$$\left( \frac{f_{ijk}}{f_i f_j f_k} - \lambda_{111} \phi_1^i \phi_1^j \phi_1^k \right) \delta^i \otimes \delta^j \otimes \delta^k$$

c'est la norme de cette différence que nous cherchons à minimiser.

Cette approche permet de voir que le problème d'approximation sous contrainte peut se ramener à une recherche d'une plus forte coordonnée dans une base particulière.

La formulation précédente est équivalente à celle qui n'utilise qu'un produit scalaire.

$$\max \sum_{i,j,k} \left( \frac{f_{ijk}}{f_i f_j f_k} \phi_p^i \phi_m^j \phi_n^k \right) f_i f_j f_k = \lambda_{p m n} \quad \text{le signe n'ayant pas}$$

d'importance puisque l'on peut modifier les signes des facteurs  $\phi$ .

*non  
de l'ordre de  
outre  
I a qm  
à pas de  
suffit  
le fait  
cette hypo  
hère est fait  
mais seul  
-21- page 22*

Cette expression  $\sum_{ijk} f_{ijk} \phi_p^i \phi_m^j \phi_n^k = \lambda_{emn}$  peut à nouveau s'interpréter

comme une corrélation entre deux fonctions sur  $I \times J \times K$ ,  $\phi_e^i$  et  $\phi_m^j \phi_n^k$  par exemple, la densité de probabilité étant définie par  $f_{ijk}$ , la valeur maximum de  $\lambda_{emn}$  est donc égale à 1, valeur qui est atteinte avec les facteurs  $1^I 1^J 1^K$ , ce triplet particulier apparaît donc, tant naturellement, en première position dans la décomposition selon le critère des moindres carrés.

Pour rechercher les fonctions de bases suivantes nous sommes amenés à poser des conditions d'orthogonalité. En effet nous désirons que les fonctions explicatives relativement à un indice soient deux à deux non-correlées. Pour le deuxième triplet explicatif cette non corrélation se traduit par le fait que les facteurs sont de moyenne nulle à moins que ce second triplet ne donne lieu à une répétition d'un des facteurs constants  $1^I 1^J 1^K$ . Ce premier triplet permet déjà d'écrire la décomposition complète de la fonction écart sur des bases orthonormées comportant les facteurs constants.

$$\frac{f_{ijk}}{f_i f_j f_k} = 1^I 1^J 1^K + \sum_{m,n} \lambda_{o,mn} 1^I \phi_m^J \phi_n^K + \sum_{p,n} \lambda_{p,0n} \phi_p^I 1^J \phi_n^K + \sum_{p,m} \lambda_{p,m,0} \phi_p^I \phi_m^J 1^K + \sum_{p,m,n} \lambda_{p,m,n} \phi_p^I \phi_m^J \phi_n^K$$

Les bases sont définies comme ci-dessous :

$$\left\{ 1^I, \phi_p^I \mid p \in 1, \text{card } I - 1 \right\} \text{ avec } \sum_i \phi^i f_i = 0, \sum_i \phi_p^i \phi_{p'}^i f_i = \delta_{pp'}$$

$$\left\{ 1^J, \phi_m^J \mid m \in 1, \text{card } J - 1 \right\} \text{ avec } \sum_j \phi^j f_j = 0, \sum_j \phi_m^j \phi_{m'}^j f_j = \delta_{mm'}$$

$$\left\{ 1^K, \phi_n^K \mid n \in 1, \text{card } K - 1 \right\} \text{ avec } \sum_k \phi^k f_k = 0, \sum_k \phi_n^k \phi_{n'}^k f_k = \delta_{nn'}$$

Elles sont analogues à celles qui ont été utilisées dans le théorème de décomposition.

## INTERPRETATION DE LA DECOMPOSITION

La formule explicative fait apparaître cinq termes de nature différente selon qu'ils contiennent ou non les facteurs constants. Dans le cas général pour une fonction sur trois ensembles d'indices nous devrions obtenir  $2^3$  termes différents, qui expriment les coordonnées de la fonction dans les sous-espaces recotés dans le théorème de décomposition.

Ici leur nombre est réduit du simple fait que les effets relatifs un indice (moyenne relative aux deux autres) sont nuls.

Pour interpréter les termes restant dans cette décomposition et donner les composantes dans les sous-espaces introduits en analyse de la variance il suffit d'effectuer des sommations relativement aux trois indices.

Sommation en  $i \Rightarrow$  disparition des termes contenant  $\phi_p^i$

$$\sum_{m,n} \lambda_{omn} \phi_m^j \phi_n^k = \sum_i \left( \frac{f_{ijk}}{f_i f_j f_k} - 1 \right) f_i = \left( \frac{f_{jk}}{f_j f_k} - 1 \right) = \text{Effet (J, K)}$$

Même résultat en  $j$  et  $k$

$$\sum_{p,n} \lambda_{pon} \phi_p^i \phi_n^k = \left( \frac{f_{ik}}{f_i f_k} - 1 \right) = \text{Effet (I, K)}$$

$$\sum_{p,m} \lambda_{pmo} \phi_p^i \phi_m^j = \left( \frac{f_{ij}}{f_i f_j} - 1 \right) = \text{Effet (I, J)}$$

Et par soustraction

$$\sum_{p,m,n} \lambda_{pmn} \phi_p^i \phi_m^j \phi_n^k = \frac{f_{ijk}}{f_i f_j f_k} - 1 - \text{Effet (I, J)} - \text{Effet (I, K)} - \text{Effet (J, K)}$$

= Interaction

Dans cette décomposition s'appuyant sur la recherche des triplets plus explicatifs on voit que dans le terme correspondant à une fonction d'écart sur deux indices certains facteurs peuvent à priori se répéter en s'associant à deux facteurs différents sur le second ensemble.

Ces répétitions justifient l'emploi d'une double sommation pour chaque terme.

On peut se demander quelles doivent être les conditions remplies par le tableau  $f_{ijk}$  pour que les facteurs sur un même ensemble soient tous orthogonaux d'un terme à l'autre.

Supposons qu'il n'y ait aucun facteur commun entre les fonctions liées aux contingences IK et JK alors nous aurions :

$$\sum_k \left( \sum_{l \in n} \lambda_{l \in n} \phi_l^I \phi_n^k \right) \left( \sum_{m' \in n'} \lambda_{m' \in n'} \phi_{m'}^J \phi_{n'}^k \right) f_k = \sum_{n' \in n} \delta_n^{n'} = 0$$

$$\sum_k \left( \frac{f_{ik}}{f_i f_n} - 1 \right) \left( \frac{f_{jk}}{f_j f_n} - 1 \right) f_k = \sum_k \frac{f_{ik} f_{jk}}{f_i f_j f_n} - \sum_k \frac{f_{ik}}{f_i} - \sum_k \frac{f_{jk}}{f_j} + \sum_k f_k$$

d'où

$$= \sum_k \left( \frac{f_{ik} f_{jk}}{f_n} \times \frac{1}{f_i f_j} \right) - 1$$

$$f_i \cdot f_j = \sum_k \frac{f_{ik} \cdot f_{jk}}{f_n} = \hat{f}_{ij}$$

$\hat{f}_{ij}$  peut s'interpréter comme la densité de fréquence sur I x J qui résulte d'une indépendance conditionnelle en i et j si k. Les densités de probabilité étant estimées par les densités de fréquence pour les tableaux IK et JK.

Ainsi l'absence de répétition signifie que les hypothèses d'indépendances conditionnelles n'apportent pas plus pour la description du tableau que les hypothèses d'indépendances totales, c'est-à-dire qu'un caractère n'aide pas à l'explication de l'association entre les deux autres.

La densité de fréquence qui est apparue dans la condition de non répétition sera souvent retrouvée dans les études postérieures et son analyse sera approfondie (chapitre IV et V).

### CALCULS PRATIQUES

- Principe de maximisation

Soit :

$$g_{ijk} = \frac{f_{ijk}}{f_i f_j f_k} - 1$$

Nous recherchons la première coordonnée  $\lambda_{111} \phi_1^i \phi_1^j \phi_1^k$  la plus proche de la fonction  $g_{ijk}$ . Cette proximité peut s'exprimer de la façon suivante :

$$g_{ijk} - \lambda_{111} \phi_1^i \phi_1^j \phi_1^k = \varepsilon_{ijk} = \text{distance de } g \text{ ou triplet.}$$

On utilisera une minimisation quadratique de cette distance

$$\| \varepsilon_{ijk} \|^2 = \sum_{i,j,k} (g_{ijk} - \lambda_{111} \phi_1^i \phi_1^j \phi_1^k)^2 f_i f_j f_k$$

minimum, ou ce qui est équivalent la maximisation suivante :

$$\sum_{i,j,k} (g_{ijk} \phi_1^i \phi_1^j \phi_1^k) f_i f_j f_k = \lambda_{111}$$

$\lambda_{111}$  apparait donc comme la coordonnée de  $g_{ijk}$  sur le triplet le plus proche. Il est bien sûr nécessaire d'imposer des conditions de normalisation a fonctions de bases.

L'équation  $\| \varepsilon_{ijk} \|^2$  minimum donne les conditions suivantes :

$$\frac{\partial}{\partial \phi_1^i} = \sum_{j,k} \phi_1^j \phi_1^k \left( g_{ijk} - \lambda_{111} \phi_1^i \phi_1^j \phi_1^k \right) f_i f_j f_k = 0 \quad \forall i$$

En tenant compte de la norme unité des fonctions de bases on a :

$$\sum_{j,k} g_{ijk} \phi_1^j \phi_1^k f_i f_j f_k = \lambda_{111} \phi_1^i f_i$$

d'où en revenant aux fréquences.

$$\sum_{j,k} \frac{f_{ijk}}{f_i} \phi_1^j \phi_1^k - \sum_{j,k} \phi_1^j \phi_1^k f_j f_k = \lambda_{111} \phi_1^i$$

A ce niveau deux cas distincts peuvent apparaitre.

— L'un des facteurs est constant.

Prenons par exemple  $\phi^j$  alors l'équation ci-dessus donne :

$$\sum_{j,k} \frac{f_{ijk}}{f_i} \phi_1^k - \sum_k \phi_1^k f_k = \lambda_{111} \phi_1^i$$

$$\sum_k \phi_1^k \frac{f_{ik}}{f_i} = \lambda_{111} \phi_1^i$$

On retrouve la formule de transition des correspondances. La dérivée par rapport à  $\phi^j$  donne :

$$\sum_{i,k} \frac{f_{ijk}}{f_i} \phi_1^i \phi_1^k - 0 = \lambda_{111} = \sum_{i,k} \left( \frac{f_{ijk}}{f_i f_j f_k} - 1 \right) \phi_1^i \phi_1^k f_i f_j f_k$$

qui est en évidence, elle exprime que  $\lambda$  est la projection de  $g_{ijk}$  sur le triplet.

Ce cas qui était prévisible puisque l'on connaissait la forme de la décomposition nous montre qu'il est possible de retrouver les facteurs d'une table de contingence marginale comme facteurs explicatifs d'un tableau sur trois caractères.

Tous les facteurs sont de moyenne nulle.

La première direction de projection appartiendra donc au sous-espace dit de l'interaction.

L'équation devient :

$$\sum_{j,k} \frac{f_{ijk}}{f_i} \phi_1^j \phi_1^k = \lambda_{111} \phi_1^i$$

elle revêt une forme tout à fait analogue

à celle de la formule de transition de l'analyse des correspondances, ici on passe d'un couple  $\phi^j \phi^k$  au facteur  $\phi^i$ .

Les équations symétriques permettent d'opérer une substitution :

$$\begin{aligned} \lambda^2 \phi^i &= \sum_{j,k} \lambda \phi^j \times \frac{f_{ijk}}{f_i} \phi^k \\ &= \sum_{j,k} \left( \sum_{j',k'} \frac{f_{ij'k'}}{f_i} \phi^{j'} \phi^{k'} \right) \frac{f_{ijk}}{f_i} \phi^k \\ &= \sum_{i'} \phi^{i'} \underbrace{\sum_{kk'} \phi^k \phi^{k'}}_3 \sum_{i'} \frac{f_{ij'k'} f_{i'jk}}{f_i f_j} \end{aligned}$$

Le facteur  $\phi^I$  est donc une fonction propre d'un opérateur  $\sigma^I$  construit sur les fréquences observées et sur le facteur associé  $\phi^K$ .

*équation analogue au  $\phi^I$  au lieu de  $\phi^K$*

Difficultés dues à la forme des équations et à la nature des directions de projection.

L'équation précédente permettant de trouver une projection dans l'ensemble des triplets construits avec des fonctions de moyenne nulle peut se résoudre ainsi :

$$\lambda^2 \langle \phi^I, \phi^I \rangle = 1$$

$$\lambda^2 \phi^I = \phi^I \circ \phi^K \circ \sigma_{IK}^{IK} \circ \phi^K$$

avec

$$\sigma_{IK}^{IK} = \left\{ \sum_j \frac{f_{ijk} f_{i'j'k'}}{f_i f_j} \mid i, i' \in I ; k, k' \in K ; j \in J \right\}$$

ou encore :

$$\lambda^2 = \phi^I \circ \phi^K \circ \sigma_{IK}^{IK} \circ \phi^K \circ \phi^I$$

$$\lambda^2 = \phi^K \circ \phi^I \circ \sigma_{IK}^{IK} \circ \phi^I \circ \phi^K$$

S est obtenu en permutant i et k dans l'expression de  $\sigma$ . Notre problème est de rechercher la projection la plus importante. Il s'agit donc de trouver le couple  $\phi^I \phi^K$  donnant sa plus grande valeur à  $\lambda$ , cette question conduisait dans le cas d'un tableau à deux dimensions à rechercher le vecteur sur lequel la projection d'un nuage de points avait la variance ou la dispersion maximum (analyse factorielle) et donc la direction propre d'une matrice de variance-covariance. Dans le cas tridimensionnel on doit s'intéresser au produit tensoriel  $\phi^I \phi^K$  que l'on peut appeler le plus contractant vis-à-vis de l'opérateur  $\sigma_{IK}^{IK}$ . Cet opérateur peut s'interpréter comme une matrice d'inertie d'un nuage de Card J points dans l'espace I x K mais ici on impose des conditions très spéciales pour le vecteur explicatif de  $R^{IK}$ .



Il doit en effet être un élément de  $R^I \times R^K$ .

La résolution des équations est certainement possible par itérations successives. On choisit un facteur  $\phi^K$  et l'on détermine le facteur  $\phi^I$  associé à la plus forte valeur propre de  $\phi^K \circ \sigma_{IK}^{IK} \circ \phi^K$ . A partir de cette solution on construit  $\phi_{IK}^I \circ S_{IK}^{IK} \circ \phi^I$ . La seconde valeur de  $\phi^K$  sera la première fonction propre de cet opérateur. La convergence de cette procédure n'est pas démontrée mais il en existe des exemples (cf l'étude de Carroll et Chang).

A cette difficulté de résolution s'ajoute celle des conditions supplémentaires qu'il y a lieu d'introduire lorsque l'on cherche les triplets suivants. Si l'on peut montrer facilement qu'une maximisation sur la fonction résiduelle  $(g_{IJK} - \lambda_{111} \phi_1^I \phi_1^J \phi_1^K)$  donnerait un triplet orthogonal au précédent, on n'est pas du tout assuré, par contre, de la complète orthogonalité.

On doit donc tenir compte du fait que le facteur suivant sur un ensemble sera identique ou orthogonal au facteur précédent.

Au problème de convergence et de durée de calcul s'ajoute donc celui de la lourdeur et de la multiplication des contraintes. Toutes ces difficultés résultent du fait qu'un élément de  $R^I \times R^J \times R^K$  n'engendre pas un sous-espace de  $R^{IJK}$ . Le modèle proposé a cependant le mérite d'utiliser des bases simples qui aideront à la compréhension des futures opérations et transformations que l'on rencontrera sur ce genre de tableau.

Pour obtenir rapidement des renseignements concrets sur la structure d'un tableau sur trois caractères on peut s'intéresser à la "proximité" entre les termes du développement si délicat à calculer.

### LA NOTION DE CORRELATION INTERNE

#### Définition et propriétés.

La question que l'on peut se poser est de comparer les fluctuations relatives d'une fonction à trois indices  $g_{IJK}$  et d'une fonction à un indice  $\phi^I$ .

Dans notre décomposition cette question revient à se demander si un facteur  $\phi^I$  apparaît dans de nombreux triplets et avec des coefficients  $\{\lambda_{pmm} \mid m \in M, n \in N\}$  importants.

Pour répondre à cette question on considère la fonction à deux indices.

$$g_{JK} = \sum_{i \in I} g_{iJK} \phi^i f_i = \langle g_{iJK}, \phi^i \rangle_{f_i}$$

Elle peut s'interpréter comme la projection de  $g_{IJK}$  sur l'espace  $R_{JK}$  perpendiculairement à  $\phi^I$ .

De même  $\phi^I g_{JK}$  sera la projection orthogonale sur  $\phi^I R^{JK}$  de la fonction  $g_{IJK}$ , on vérifie aisément que  $(g_{IJK} - \phi^I g_{JK})$  est orthogonal à tout élément de  $\phi^I R^{JK}$  car  $\langle g_{IJK} - \phi^I g_{JK}, \phi^I \rangle_{f_i} = 0$ . ) o w

La proximité entre la fonction et sa projection peut s'exprimer par un rapport de norme, on définit donc le coefficient de corrélation interne  $\rho$  par la formule suivante :

$$\rho = \frac{\|\phi^I g_{JK}\|}{\|g_{JK}\|} = \frac{\| \langle g_{JK}, \phi^I \rangle \|}{\|g_{JK}\|}$$

Cette corrélation interne peut s'interpréter facilement dans le cas où  $\phi^I$  est une des fonctions de bases  $\phi_p^I$  dans la décomposition de  $g_{IJK}$ .

$$g_{IJK} = \sum_{p,m,n} \lambda_{p,m,n} \phi_p^I \phi_m^J \phi_n^K$$

effectuons les calculs menant à  $g_{JK}$  et  $\rho$  en tenant compte de l'orthogonalité des fonctions de bases.

$$g_{JK} = \sum_{m',n'} \lambda_{p,m',n'} \phi_{m'}^J \phi_{n'}^K$$

Les indices  $m' n'$  désignent les facteurs sur J et K associés à  $\phi_p^I$  dans la décomposition. On pourrait conserver le même ensemble de variation pour ces indices en admettant que certains coefficients  $\lambda_{p,m,n}$  peuvent être nuls.

$$\|\phi_p^I g_{JK}\|^2 = \|g_{JK}\|^2 = \sum_{m,n} \lambda_{p,m,n}^2$$

C'est donc la somme des carrés des coordonnées associées au facteur  $\phi_p^I$ , en terme statistique cette somme peut s'interpréter comme la variance associée au facteur sur I correspondant à l'indice  $p$ .

$$\|g_{IJK}\|^2 = \sum_{p,m,n} \lambda_{p,m,n}^2$$

Variance totale de la fonction écart.

$$\rho^2 = \frac{\text{Variance associée à l'indice factoriel } p}{\text{Variance totale}}$$

Ce coefficient de corrélation interne est donc compris entre zéro et un, il représente bien l'importance que le facteur  $\phi_p^I$  a dans la décomposition vis-à-vis des autres facteurs sur l'ensemble I.

On pourrait étendre cette définition pour comparer un groupe de facteurs  $\{\phi_1^I, \phi_2^I, \dots\}$  orthogonaux à une fonction  $g_{IJK}$ . Alors le carré de la corrélation associée au groupe serait la somme des carrés des corrélations définies par chaque facteur.

On pourrait aussi approcher le problème de la décomposition en recherchant les facteurs sur un ensemble, les mieux corrélés avec la fonction à expliquer. Cette approche associant un caractère au produit des deux autres a déjà donné lieu à de nombreux résultats (cf. Benzecri, Tucker).

#### Résultats sous forme d'une table de corrélation.

La corrélation interne que l'on vient de définir et la corrélation classique entre facteurs sur un même ensemble permettent de juger facilement de la proximité entre facteurs et fonctions à un ou plusieurs indices.

Ainsi pour éviter les difficultés pratiques liées au calcul d'une décomposition complète on pourra se contenter de calculer certains facteurs significatifs, les plus influents dans les contingences marginales en particulier, et juger de leur corrélation entre eux et avec l'interaction qui n'aura pas été décomposée.

En effet les facteurs extraits dans une décomposition binaire sont des combinaisons linéaires des facteurs qui correspondraient à la décomposition ternaire. Ainsi lorsqu'on étudie les contingences IK et JK on obtient des facteurs sur K qui ne sont pas, sauf cas particulier, identiques ou orthogonaux. La seule façon de les comparer est de calculer leur corrélation.

Pour l'interaction on calcule toutes les corrélations internes avec les facteurs qui expliquent le plus de variance sur les marges entre deux caractères. Si la somme des carrés des corrélations internes correspondant à un groupe de facteurs orthonomés sur un indice est proche de 1, on pourra estimer que l'on peut expliquer les variations de l'interaction pour cet indice à l'aide de ce groupe de facteurs.

Les tableaux de résultats pourront se présenter ainsi :

Facteur sur I

|                   |               | $\phi_1^I$  | $\phi_2^I$  | $\phi_1^{I'}$ | $\phi_2^{I'}$ | Interaction | Valeur Propre (Coordonnée) |
|-------------------|---------------|-------------|-------------|---------------|---------------|-------------|----------------------------|
| Contingence<br>IJ | $\phi_1^I$    | 1           | 0           | $\rho_1$      | $\rho_3$      | $\rho_{11}$ | $\lambda_1$                |
|                   | $\phi_2^I$    | 0           | 1           | $\rho_2$      | $\rho_4$      | $\rho_{12}$ | $\lambda_2 < \lambda_1$    |
| Contingence<br>IK | $\phi_1^{I'}$ | $\rho_1$    | $\rho_2$    | 1             | 0             | $\rho_{13}$ | $\lambda'_1$               |
|                   | $\phi_2^{I'}$ | $\rho_3$    | $\rho_4$    | 0             | 1             | $\rho_{14}$ | $\lambda'_2 < \lambda'_1$  |
| Interaction       |               | $\rho_{11}$ | $\rho_{12}$ | $\rho_{13}$   | $\rho_{14}$   |             |                            |

Si  $\rho_{11}^2 + \rho_{12}^2 \approx 1$ . La variance de l'intéraction peut être expliquée par les facteurs  $\phi_1^I$  et  $\phi_2^I$ .

Dans une décomposition totale l'intéraction  $(\sum_{\substack{e \neq 0 \\ m \neq 0 \\ n \neq 0}} \lambda_{emn} \phi_e^i \phi_m^j \phi_n^k)$  sera essentiellement constituée par des facteurs intervenant dans  $\phi_1^I = \sum_{e \in L_1} \alpha_e \phi_e^i$

et  $\phi_2^I = \sum_{e \in L_2} \beta_e \phi_e^i$ .

Deux tables analogues pour les facteurs sur J et sur K compléterai la description de toutes les proximités existantes entre les différentes composant de la fonction écart.

#### I4 - AUTRES DECOMPOSITIONS

La solution que nous avons choisie pour expliquer la fonction d'écart s'appuie à la fois sur un modèle explicatif d'une forme très particulière (bases orthonormées avec des fonctions constantes ou de moyenne nulle) et sur le critère du pouvoir explicatif maximum pour la détermination des triplets successifs.

Certaines approches de ce problème de décomposition ont déjà trouvé leur solution et sont couramment utilisées. Nous exposerons les trois principales qui ne conduisent pas à des calculs prohibitifs.

Ces approches posent un nombre réduit de conditions a priori sur le modèle explicatif ce qui n'entraîne jamais de trop grosses difficultés dans la recherche numérique d'une solution particulière.

Les deux premières (L.R. TUCKER et J.P. BENZECRI) donnent des bases de facteurs orthonormées, la seconde (J.D. CARROLL et J.J. CHANG) avec une solution du type des moindres carrés, recherche un pouvoir explicatif maximum.

#### I4I - Modèle fondamental pour l'Analyse Factorielle à trois modes (LR TUC

Les données sont constituées par un tableau à trois entrées: on notera  $x_{ijk}$  un de ses éléments, les caractères permettant la classification d'un résultat de comptage ou de mesure seront toujours notés I, J et K.

Le modèle d'analyse factorielle retenu par L.R. TUCKER est le suivant :

$$x_{ijk} = \hat{x}_{ijk} + e_{ijk}$$

$\uparrow$                                    $\uparrow$   
 Observation          Ecart au modèle

$$\hat{x}_{ijk} = \sum_e \sum_m \sum_n a_{ie} b_{jm} c_{kn} g_{emn}$$

Formellement il y a équivalence entre cette écriture et la décomposition du paragraphe précédent avec  $a_{ie} = \phi_e^i$  ;  $b_{jm} = \phi_m^j$   
 $c_{kn} = \phi_n^k$  ;  $g_{emn} = \lambda_{emn}$ .

Ce modèle permet de passer des caractères I, J, K aux caractères L M N (ensemble de variation des indices e, m, n) que l'on espère être plus intéressants pour l'explication du tableau étudié. On peut même supposer qu'ils représentent des caractères idéaux selon lesquels les modalités observées se rangent de façon facilement interprétable, et surtout on espère que le modèle conduira à un nombre très réduit de nouvelles modalités e, m et n.

La fonction de trois indices  $\hat{x}_{ijk}$  est supposée égale au produit de trois fonctions à deux indices et d'une fonction  $g_{emn}$  à trois indices qui définit un coefficient d'amplitude pour les associations des nouveaux caractères.



La détermination des éléments factoriels ne se fait pas selon un critère de pouvoir explicatif maximum vis-à-vis de la variance de la fonction  $\hat{x}_{ijk}$ . La méthode utilisée pourrait s'appeler la méthode des directions propres des trois matrices d'inertie.

En effet on remarque que l'on peut écrire :

$$\hat{x}_{ijk} = \sum_e a_{ie} \sum_m \sum_n g_{emn} b_{jm} c_{kn}$$

$${}_I X_{JK} = {}_I A_{L \times L} A'_{JK}$$

Le tableau à trois entrées donne naissance à une matrice carrée dont les lignes sont définies par les indices  $i$ , et les colonnes par les couples aux associations  $j, k$ , en introduisant les matrices transposées on a :

$${}_I J^{X_{JK}} \times JK^{X_I} = {}_I A_{L \times L} A'_{JK} \times JK^{A'_{L \times L}} A_I$$

Sous certaines conditions de centrage, le premier terme peut s'interpréter comme la matrice d'inertie de Card  $J \times$  Card  $K$  points dans un espace de dimension card  $I$  ou comme la matrice de variance-covariance observée de Card  $I$  variables aléatoires dont on possède Card  $J \times$  card  $K$  réalisations.

Le second terme peut se construire de plusieurs façons, on reconnaît là une décomposition classique de l'analyse factorielle d'un nuage de points. Une solution très intéressante est celle des vecteurs propres de la matrice d'inertie. Elle donne les résultats suivants :

${}_{L'} A'_{JK} \times JKA'_{L'}$  est diagonale avec  $\lambda_{\rho}^2$  inertie sur l'axe factoriel d'indice  $\rho$

${}_{I'} A_{L'}$  définit les directions des card L axes factoriels dans l'espace à Card I dimensions.

On a  ${}_{L'} A_{I'} \times {}_{I'} A_{L'} = {}_{L'} I_{L'}$  matrice unité de dimension card L x card L.

Les axes factoriels sont orthogonaux.

$\text{Rang } {}_{I'} A_{L'} = \text{Rang } ({}_{I'} X_{JK} \text{ } {}_{I'} X_{JK})$  d'où  $\text{card } L \leq \min(\text{card } I, \text{card } J \times \text{card } K)$ .

Les matrices  ${}_{J'} B_M$  et  ${}_{K'} C_N$  peuvent se calculer selon le même principe. On obtiendra des matrices de vecteurs orthonormés de rang card I et card N tel que :

$$\text{card } M \leq \min(\text{card } J, \text{card } I \times \text{card } K) = \text{card } J$$

$$\text{card } N \leq \min(\text{card } K, \text{card } I \times \text{card } J) = \text{card } K$$

On a donc ainsi constitué trois bases de fonctions sur I, J et K. Les associations ternaires de ces fonctions de bases constituent une base des fonctions à trois indices, avec ce principe de détermination on aboutit donc à  $\min(\text{card } I, \text{card } J \times \text{card } K) \times \text{card } J \times \text{card } K$  triplets au maximum.

La matrice des coefficients  $g_{emn}$  se détermine aisément en tenant compte de l'orthonormalité des vecteurs de A, B et C.

$$g_{pmn} = \sum_i \sum_j \sum_k \tilde{x}_{ijk} a_{if} b_{jm} c_{kn}$$

Puisque

$$\sum_i a_{if} a_{if'} = \delta_{f'f}^p$$

$$\sum_j b_{jm} b_{jm'} = \delta_{m'm}^m$$

$$\sum_k c_{kn} c_{kn'} = \delta_{n'n}^n$$

Une autre conséquence importante de cette propriété concerne les traces des matrices :

$$p A'_{jk} x_{jk} A'_{f'} = \sum_{jk} \sum_{\substack{m,n \\ m,n}} g_{pmn} g_{p'm'n'} b_{jm} c_{kn} b_{j'm'} c_{k'n'} = \sum_{m,n} g_{pmn} g_{p'm'n'}$$

$$\text{Trace}({}_I X_{jk} x_{jk} X_I) = \text{Trace}({}_L A'_{jk} x_{jk} A'_L)$$

$$\text{car } ({}_I A_L)^{-1} = {}_L A_I$$

d'où

$$\sum_i \sum_{jk} x_{ijk}^2 = \sum_f \sum_{m,n} g_{fpmn}^2$$

Les matrices X et G représentent bien la même variance totale.

Appliquons la méthode préconisée par L.R. TUCKER à l'analyse de notre fonction écart.

Les facteurs  $\phi_p^I, \phi_m^J$  et  $\phi_n^K$  doivent être orthonormés selon "métrique des poids" c'est-à-dire relativement à  $f_i, f_j$  et  $f_k$

$$g_{ijk} = \frac{f_{ijk}}{f_i f_j f_k} = \sum \lambda_{pmn} \phi_p^i \phi_m^j \phi_n^k$$

$$\sum_{jk} g_{ijk} g_{i'jk} f_j f_k = \sum_{jk} \sum_{\substack{p,mn \\ p',m'n'}} \lambda_{pmn} \lambda_{p'm'n'} \phi_p^i \phi_{p'}^{i'} (\phi_m^j \phi_{m'}^j f_j) (\phi_n^k \phi_{n'}^k f_k)$$

$$\sum_{jk} \frac{f_{ijk} f_{i'jk}}{f_i f_j f_k f_{i'} f_j f_k} = \sum_{p',m,n} \lambda_{pmn} \lambda_{p'm'n'} \phi_p^i \phi_{p'}^{i'} = \lambda_\alpha^2 \phi_\alpha^i \phi_\alpha^{i'}$$

$\sum_p \lambda_{pmn} \phi_p^i = \psi_{mn}^i$   
 $\sum_{m,n} \psi_{mn}^i \psi_{mn}^{i'}$

Il y a donc lieu d'effectuer une décomposition de l'opérateur

$$\sigma_{i,i'} = \left\{ \sum_{jk} \frac{f_{ijk} f_{i'jk}}{f_i f_j f_k f_{i'} f_j f_k} \mid j \in J, k \in K, i, i' \in I \right\}$$

Les facteurs  $\phi^I$  que l'on obtient permettent le développement suivant :

$$\frac{f_{ijk}}{f_i f_j f_k} = \sum_\alpha \lambda_\alpha \phi_\alpha^i \phi_\alpha^{jk}$$

avec  $\lambda_\alpha \phi_\alpha^{jk} = \sum_i \frac{f_{ijk}}{f_j f_k} \phi_\alpha^i$

On a donc affaire à une décomposition du type de celle de l'analyse de correspondances binaires.

Elle permet d'écrire la fonction  $g_{IJK}$  comme une somme de produits de fonctions de  $R^I$  et de  $R^J \times K$  munis respectivement des métriques  $f_I$  et  $f_I \otimes f_K$ .

Le fait que la métrique utilisée sur  $R^{JK}$  ne soit pas la métrique  $f_{JK}$  qui se déduit de la densité  $f_{IJK}$  complique considérablement les résultats que peut donner cette décomposition.

En effet le premier couple de facteurs n'est pas constitué de fonctions constantes.

$$\sum_{ijk} \frac{f_{ijk}}{f_i f_j f_k} \phi_\alpha^i \phi_\alpha^j \phi_\alpha^k f_i f_j f_k = \sum_{ijk} f_{ijk} \phi_\alpha^i \phi_\alpha^j \phi_\alpha^k = \lambda_\alpha$$

n'est pas une corrélation car

$$\|\phi^{jk}\|_{f_{ijk}}^2 = \sum_{ijk} f_{ijk} \phi^{jk} \phi^{jk} \neq \sum_{ik} f_i f_k \phi^{jk} \phi^{jk} = 1$$

et  $\lambda$  n'est pas nécessairement inférieur ou égal à 1 pour toutes autres fonctions que  $1^I$  et  $1^{JK}$ .

On peut voir immédiatement que les facteurs sur I ne sont pas tous de moyenne nulle pour la décomposition de  $g'_{ijk} = \frac{f_{ijk}}{f_i f_j f_k} - 1$

$$\sum_i g'_{ijk} f_i = \frac{f_{jk}}{f_j f_k} - 1 = \sum_i \lambda_\alpha \phi_\alpha^{jk} \text{ moy}(\phi_\alpha^i) \neq 0$$

Cette procédure ne pose donc pas de problème de calcul mais elle fournit des facteurs qui ne sont pas aussi facilement interprétables que dans le cas précédent.

On peut remarquer que cette méthode ne fait que rechercher des facteurs  $\phi^I$  tels que :

$$\sum_i g_{iik} \phi^i = \lambda \phi^{ik} \quad \text{avec} \quad \|\phi^{ik}\|_{f_i f_k} = 1 \quad \text{et} \quad \lambda \quad \text{maxi}$$

donc

$$\| \langle g_{IJK}, \phi^I \rangle_{f_I} \|_{f_J f_K} = \text{corrélation interne } (g, \phi^I) \text{ maximum}$$

Et qu'avec card J card K min (card I, card J x card K) triplets elle permet de reconstituer une fonction de  $R^{IJK}$ ; en effet on a :

$$g_{IJK} = \sum_p \lambda_p \phi_p^I \phi_p^{JK}$$

et ayant card J facteurs  $\phi^J$  et card K facteurs  $\phi^K$  on peut toujours décomposer les fonctions  $\phi_p^{JK}$  de  $R^{JK}$

$$\phi_p^{JK} = \sum_m \mu_{mn} \phi_m^J \phi_n^K$$

D'où

$$g_{IJK} = \sum_{p,m,n} \lambda_{p,m,n} \phi_p^I \phi_m^J \phi_n^K$$

avec généralement

$$\text{Card L} \times \text{Card M} \times \text{Card N} \ll \text{Card I} \times \text{Card J} \times \text{Card K}$$

Le professeur J.P. BENZECRI conseille d'étudier les tables de contingences entre trois caractères à l'aide de trois analyses factorielles.

Chaque analyse étudie une contingence entre un caractère et l'association des deux autres. En effet à partir d'une table (card I) x (card J) x (card K) on peut construire trois tables carrées (card I) x (card J x card K) ; (card J) x (card I x card K) ; (card K) x (card I x card J) sans perdre aucune information sur les données.

Il est alors naturel de décomposer les fonctions d'écart

$\frac{f_{ijk}}{f_i f_{j.} f_{.k}}$ ,  $\frac{f_{ikj}}{f_i f_{.k} f_{j.}}$ ,  $\frac{f_{kij}}{f_{.k} f_{i.} f_{j.}}$  entre l'observation et la densité de fréquence résultant de l'hypothèse d'indépendance entre un caractère et les deux autres.

Chaque décomposition définit un ensemble de facteurs ortho-normés définis sur un caractère.

On dispose donc en fin de calcul de :

$$\{ \phi_1^I, \phi_2^I \dots \phi_L^I \mid L = \min(\text{card I}, \text{card J} \times \text{card K}) - 1 \}$$

$$\{ \phi_1^J, \phi_2^J \dots \phi_M^J \mid M = \min(\text{card J}, \text{card I} \times \text{card K}) - 1 = \text{card J} - 1 \}$$

$$\{ \phi_1^K, \phi_2^K \dots \phi_N^K \mid N = \min(\text{card K}, \text{card I} \times \text{card J}) - 1 = \text{card K} - 1 \}$$

Tous les premiers facteurs sont des fonctions constantes, puisqu'ils sont obtenus dans des décompositions binaires classiques.

Pour effectuer ces décompositions on utilise des métriques ou pondérations telles que  $f_{IJ}$ , dont certains éléments peuvent être nuls. Il y a donc une possibilité de dégénérescence, ce qui était impossible auparavant puisque les pondérations  $f_I f_J f_K$  représentaient l'importance de modalités de I, J et K et que l'on supposait que ces modalités étaient effectivement rencontrées.

Cette difficulté théorique n'a pas d'incidence sur la détermination des facteurs sur I.

Les facteurs vérifient l'équation

$$\sum_i \sum_{i'k} \frac{f_{iik} f_{i'ik}}{f_i f_{i'k}} \phi^i = \lambda^2 \phi^{i'} = \sum_i \sigma_i^{i'} \phi^i$$

Si  $f_{jk} = 0$  à fortiori  $f_{ijk}$  et  $f_{i'jk}$  sont nuls le couple  $jk$  fournira donc une contribution nulle au terme  $\sigma_i^{i'}$ , et les facteurs sur I pourront être déterminés sans problème.

La solution de l'analyse des correspondances ternaires semble donc éviter les difficultés rencontrées dans la méthode de L.R. TUCKER.

Pour pouvoir affirmer que cette procédure répond mieux que la précédente à la question que nous nous étions posée, il faut vérifier que les ensembles de facteurs orthonormés permettent bien une décomposition complète de la fonction écart que nous voulions étudier.



Nous disposons en fin de calcul de la décomposition

$$\frac{f_{ijk}}{f_i f_{jk}} = \sum_p \lambda_p \phi_p^i \phi_p^{jk}$$

d'une base complète de  $R^J = \{ \delta^j, \phi_m^j \mid m \in M \}$

d'une base complète de  $R^K = \{ \delta^k, \phi_n^k \mid n \in N \}$

$$\begin{aligned} \text{d'où } \frac{f_{ijk}}{f_i f_j f_k} &= \frac{f_{ijk}}{f_i f_{jk}} \times \frac{f_{jk}}{f_j f_k} = \sum_p \lambda_p \phi_p^i \left( \phi_p^{jk} \frac{f_{jk}}{f_j f_k} \right) \\ &= \sum_p \lambda_p \phi_p^i \phi_p^{jk} \end{aligned}$$

$\phi^{JK}$  est décomposable sur  $\{ 1^j 1^k, 1^j \phi_n^k, 1^k \phi_m^j, \phi_m^j \phi_n^k \mid m \in M, n \in N \}$

La seule condition pour que cette procédure donne tous les résultats souhaités est que nous ayons des bases complètes pour les fonctions sur les deux ensembles de modalités d'effectifs les plus faibles.

C'est généralement ce qui se produit car le cas contraire supposerait des relations linéaires entre les card I x card K réalisations d card J variables aléatoires correspondant au caractère J, ou entre les card I x card J réalisations liées aux card K modalités k.

Ces conditions s'imposent de la même façon dans la méthode précédente.

La dernière inquiétude que l'on peut avoir ne concerne plus les conditions imposées dans le modèle. Elle concerne la qualité de l'interprétation que l'on peut faire des facteurs sur un ensemble sachant qu'ils ont été définis dans une contingence avec les deux autres ensembles.

Par hypothèse les tableaux étudiés s'appuient sur trois ensembles homogènes de modalités, mais rien ne garantit le sens des associations de modalités sur deux caractères différents.

#### I43 - DECOMPOSITION CANONIQUE D'UN TABLEAU A TROIS ENTREES

Cette appellation et la méthode de calcul proposées sont dues à MM. J.D. CARROLL et JJ. CHANG.

Le modèle est le suivant :

$$x_{ijk} = \sum_{t \in T} a_{it} b_{jt} c_{kt}$$

On voit donc qu'il n'y a plus qu'un indice pour classer les facteurs sur les trois ensembles I, J et K. Aucune condition d'orthogonalité n'est imposée aux facteurs a, b et c. La détermination des éléments multiplicatifs se fait par une méthode dite NILES (Nonlinear Iterative Least Squares) elle consiste à choisir arbitrairement les deux matrices  $B = \{b_{jt}\}$  et  $C = \{c_{kt}\}$  et déterminer  $A = \{a_{it}\}$  par la condition

$$\| X - A \times (B \otimes C) \| \text{ minimum}$$

avec  $X = \{ x_{ijk} \}$  et  $B \otimes C = Q = \{ b_{je} c_{ke} \}$

X est de dimension card I x (card J x card K)

Q..... (card J x card K) x card T

$$\| X - AQ \|^2 = \sum_{ijk} \left( x_{ijk} - \sum_e a_{ie} \times (b_{je} c_{ke}) \right)^2$$

La solution des moindres carrés donne :

$$A = (X \times Q)(Q^T Q)^{-1}$$

avec

$$X \times Q = \left\{ p_{ie} = \sum_j \sum_k x_{ijk} b_{je} c_{ke} \right\}$$

$$(Q^T Q) = \left\{ q_{ee'} = \sum_j b_{je} b_{je'} \times \sum_k c_{ke} c_{ke'} \right\}$$

A partir de cette matrice A et de la matrice C arbitraire on peut calculer une matrice B par la même méthode des moindres carrés. A l'aide de A et B on trouve C optimum et l'on répète ces opérations jusqu'à ce qu'on observe une certaine stabilité dans les résultats. La convergence de ces itérations n'est pas démontrée, mais elle a été observée sur de nombreux exemples jusqu'à des tables de contingence entre sept caractères.

Cette procédure est analogue à la régression multiple et les calculs itératifs sur les matrices peuvent s'identifier à la recherche du "produit tensoriel le plus contractant" du paragraphe I33.

La technique NILES peut facilement s'appliquer à notre fonction écart  $(f_{ijk}/f_i f_j f_k)$ , il faudra seulement tenir compte de la nouvelle forme de la métrique.

Ceci peut se faire en effectuant des homothéties sur les axes de  $R^{IJK}$ . Les fonctions :

$$\phi_{\epsilon}^i = \phi_{\epsilon}^i \sqrt{f_i} \quad , \quad \phi_{\epsilon}^j = \phi_{\epsilon}^j \sqrt{f_j} \quad , \quad \phi_{\epsilon}^k = \phi_{\epsilon}^k \sqrt{f_k}$$

seront des éléments de bases orthogonaux dans une métrique définie par une matrice unité I .

$$d'où \quad \frac{f_{ijk}}{(f_i f_j f_k)^{1/2}} = \sum_{\epsilon} \phi_{\epsilon}^i \phi_{\epsilon}^j \phi_{\epsilon}^k$$

On pourrait tenter de retrouver notre décomposition initiale en posant :

$$t = p + (\text{card } L-1) m + (\text{card } L-1) \times (\text{card } M-1) n$$

$$\phi_n^k = \phi_{t'}^k \quad t' = t \bmod [(\text{card } L-1) \times (\text{card } M-1)]$$

$$\phi_m^j = \phi_{t''}^j \quad t'' = (t - t' (\text{card } L-1) \times (\text{card } M-1)) \bmod (\text{card } L-1)$$

$$\lambda_{emn}^i \phi_{\epsilon}^i = \phi_{t'''}^i, t''' = t - t' (\text{card } L-1) \times (\text{card } M-1) - t'' (\text{card } L-1)$$

En supposant que les facteurs sur J et K sont normés et que le coefficient d'amplitude  $\lambda_{emn}^i$  n'agit que sur le facteur sur I.

On peut s'apercevoir rapidement que l'absence de condition de complète orthogonalité entre les triplets fait que l'on ne peut pas obtenir la même décomposition par les deux méthodes.

Considérons :

$$\frac{f_{ijk}}{f_i f_j f_k} = 1^i 1^j 1^k + \lambda 1^i \phi_1^j \phi_1^k + \lambda' \phi_2^i \phi_1^j \phi_1^k$$

une possible décomposition dans la base dite de l'analyse de la variance. Il est nécessaire d'introduire les deux triplets  $1^i \phi_1^j \phi_1^k$  et  $\phi_2^i \phi_1^j \phi_1^k$  qui sont orthogonaux. La variance expliquée est  $\lambda^2 + \lambda'^2$ .

Dans la méthode des moindres carrés un seul triplet suffirait on aurait :

$$\frac{f_{ijk}}{f_i f_j f_k} = 1^i 1^j 1^k + \phi_2^i \phi_1^j \phi_1^k \quad \text{avec} \quad \phi_2^i = \lambda' \phi_1^i + \lambda 1^i$$

La variance expliquée sera :

$$\| \phi_2^i \phi_1^j \phi_1^k \|^2 = \lambda'^2 + \lambda^2$$

Evidemment la condition manquante est :

$$\langle \phi_2^i, 1^i \rangle = 0 \quad \text{ou} \quad 1$$

Pour comparer le pouvoir explicatif des différents modèles remarquons que d'une façon générale lorsque l'on recherche des bases sur  $R^I, R^J$  et  $R^K$  et que l'on dispose déjà de :

$$\left\{ \begin{array}{l} \phi_1^I \phi_1^J \phi_1^K \\ \phi_2^I \phi_2^J \phi_2^K \\ \phi_3^I \phi_3^J \phi_3^K \end{array} \right.$$

Il y a lieu d'introduire les coordonnées de

$$\frac{f_{ijk}}{f_i f_j f_k} \quad \text{sur} \quad \left\{ \begin{array}{l} \phi_1^I \phi_1^J \phi_2^K \\ \phi_1^I \phi_1^J \phi_3^K \\ \text{etc...} \end{array} \right.$$

Pour la méthode des moindres carrés il ne peut apparaître qu'un seul couple comportant par exemple les deux fonctions  $\phi_1^J \phi_1^K$ . Il pourrait être de la forme

$$\phi_1^J \phi_1^K (\lambda_{111} \phi_1^I + \lambda_{211} \phi_2^I + \dots + \lambda_{L11} \phi_L^I) = \lambda_{111} \phi_1^I \phi_1^J \phi_1^K$$

$$\text{avec} \quad \phi_1^I = \left\langle \frac{f_{ijk}}{f_i f_j f_k}, \phi_1^J \phi_1^K \right\rangle$$

La décomposition de CARROLL et CHANG aura au maximum  $\min(\text{card } I \times \text{card } J ; \text{card } I \times \text{card } K ; \text{card } J \times \text{card } K) = (\text{card } J \times \text{card } K)$  termes, mais malheureusement les suites de facteurs sur un ensemble n'auront pas les propriétés d'orthonormalité nécessaire à une bonne interprétation des résultats.

## II - VARIANCES PARTIELLES ET SOUS-HYPOTHESES

### - Conséquences de la décomposition en sous-espaces orthogonaux

L'étude de la fonction d'écart dont la variance est proportionnelle au  $\chi^2$  de contingence du tableau à trois entrées, nous a conduit à décomposer la décomposition suivante :

$$\left( \frac{f_{ijk}}{f_i f_j f_k} - 1 \right) = \sum_{p,m} \lambda_{p m 0} \phi_p^i \phi_m^j 1^k + \sum_{p,n} \lambda_{p 0 n} \phi_p^i 1^j \phi_n^k + \sum_{m,n} \lambda_{0 m n} 1^i \phi_m^j \phi_n^k + \sum_{\substack{p \neq 0 \\ m \neq 0 \\ n \neq 0}} \lambda_{p m n} \phi_p^i \phi_m^j \phi_n^k$$

Cette décomposition s'est effectuée à l'aide des bases :

$$\{ 1^I, \phi_p^I \mid p \in 1, \text{Card } I - 1 \}$$

$$\{ 1^J, \phi_m^J \mid m \in 1, \text{Card } J - 1 \}$$

$$\{ 1^K, \phi_n^K \mid n \in 1, \text{Card } K - 1 \}$$

pour les espaces de fonctions  $R^I, R^J, R^K$

Ces bases sont respectivement orthonormées pour les métriques définies par  $f_I, f_J$  et  $f_K$ .

La variance d'une fonction de  $R^{IJK}$  est liée aux coefficients d'amplitude  $\lambda_{emn}$ , qui sont ses coordonnées sur les triplets associés aux indices  $e, m, n$ .

En effet considérons le  $\chi^2$  de contingence qui permet d'accepter ou de rejeter l'hypothèse d'indépendance globale.

$$\chi_{ijk}^2 = N \sum_{ijk} \left( \frac{f_{ijk}}{f_i f_j f_k} - 1 \right)^2 f_i f_j f_k = N \left( \sum_{pm} \lambda_{pm}^2 + \sum_{pn} \lambda_{pn}^2 + \sum_{mn} \lambda_{mn}^2 + \sum_{\substack{pmn \\ p \neq 0}} \lambda_{pmn}^2 \right)$$

N désigne ici l'effectif total du tableau de comptage.

Dans le calcul de cette variance qui n'est autre que le carré scalaire de la fonction écart tous les produits scalaires de triplets de bases différents sont nuls. En effet un tel produit scalaire fait apparaître au moins un produit entre deux facteurs différents sur un même espace.

Les carrés scalaires sont égaux à un du fait de la normalité des fonctions de base, leur coefficient est égal au carré de la coordonnée de la fonction écart sur le triplet considéré.

La variance totale de la fonction écart ou ce qui est équivalent le  $\chi^2$  de contingence ternaire se décompose en quatre termes associés à des sous-espaces de projection particuliers.

Prenons le premier terme qui peut être considéré comme le carré de la longueur de la projection de  $\left\{ \frac{f_{ijk}}{f_i f_j f_k} - 1 \in R^{ijk} \right\}$  sur le sous-espace  $R^{IJ}$  des fonctions ne dépendant que des indices  $i$  et  $j$ , il peut se réécrire comme suit :

$$\begin{aligned} \sum_{p,m} \lambda_{pm}^2 &= \sum_{ijk} \left( \sum_{p,m} \lambda_{pm} \phi_p^i \phi_m^j 1^k \right) \left( \sum_{p,m} \lambda_{pm} \phi_p^i \phi_m^j 1^k \right) f_i f_j f_k \\ &= \sum_{ijk} \left( \frac{f_{ij}}{f_i f_j} - 1 \right) \left( \frac{f_{ij}}{f_i f_j} - 1 \right) f_i f_j f_k = \sum_{ij} \left( \frac{f_{ij}}{f_i f_j} - 1 \right)^2 f_i f_j = \chi_{IJ}^2 \end{aligned}$$



On reconnaît là le  $\chi^2$  de contingence binaire qui correspond à la face I x J du tableau à trois entrées. Les deux termes suivants donnent  $\chi^2_{JK}$  et  $\chi^2_{IK}$  correspondant aux deux autres faces. Le quatrième terme est un résidu fonction de ce que nous avons défini comme l'interaction entre les trois caractères. Cette variance de l'interaction mesure l'importance de la projection de la fonction écart dans le sous espace que nous avons appelé  $(V^I \otimes V^J \otimes V^K)$ . C'est ce qu'il manque pour pouvoir expliquer une table à trois dimensions à l'aide de ses trois faces bidimensionnelles.

### Sous-Hypothèses de l'indépendance totale

La statistique numérique qui permet d'apprécier le bien fondé de l'hypothèse d'indépendance totale, se décompose donc en quatre termes que l'on peut facilement interpréter et qui se rattachent à certaines hypothèses concernant des sous-ensembles du tableau à trois entrées. Ainsi les trois premiers termes concernent les tableaux marginaux de contingence entre deux caractères ils constituent des statistiques permettant de tester les hypothèses d'indépendance entre deux caractères.

Cette décomposition de la variance incite à définir des sous-hypothèses additives qui correspondraient aux différentes parties.

$$\chi^2_{IJK} = \chi^2_{IJ} + \chi^2_{IK} + \chi^2_{JK} + \chi^2_{Int}$$

$$\begin{array}{l} \text{Indépendance} \\ I \times J \times K \end{array} = \begin{array}{l} \text{Indépend.} \\ I \times J \end{array} \cup \begin{array}{l} \text{Indépend.} \\ I \times K \end{array} \cup \begin{array}{l} \text{Indépend.} \\ J \times K \end{array} \cup \begin{array}{l} \text{Non Intéragi} \\ I \times J \times K \end{array}$$

Cette définition est admissible car on a bien l'implication

$$\begin{array}{l}
 \text{Indépendance} \\
 \text{totale} \\
 p_{ijk} = p_i p_j p_k
 \end{array}
 \left. \vphantom{\begin{array}{l} \text{Indépendance} \\ \text{totale} \\ p_{ijk} = p_i p_j p_k \end{array}} \right\} \Rightarrow \left\{ \begin{array}{ll}
 \text{Indép. IJ} & p_{ij} = p_i p_j \\
 \dots IK & p_{ik} = p_i p_k \\
 \dots JK & p_{jk} = p_j p_k
 \end{array} \right.$$

Mais non la réciproque il y a donc lieu d'introduire une hypothèse de non interaction et de rechercher sa formulation précise en fonction la densité  $p_{ijk}$  pour que l'implication ci-dessus puisse être à double sens.

On dispose d'une certaine liberté dans le regroupement des termes de la décomposition, en effet les sous-hypothèses précédentes étaient plus fines que nous pouvions énoncer à partir des projections en sous-espaces orthogonaux. Si l'on regroupe le  $\chi^2$  d'interaction avec un des  $\chi^2$  relatif à une face, il faudra rechercher l'hypothèse complémentaire des deux indépendances I x K et J x K par exemple. L'hypothèse minimale qui conduit à l'indépendance totale est celle qui suppose que la dépendance qui doit exister en I et J (pour ne pas retomber dans le cas précédent) est due seulement au caractère K, c'est-à-dire qu'il y a indépendance conditionnelle relativement à K entre I et J.

$$\begin{array}{rcl}
 \chi^2_{IJK} & = & \chi^2_{IK} + \chi^2_{JK} + \chi^2_{(IJ/K)} \\
 \text{Indép.} & = & \text{Indép.} \quad \text{Indép.} \quad \text{Indép.} \\
 \text{totale} & & I \times K \quad \cup \quad J \times K \quad \cup \quad I \times J \text{ si } K
 \end{array}$$

Si l'on essaie de regrouper les deux premiers termes du membre de droite on ne peut pas aboutir à une hypothèse concernant  $p_{ijk}$  puisqu'on dispose que de :

$$p_{ik} = p_i p_k \quad \text{et} \quad p_{jk} = p_j p_k$$

Pour obtenir une nouvelle décomposition de l'hypothèse globale on doit regrouper l'interaction avec deux hypothèses relatives aux faces, ce qui revient au même une hypothèse d'indépendance conditionnelle et une hypothèse d'indépendance entre deux caractères alors :

$$\chi^2_{IJK} = \chi^2_{Ij} + \chi^2_{(I \times k)}$$

L'hypothèse complémentaire a déjà été rencontrée dans les correspondances ternaires, c'est l'indépendance entre un caractère et l'association des deux autres :

$$p_{ijk} = p_{ij} p_k$$

En effet on a l'implication double

$$H : p_{ijk} = p_i p_j p_k \quad \Leftrightarrow \quad \left\{ \begin{array}{l} H_1 : p_{ij} = p_i p_j \\ \text{et} \\ H_2 : p_{ijk} = p_{ij} p_k \end{array} \right\} H_1 \cup H_2$$

Cette sous-hypothèse était testée à l'aide de la quantité

$$\chi^2 = N \sum_{ijk} \left( \frac{f_{ijk}}{f_{i.} f_{.j}} - 1 \right)^2 f_{i.} f_{.j} = N \sum_{ijk} \frac{(f_{ijk} - f_{i.} f_{.j})^2}{f_{i.} f_{.j}}$$

Alors que la quantité  $\chi^2_{ijk} / N$  représente la norme de la fonction :  $\left( \frac{f_{ijk}}{f_i f_j f_k} - 1 \right) - \left( \frac{f_{ij}}{f_i f_j} - 1 \right)$   $\chi^2_{ijk} = N \sum_{ijk} \frac{(f_{ijk} - f_i f_j f_k)}{f_i f_j f_k}$

Les deux statistiques diffèrent légèrement, dans le second cas on utilise en effet la pondération associée à l'hypothèse d'indépendance globale et non à la sous-hypothèse.

La même remarque s'applique aux cas précédents. Il y a donc lieu de faire certaines réserves sur l'utilisation des variances partielles pour les tests de sous-hypothèses.

Loi des variances. Utilisation pratique des résultats.

L'analyse de la variance d'une table tridimensionnelle a fait apparaître des variances partielles de loi connue dans une certaine hypothèse probabiliste. Il serait intéressant de déterminer la loi de la variance liée à l'interaction, et les lois des regroupements de variance que nous avons attribués à des sous-hypothèses particulières. Pour ceci il faut s'intéresser à la distribution du vecteur aléatoire dont on calcule la norme et aussi à l'indépendance statistique des termes de la décomposition.

Cette question se résoud en reprenant la démonstration de H. Cramer sur les  $\chi^2$  d'adéquation. On dispose de T classements possibles chacun ayant la probabilité  $\{p_t, t \in T\}$  d'être observé.

Un échantillon nous fournit pour chaque case  $t$  un nombre de réalisations  $n_t$ . L'ensemble de ces nombres suit une loi multinomiale.

On construit à partir de l'hypothèse probabiliste  $\{p_t \mid t \in T\}$  et de l'observation une statistique numérique :

$$\chi^2 = \sum_t \frac{(n_t - np_t)^2}{np_t} \quad n = \sum_t n_t$$

qui sera utilisé pour juger de l'adéquation de l'hypothèse et de l'observation.

En effet cette quantité suit une loi de  $\chi^2$  à Card  $T-1$  degrés de liberté lorsque les probabilités des classements sont bien les  $p_t$ . Pour démontrer cette propriété on considère le vecteur aléatoire  $\{x_t \mid t \in T\}$

$$x_t = \frac{n_t - np_t}{\sqrt{np_t}}$$

tel que 
$$\chi^2 = \sum_t x_t^2$$

dont la seconde fonction caractéristique tend pour  $n \rightarrow \infty$  vers celle d'un vecteur normal multidimensionnel de moyenne nulle et de matrice de variance-covariance

$$I - pp^T \quad \text{avec} \quad p = (\sqrt{p_1}, \dots, \sqrt{p_t})$$

Autrement dit dans toute direction orthogonale à ce vecteur  $p$  la variance est égale à 1. La dimension de ce sous-espace étant (card  $T-1$ ) il est possible d'effectuer un changement de base tel que le  $\chi^2$  apparaisse comme le carré de Card  $T-1$  variables normales centrées et réduites.

Considérons à présent  $T = I \times J \times K$ , alors dans l'hyperplan orthogonal à  $p = (\sqrt{p_{ijk}} \mid ijk \in T)$  il est possible de choisir une base particulière mettant en évidence les sous-espaces orthogonaux que nous avons déjà rencontrés.

Ici l'orthogonalité (ou non corrélation) est définie par une matrice unité, pour retrouver notre métrique  $p_i \times p_j \times p_k$  ou dans  $R^{IJK}$  il suffit d'effectuer une affinité selon les axes de coordonnées de  $R^{IJK}$ .

$$x_t = \frac{n_t - n p_t}{\sqrt{n p_t}} \quad \text{devient} \quad \frac{n_t - n p_t}{\sqrt{n} p_t}$$

$$P = (\sqrt{p_{ijk}} \mid ijk \in T) \quad \text{devient} \quad P = (1 \mid \forall ijk)$$

L'hyperplan orthogonal à  $p$  peut se décomposer dans une base ainsi constituée :

$$e_p = \left\{ e_{ijk} = \phi_p^i \mid p \in 1, \text{card } I - 1 \right\}$$

Vecteur dépendant des seuls indices  $i$ , orthogonaux à  $p$  donc de moyenne nulle relativement à la pondération  $p_{ijk}$  ou ce qui est équivalent :  $p_i$

De même pour  $J$  et  $K$  :

$$\left\{ e_m \mid m \in (1, \text{card } J - 1) \right\}$$

$$\left\{ e_n \mid n \in (1, \text{card } K - 1) \right\}$$

Ensuite on pourra prendre des vecteurs ne dépendant que de deux indices, c'est-à-dire constants sur les coordonnées de  $R^{IJK}$  ne différant par exemple que de l'indice k. Ces vecteurs sont orthogonaux à  $\{e_n\}$  les précédents vecteurs de base de moyenne nulle sur  $p_k$ . Ils doivent aussi être orthogonaux à  $p, e_e, e_m$ . D'un espace à Card I card J dimensions on passe à :

$$\begin{aligned} \text{Card I} \times \text{card J} - (\text{card I} - 1) - (\text{card J} - 1) - 1 &= \\ \text{card I} \times \text{card J} - \text{card I} - \text{card J} + 1 &= \\ (\text{card I} - 1) (\text{card J} - 1) \text{ dimensions} & \end{aligned}$$

On aurait des résultats analogues pour les vecteurs constants sur I et J.

Le résidu se trouve dans un espace à  $(\text{card I} \times \text{card J} \times \text{card K} - (\text{card I} - 1) - (\text{card J} - 1) - (\text{card K} - 1) - (\text{card I} - 1) \times (\text{card J} - 1) - (\text{card I} - 1) \times (\text{card K} - 1) - (\text{card J} - 1) \times (\text{card K} - 1) = (\text{card I} - 1) (\text{card J} - 1) (\text{card K} - 1)$

Si  $e_\alpha$  représente un vecteur normé de cette base la variable aléatoire  $\langle X, e_\alpha \rangle$  coordonnée du vecteur normal de  $R^{\text{card T}-1}$  sur le vecteur  $e_\alpha$  suit une loi normale centrée et réduite. De plus la corrélation entre deux coordonnées différentes  $\langle X, e_\alpha \rangle$  et  $\langle X, e_\beta \rangle$  est nulle. Dans cette base,  $X$  a une matrice de variance covariance unité, et les variables calculées dans les sous-espaces orthogonaux sont indépendantes en probabilité. C'est vrai en particulier pour la somme des carrés des coordonnées qui suivent des lois  $\chi^2$  (les degrés de liberté sont égaux au nombre de coordonnées intervenant dans la somme).

$$\sum_{ijk} \left( \frac{n_{ijk}}{n p_i p_j p_k} - 1 \right)^2 p_i p_j p_k = \sum_i \left( \frac{n_i}{n p_i} - 1 \right)^2 p_i + \sum_j \left( \frac{n_j}{n p_j} - 1 \right)^2 p_j$$

$$\chi^2_{\text{Card I} \times \text{Card J} \times \text{Card K} - 1} = \chi^2_{\text{Card I} - 1} + \chi^2_{\text{Card J} - 1}$$

$$+ \sum_k \left( \frac{n_k}{n p_k} - 1 \right)^2 p_k + \sum_{ij} \frac{(n_{ij} - n_i p_j - n_j p_i + n p_i p_j)^2}{n p_i p_j} + \text{etc} \dots$$

$$\chi^2_{\text{Card K} - 1} + \chi^2_{(\text{Card I} - 1)(\text{Card J} - 1)} + \dots$$

La décomposition fait bien apparaitre les  $\chi^2$  d'adéquation sur une marge mais les  $\chi^2$  d'indépendances n'ont pas la forme habituelle. Le modèle additif que nous avons choisi ne s'adapte pas parfaitement à la forme multiplicative des hypothèses habituelles.

Lorsque les marges  $p_i p_j$  et  $p_k$  intervenant dans l'hypothèse d'indépendance totale sont estimées on peut obtenir les mêmes propriétés de loi et d'indépendance entre les termes de la décomposition de la variance (cf. LANCASTER, 1ère référence, chapitre 12). Mais toutes les variancées calculées suivent des loi de  $\chi^2$  indépendantes avec pour nombre de degrés de liberté la dimension du sous-espace de projection associé, uniquement dans l'hypothèse d'indépendance totale. C'est ainsi que RL Plackett a pu constater la procédure de test d'interaction qui consiste à considérer la variance d'interaction (différence entre  $\chi^2_{ijk}$  et  $\chi^2_{ij} + \chi^2_{jk} + \chi^2_{ik}$ ) comme une statistique permettant de tester la sous-hypothèse de non-interaction.



Les variances calculées dans ce modèle ne sont que des valeurs indicatrices de la nature de l'écart à l'indépendance. Dans la pratique ces calculs très simples à effectuer, permettront de choisir la contingence entre deux caractères qu'il serait intéressant de décomposer. On n'essiera d'interpréter que les associations entre caractères qui se seront manifestées par un  $\chi^2$  significatif.

La variance de l'interaction pourra être utilisée pour choisir une sous-hypothèse, d'indépendance entre un caractère et les deux autres ou d'indépendance conditionnelle, qui méritera d'être testée à part, avec des moyens plus appropriés.

Ainsi cette décomposition de variance permettra de simplifier et de mieux appréhender la structure interne du tableau à trois dimensions.

#### Compléments à l'étude des sous-hypothèses

Notre étude d'une table de contingence entre trois caractères au voisinage de l'hypothèse d'indépendance (c'est-à-dire en utilisant une métrique très particulière) à fait ressortir par l'intermédiaire d'une décomposition d'une fonction d'écart ou d'une variance totale des propriétés intrinsèques du tableau étudié et de l'hypothèse utilisée. Ainsi il s'est avéré possible d'approcher la répartition interne du tableau par la répartition observée sur ses faces et de scinder l'écart à une hypothèse très forte en une somme d'écarts à des hypothèses moins contraignantes.

Nous retrouvons ainsi une façon d'aborder un problème statistique courante dans la théorie des tests. Nous ne pouvons toutefois l'exploiter pleinement à cause de l'hypothèse très forte utilisée pour la décomposition et qui est la seule à donner une métrique ayant toutes les propriétés désirées.

Le choix d'un modèle explicatif avec effets et interactions additifs ne nous a pas permis de retrouver dans les décompositions de variances les grandeurs habituellement affectées aux sous-hypothèses ou aux sous-ensembles du tableau.

La forme multiplicative des modèles probabilistes et les décompositions du type de l'analyse de la variance ne sont pas parfaitement compatibles et il serait intéressant de remplacer les variances par de nouveaux instruments de mesures adaptés aux modèles multiplicatifs comme les entropies ou les quantités d'information.

Le complément naturel de cet étude est donc constitué par deux approches différentes des problèmes dans les tables de contingence. Celle de M. SN ROY, qui sous l'angle de la théorie des tests traite toutes les sous-hypothèses que nous avons rencontrées, et celle de M. S KULLBACK qui applique les résultats de la théorie de l'information à ce genre de données statistiques, et qui définit de nouvelles structures pour les tableaux tridimensionnels.

La méthode de la "décomposition des écarts" et ces deux approches ont des objectifs très différents mais elles peuvent se compléter et s'enrichir mutuellement. Ceci sera particulièrement vrai pour l'étude de l'indépendance conditionnelle, notion très intéressante pour des contingences entre variables qualitativement différentes et qui permet d'approfondir la notion de facteur explicatif dans une analyse factorielle.

### III - LES TESTS DE $\chi^2$ sur les tables de contingence

#### III1 Présentation

Les résultats énoncés dans ce chapitre proviennent essentiellement de deux articles de M. S.N. ROY :

- Some non-parametric generalizations of analysis of variance and multivariate analysis.

- On the hypothesis of no "interaction" in a multiway contingency table.

Qui traitent de façon très complète des tables de contingence résultant de comptages ou construites dans le but de vérifier une hypothèse. Nous ne présenterons pas les résultats afférents à ce second type de table qui concernent plutôt les techniques de plan d'expérience.

Pour les tables résultant d'un modèle probabiliste et de tirage aléatoires indépendants, nous résumerons les différentes étapes théoriques qui conduisent à faire un test du  $\chi^2$  et nous présenterons tous les résultats pratiques qui en découlent.

#### III2 Approche et Résultats théoriques

Les notations utilisées seront les mêmes que celles du premier chapitre,  $n$  représentant un nombre,  $p$  une probabilité et les indices déterminent une case dans une table à deux ou trois entrées.

Les fonctions de vraisemblance sont définies à partir d'un échantillon observé, par exemple :  $\{n_{ijk} | i \in I, j \in J, k \in K\}$  et d'une densité de probabilité  $p_{ijk}$  qui peut être soumise à de nombreuses conditions selon l'hypothèse que l'on désire tester. On aura généralement deux fonctions de vraisemblance.

$$\phi = \frac{n!}{\prod_{ijk} n_{ijk}!} \prod_{ijk} p_{ijk}^{n_{ijk}}$$

$$\phi_0 = \frac{n!}{\prod_{ijk} n_{ijk}!} \prod_{ijk} p_{0ijk}^{n_{ijk}}$$

Ces quantités sont analogues à des probabilités à priori des échantillons observés pour des lois multinomiales particulières. La première fonction correspond à une densité quelconque (une seule condition  $\sum p_{ijk} = 1$  qui représentera la contre hypothèse la plus générale. L'hypothèse, que l'on soumettra au test, définit sur notre tableau la densité  $p_0$  qui pourra avoir de nombreuses formes dans un tableau à trois dimensions où les sous-hypothèses de l'hypothèse d'indépendance totale sont multiples.

Il faut remarquer que si nous avons considéré un des indices non pas comme une simple modalité aléatoire d'un caractère mais comme un mode de classement des réalisations, c'est-à-dire si nous avons fixé la marge (nombre total d'observations) correspondant à cet indice, la forme de la fonction de vraisemblance aurait été profondément modifiée. En effet dans les techniques d'expérimentation les caractères aléatoires des observations sont bien différents puisque par exemple il n'y a pas nécessairement indépendance entre deux tirages successifs, les groupes de classement peuvent être connus à l'avance.

Lorsque l'on dispose de deux fonctions de vraisemblance correspondant aux densités  $p_0$  et  $p_1$  il est possible de définir un test le plus puissant à risque de première espèce fixé. L'application du lemme fondamental de Neyman et Pearson nous définit une région de refus de notre hypothèse  $p_0$ . En fait le lemme nous permet de construire une statistique (grandeur numérique calculée à partir de l'échantillon) jouissant de propriétés optimales au sens où si l'on admet que l'on peut rejeter  $\alpha$  fois sur cent l'hypothèse  $H_0$  lorsqu'elle est vraie alors la probabilité d'acceptation  $\beta/100$  de  $H_0$  lorsqu'elle est fautive est minimale.

Le lemme indique que la région de refus de  $H_0$  est telle que :

$$\frac{\phi_1(n)}{\phi_0(n)} = \frac{P_r(n/p_1)}{P_r(n/p_0)} \geq C \quad ; \quad n \text{ symbolisant l'échantillon observé}$$

La constante  $c$  étant définie par le risque de première espèce et l'effectif total de l'échantillon.

Nous avons ici une inégalité entre une fonction de l'échantillon de deux densités et une constante ceci définit un intervalle numérique pour une statistique, le problème qui se pose est de passer d'une contre hypothèse  $p_1$  à une contre hypothèse  $p \neq p_0$ . Pour conserver un risque de première espèce de  $\alpha$  pour cent et des conditions d'optimalité quant au risque de seconde espèce, il est possible de prendre pour domaine de refus l'union de domaines définis par  $p_1 \neq p_0$  lorsque  $p_1$  varie.

Dans le cas particulier de nos fonctions de vraisemblance on travaille bien sûr sur  $-\log(\phi_1/\phi_0)$ , le domaine de rejet est de la forme  $f(n, p_0, p_1) \geq C$ .

L'opération d'union des intervalles de refus amène à écrire :

$$\sup_{p \neq p_0} f(n, p_0, p) \geq C \Rightarrow \text{refus de } H_0$$

c'est-à-dire que l'on acceptera  $H_0$  et donc la densité  $p_0$  que si toutes les contre-hypothèses possibles vérifient les conditions imposées à  $p_1$  :

$$\begin{aligned} &\text{Acceptation pour } p_1 \text{ si } f(n, p_0, p_1) \leq C \\ &\text{de } H_0 \text{ pour } p_0 \text{ si } f(n, p_0, p_1) \leq C \quad \forall p \neq p_0 \end{aligned}$$

d'où en particulier :  $\sup f(n, p_0, p_1) \leq C$

Cette écriture sous forme d'un extremum a l'avantage de faire disparaître dans la statistique la densité  $p$  par rapport à laquelle on effectue l'opération de maximisation. Pour un tableau unidimensionnel on obtient un extremum de la forme  $\sum_{i \in I} (n_i - np_{0i})^2 / np_{0i}$ , dont le carré est distribué selon un  $\chi^2$  à  $(\text{Card } I - 1)$  degrés de liberté dans l'hypothèse probabiliste correspondant à  $H_0$ . Connaissant la loi de la statistique calculée, on définit immédiatement un intervalle à  $\alpha$  pour 100 d'une forme déterminée.

Ces considérations permettent de retrouver le célèbre test d'adéquation d'un échantillon à une loi de probabilité.

Des cas très fréquents de tests d'hypothèse portent sur des densités de probabilité dépendant d'un certain nombre de paramètres. On a alors affaire aux deux hypothèses suivantes :

$$H_0 : \left\{ p_i = p_i(\theta_1, \dots, \theta_r) \right\}_{(\theta_1, \dots, \theta_r) \in \Omega}$$

$$H_1 \neq H_0$$

$$r = \text{card } R < \text{card } I$$

pour tout ensemble  $(\theta_1^*, \dots, \theta_r^*)$  on peut appliquer la procédure précédente, en remarquant que rejeter  $p_i(\theta_1, \dots, \theta_r)$  revient à rejeter  $p_i(\theta_1^*, \dots, \theta_r^*)$  pour tout  $(\theta_1^*, \dots, \theta_r^*) \in \Omega$  il est plausible de construire un intervalle de refus qui sera l'intersection des intervalles définis par un ensemble de paramètres déterminés.

Nous avons

$$\sum_{i \in I} (n_i - n p_i(\theta^0))^2 / n p_i(\theta^0) > c^2$$

l'intersection de ces régions revient à écrire

$$\inf_{\theta^0} \sum_i (n_i - n p_i(\theta^0))^2 / n p_i(\theta^0) > c^2$$

Le rejet de  $H_0$  aura bien lieu contre  $H_1 \neq H_0$  que s'il a lieu quelque soit l'ensemble de paramètres  $\theta^0$  permettant de définir  $H_0$ .

La justification de ces principes d'union et d'intersection de régions de refus selon que la contre-hypothèse ou l'hypothèse est variable se trouve dans le livre de S. N. ROY "Some Aspects of Multivariate Analysis".

La détermination des paramètres permettant d'effectuer le test se fait donc par minimisation d'un  $\chi^2$  d'adéquation, mais M. H Cramer a montré que pour des échantillons importants les paramètres qui maximisent la fonction de vraisemblance de l'échantillon sous l'hypothèse  $H_0$  ont les mêmes propriétés optimales.

Avec la fonction :

$$\phi_0 = \frac{n!}{\prod_i n_i!} \prod_i p_i^{n_i}(\theta_1, \dots, \theta_r) \quad \text{par exemple}$$

On utilisera donc plutôt les conditions

$$0 = \frac{\partial}{\partial \theta_j} \text{Log } \phi_0 = \sum_{i \in I} \frac{n_i}{n p_i} \frac{\partial p_i}{\partial \theta_j} = \sum_{i \in I} \frac{n_i - n p_i}{n p_i} \frac{\partial p_i}{\partial \theta_j} \quad j \in R$$

Le théorème d'adéquation d'un échantillon à une loi peut donc se généraliser au cas où l'on estime les paramètres définissant la loi de référence.

Avec une fonction de vraisemblance définie comme ci-dessus et Card I probabilités, fonctions de card R paramètres, telles que pour tout point d'un intervalle A non-dégénéré de l'espace des paramètres on ait :

$$\sum_{i \in I} p_i = 1 \quad \text{et} \quad p_i(\theta_1, \dots, \theta_r) > 0 \quad \forall i \in I$$

$$\frac{\partial p_i}{\partial \theta_j} \quad \text{et} \quad \frac{\partial p_i}{\partial \theta_j \partial \theta_k} \quad \text{continues} \quad , \quad \text{Rang} \left( \frac{\partial p_i}{\partial \theta_j} \right) = \text{Card } R$$



Alors si les paramètres réels de la loi observée appartiennent l'intérieur de l'intervalle A alors il n'existe qu'une solution aux équations du maximum de vraisemblance, cette solution converge en probabilité pour  $n \rightarrow \infty$  vers les valeurs réelles du paramètre et la statistique d'adéquation construite avec cette estimation tend vers un  $\chi^2$  à card I - 1 - card degrés de liberté dans l'hypothèse  $H_0$  qui définit la fonction de vraisemblance servant aux calculs.

Le complément naturel à ces deux théorèmes d'adéquation est l'étude du cas où les probabilités ou paramètres qui s'introduisent dans les modèles sont liés par des conditions supplémentaires.

Pour le premier théorème on peut imposer des conditions du type suivant :

$$f_j(p_1, \dots, p_{\text{card I}}) = 0 \quad j = 1, \text{card S}, j \in S$$

l'une d'entre elles étant  $\sum_i p_i = 1$  pour permettre de calculer les valeurs  $p_i$  plutôt que de les imposer dans l'hypothèse  $H_0$ .

On suppose  $\frac{\partial f_j}{\partial p_i}$  et  $\frac{\partial^2 f_j}{\partial p_i \partial p_j}$  continues,  $\text{rang} \left( \frac{\partial f_j}{\partial p_i} \right) = \text{Card S}$

Les probabilités sont calculées à partir des équations suivantes :

$$\begin{cases} \frac{n_i - n p_i}{n p_i} + \sum_{j \in S} \mu_j \frac{\partial f_j}{\partial p_i} = 0 & i \in I \\ f_j(p_1, \dots, p_{\text{card I}}) = 0 \end{cases}$$

qui correspondent au maximum de vraisemblance sous contraintes.

Si la répartition  $p_0$  est un point intérieur du domaine de variation des probabilités  $\{p_i | i \in I\}$  alors il n'existe qu'un estimateur qui converge en probabilité vers  $p_0$ , et la statistique d'adéquation construite avec cet estimateur  $p^1, \sum_i (n_i - n p_i)^2 / n p_i$  tend vers un  $\chi^2$  à  $(\text{Card } I - \text{Card } S)$  degrés de liberté. ( $\text{Card } I$  cases,  $(\text{Card } I - \text{Card } S)p_i$  indépendants, i.e.  $\text{Card } I$  valeurs et  $\text{Card } S$  contraintes).

Pour le second théorème on impose des conditions sur les paramètres.

$$f_k(\theta_1, \dots, \theta_{\text{Card } R}) = f_k^0 \quad k = 1, \dots, \text{Card } T$$

et  $\frac{\partial f_k}{\partial \theta_j}, \frac{\partial f_k}{\partial \theta_j \partial \theta_l}$  continues,  $\text{rang} \left( \frac{\partial f_k}{\partial \theta_j} \right) = \text{Card } T$

Les équations du maximum de vraisemblance

$$\left\{ \begin{array}{l} \sum_{i \in I} \frac{n_i - n p_i(\theta_1, \dots, \theta_r)}{n p_i(\theta_1, \dots, \theta_r)} + \sum_{k \in T} \lambda_k \frac{\partial f_k}{\partial \theta_j} = 0 \quad j \in R \\ f_k(\theta_1, \dots, \theta_r) = f_k^0 \quad k \in T \end{array} \right.$$

fournissent un seul estimateur  $\hat{\theta}^1$  du vrai vecteur de paramètres  $\theta^0$  si celui-ci appartient à l'ensemble de variation de  $\theta$  et la statistique

$$\sum_{i \in I} \frac{(n_i - n p_i(\hat{\theta}))^2}{n p_i(\hat{\theta})} \quad \text{tend vers un } \chi^2 \text{ à :}$$

$(\text{Card } I - 1) - (\text{Card } R - \text{Card } T)$  degrés de liberté.

La conclusion générale de ces théorèmes d'adéquation avec ou sans paramètre, avec ou sans relation entre les probabilités ou les paramètres s'énonce de la façon suivante :

$$\sum_{i=1}^{\text{Card I}} \frac{(n_i - n p_i(\theta_1, \dots, \theta_i, \dots, \theta_{\text{Card R}}))^2}{n p_i(\theta_1, \dots, \theta_{\text{Card R}})} \rightarrow \chi_p^2 \quad j \in R$$

avec  $p = (\text{card I} - \text{card T}) - (\text{Card R} - \text{card T}) - (\text{Card I} - \text{card S})$

Card S étant le nombre de relations indépendantes et régulières définissant les  $p_i$ , et card T celui des relations entre les  $\theta_j$

### III3 Application aux tables de contingence

Pour une table entre deux caractères la fonction de vraisemblance générale est :

$$\phi = \frac{n!}{\prod_{ij} n_{ij}!} \prod_{ij} p_{ij}^{n_{ij}}$$

et l'hypothèse composite qu'il est intéressant de tester est l'indépendance entre les deux caractères c'est à dire :  $H_0 p_{ij} = p_i p_j$  avec  $p_i = \sum_j p_{ij}$   
 $p_j = \sum_i p_{ij}$  contre  $H \neq H_0$ . Cette hypothèse correspond à l'hypothèse de non corrélation dans une population normale à deux variables.

Dans l'hypothèse  $H_0$  la vraisemblance s'écrit :

$$\phi_0 = \frac{n!}{\prod_{ij} n_{ij}!} \prod_{ij} (p_i p_j)^{n_{ij}} = \frac{n!}{\prod_{ij} n_{ij}!} \prod_i p_i^{n_i} \prod_j p_j^{n_j}$$

$$n_i = \sum_j n_{ij} \quad , \quad n_j = \sum_i n_{ij}$$

L'estimation des  $p_i$  et  $p_j$  qui jouent le rôle de paramètre se fait en maximisant  $\log \phi_0$  en introduisant les contraintes  $\sum_i p_i = 1$ ,  $\sum_j p_j = 1$  à l'aide de multiplicateurs de Lagrange.

Les solutions sont  $p_i^1 = \frac{n_i}{n}$   $p_j^1 = \frac{n_j}{n}$  qui ne sont autres que les fréquences observées pour les modalités des deux caractères.

La statistique qui tend vers un  $\chi^2$  est donc

$$\sum_{i,j} \frac{(n_{ij} - n p_i^1 p_j^1)^2}{n p_i^1 p_j^1} = \sum_{i,j} \frac{(n_{ij} - n_i n_j / n)^2}{n_i n_j / n}$$

On retrouve le  $\chi^2$  de contingence. Pour son nombre de degrés de liberté on utilise les théorème généraux. Il y a card I card J probabilités  $p_{ij}$  liées par une relation  $\sum_{i,j} p_{ij} = 1$  ; on a introduit card I + card J paramètres liés par deux relations  $\sum_j p_i = 1$   $\sum_i p_j = 1$  d'où

$$\begin{aligned} p &= (\text{Card I} \times \text{Card J} - 1) - (\text{Card I} + \text{card J} - 2) \\ &= \text{Card I} \times \text{Card J} - \text{Card I} - \text{card J} + 1 \\ &= (\text{Card I} - 1)(\text{Card J} + 1) \end{aligned}$$

Dans une table à trois dimensions il est possible d'étudier plusieurs hypothèses qui sont déjà apparues dans les approches précédentes.

L'indépendance totale entre les trois caractères ;

$$H_0 p_{ijk} = p_i p_j p_k \text{ contre } H \neq H_0$$

$$\text{Alors } \phi_0 = \frac{n!}{\prod n_{ijk}!} \times \prod_{ijk} (p_i p_j p_k)^{n_{ijk}}$$

L'estimation des paramètres est très semblable à la précédent et l'on trouve  $\hat{p}_i = \frac{n_i}{n}$  ;  $\hat{p}_j = \frac{n_j}{n}$  ;  $\hat{p}_k = \frac{n_k}{n}$  les fréquences observé des modalités i, j et k.

Les paramètres représentant des probabilités associées aux modalités on a les conditions suivantes:

$$\sum_i p_i = \sum_j p_j = \sum_k p_k = 1$$

Cette hypothèse de totale indépendance correspondant pour une population normale multidimensionnelle à l'hypothèse donnant une matrice de variance covariance diagonale.

Le  $\chi^2$  de contingence s'écrit :

$$\sum_{ijk} \left( n_{ijk} - \frac{n_i \cdot n_j \cdot n_k}{n^2} \right)^2 / \frac{n_i \cdot n_j \cdot n_k}{n^2}$$

son nombre de degrés de liberté se calcule comme suit : card I \* card J \* card K - (card I + card J + card K) paramêtres  $p_i$   $p_j$  et  $p_k$  cases dans le tableau ou  $p_{ijk}$

- 1 relations  $\sum_{ijk} p_{ijk} = 1$

(card I + card J + card K) paramêtres  $p_i$   $p_j$  et  $p_k$

- 3 relations entre les paramètres d'où

$$l = \text{Card I} \times \text{Card J} \times \text{Card K} - \text{Card I} - \text{Card J} - \text{card K} + 2$$

Il est intéressant de décomposer cette hypothèse globale en plusieurs sous-hypothèses additives par exemple  $H_{01} p_{ik} = p_i p_k$  indépendance dans un tableau marginal (analogue à la corrélation totale nulle entre deux variables normales) qui se teste comme l'indépendance d'un tableau à deux dimensions.

$H_{02} p_{jk} = p_j p_k$  même type d'hypothèse

$H_{03} p_{ijk} = \frac{p_{ik} p_{jk}}{p_k}$  l'hypothèse complémentaire qui est

nécessaire pour retrouver l'indépendance totale elle peut se réécrire :

$$\frac{p_{ijk}}{p_k} = \frac{p_{ik}}{p_k} \times \frac{p_{jk}}{p_k} =$$

Proba (ij/k) = Proba (i/k) x Proba (j/k) il s'agit donc d'une indépendance conditionnelle entre les deux variables i et j que l'on sait déjà indépendantes de la variable k. Ceci correspond à la corrélation partielle entre deux variables normales x et y si z, dans une densité à trois dimensions.

La vraisemblance de l'échantillon associée à cette hypothèse est équivalente à :

$$\prod_{ijk} (p_{ik} p_{jk} / p_k)^{n_{ijk}}$$

De part la nature des paramètres il y a lieu d'introduire les relations suivantes :

$$\sum_i p_{ik} = \sum_j p_{jk} = p_k \quad \forall k \in K$$

$$\sum_k p_k = 1$$

La solution en  $p_{ik}$   $p_{jk}$   $p_k$  du maximum de vraisemblance donne encore les fréquences observées :

$$\hat{p}_{ik} = \frac{n_{ik}}{n} ; \hat{p}_{jk} = \frac{n_{jk}}{n} ; \hat{p}_k = \frac{n_k}{n}$$

La statistique à comparer à un seuil calculé sur une loi de  $\chi$  est la suivante :

$$\sum_{ijk} \left( n_{ijk} - \frac{n_{ik} n_{jk}}{n_k} \right)^2 / \frac{n_{ik} n_{jk}}{n_k}$$

Son nombre de degrés de liberté est :

$$f = \text{Card K} (\text{Card I} - 1) (\text{Card J} - 1)$$

Il y a en effet :

Card I card J card K valeur  $p_{ijk}$  liées par une relation et  $(\text{Card I} \text{ card K} + \text{card J} \text{ card K} + \text{card K})$  paramètres liés par 2 card K relations pour les  $p_{ik}$  et  $p_{jk}$  et 1 relation pour les  $p_k$  d'où

$$(\text{Card I} \text{ card J} \text{ card K} - 1) - (\text{Card I} \text{ card K}) - (\text{Card J} \text{ card K}) - \text{card K} + 2 \text{ card K} + 1 = \text{Card K} (\text{Card I} \text{ card J} - \text{Card I} - \text{card J} + 1)$$

L'hypothèse d'indépendance totale peut se décomposer d'une seconde façon :

$$H_0 p_{ijk} = p_i p_j p_k \quad H_{\alpha} p_{ijk} = p_{ij} p_k \quad \text{et} \quad p_{ij} = p_i p_j \quad H_{ob}$$

$H_{ob}$  a déjà été étudiée,  $H_{o\alpha}$  correspond à l'absence de corrélation multiple entre le couple (x,y) de deux variables normales par rapport à une troisième z.

On peut remarquer que pratiquement si l'on considère que le couple (i,j) indice un ensemble d'effectif total Card I x Card J on est ramené à un problème bidimensionnel et en effet le résultat final est :

$$\chi^2 = \sum_{ijk} \left( n_{ijk} - \frac{n_{ij} n_k}{n} \right)^2 / \frac{n_{ij} n_k}{n} \quad \text{avec } (Card I \times Card J - 1) (Card K - 1)$$

degrés de liberté.

#### III4 La non interaction

La troisième façon de décomposer l'hypothèse d'indépendance totale est de rechercher l'hypothèse complémentaire des hypothèses d'indépendance entre deux caractères.

$$H_{o'} p_{ij} = p_i p_j$$

$$H_{o''} p_{ik} = p_i p_k$$

$$H_{o'''} p_{jk} = p_j p_k$$

$$H_{o''''} \text{ non-interaction entre } i, j \text{ et } k$$

Ou ce qui revient au même l'hypothèse qui permet de compléter l'implication suivante :

$$\left. \begin{array}{l} \text{Indépendance entre I et K, } p_{ik} = p_i p_k, H_{o'''} \\ \text{Indépendance entre J et K, } p_{jk} = p_j p_k, H_{o''''} \end{array} \right\} H_{o\alpha} p_{ijk} = p_{ij} p_k \quad \text{Indépendance entre (I,J) et K}$$

Non Interaction H



Il est à remarquer que  $H_{0\alpha}$  implique  $H_{0'}$  et  $H_{0''}$ , mais que la réciproque n'est pas vraie, alors que pour l'analogie sur les lois normales multidimensionnelles, une corrélation totale nulle entre x et z et entre y et z entraîne une corrélation multiple nulle entre (xy) et z.

L'hypothèse  $H$  de non interaction peut s'exprimer par un ensemble de conditions sur la densité de probabilité, une première solution qui semble naturelle est :

$$\left\{ p_{ijk} = \frac{p_{ij} p_{ik} p_{jk}}{p_i p_j p_k} \quad \begin{array}{l} i \in I \\ j \in J \\ k \in K \end{array} \right\} H$$

qui complète bien les hypothèses précédentes pour donner  $H_0 p_{ijk} = p_i p_j p_k$  dans le premier cas et  $H_{0\alpha} p_{ijk} = p_{ij} p_k$  dans le second.

Cette solution ne sera pas retenue à cause de sa lourdeur mathématique mais surtout parce qu'elle ne se contente pas de compléter les autres sous-hypothèses. En effet elle implique des conditions très nombreuses sur le tableau  $I \times J \times K$  par le fait qu'elle fait intervenir les densités à une et deux dimensions dans son expression :

$$H \Rightarrow p_{ij} = \sum_k p_{ijk} = \frac{p_{ij}}{p_i p_j} \sum_k \frac{p_{ik} p_{jk}}{p_k}$$

$$\text{Soit } \forall_{ij} p_i p_j = \sum_k \frac{p_{ik} p_{jk}}{p_k}$$

et les conditions analogues en permutant les indices

Une seconde solution consiste à conserver la même forme mais sans faire intervenir les densités marginales :

$$p_{ijk} = \frac{q_{ij} q_{jk} q_{ik}}{q_i q_j q_k}$$

Les quantités du second membre étant arbitraires, seule la forme multiplicative de l'expression de  $p_{ijk}$  est importante. Une autre expression équivalente serait  $p_{ijk} = q_{ij} q_{ik} q_{jk}$ , elle est aussi complète les sous-hypothèses pour donner le résultat recherché, de plus elle n'entraîne pas de contradiction non désirable sur les probabilités marginales.

Pour une telle densité les paramètres  $q$  ne sont pas estimés par la méthode du maximum de vraisemblance, ils sont éliminés des  $\text{Card I} \times \text{Card J} \times \text{Card K}$  équations pour donner un ensemble de  $(\text{Card I} - 1) (\text{Card J} - 1) (\text{Card K} - 1)$  relations indépendantes entre les  $p_{ijk}$ . Ce sont donc les probabilités  $p_{ijk}$  qui sont estimées par la méthode du maximum de vraisemblance sous contraintes. Nous savons d'après les théorèmes généraux que la statistique finale tendra vers un  $\chi^2$  à  $(\text{Card I} - 1) (\text{Card J} - 1) (\text{Card K} - 1)$  degrés de liberté, nombre égal à celui des contraintes imposées par notre hypothèse.

Pour un tableau  $2 \times 2 \times 2$  on obtient une condition :

$$(1) \frac{p_{111} p_{221}}{p_{211} p_{121}} = \frac{p_{112} p_{222}}{p_{212} p_{122}}$$

c'est l'hypothèse de non interaction de Bartlett. La maximisation porte sur

$$\Phi = \prod_{ijk} p_{ijk}^{n_{ijk}}$$

avec les contraintes  $p_{111} p_{221} p_{212} p_{122} - p_{211} p_{122} p_{112} p_{222} = 0$

$$\text{et } \sum_{ijk} p_{ijk} = 1$$

Ce qui donne les conditions :

$$\frac{n_{ijk}}{p_{ijk}} + \frac{\lambda}{p_{ijk}} + M = 0 \quad \text{pour } ijk = 111, 221, 212, 122$$

$$\frac{n_{ijk}}{p_{ijk}} - \frac{\lambda}{p_{ijk}} + M = 0 \quad \text{pour les autres groupes d'indices}$$

Finalement  $M = -n$  et  $\hat{p}_{ijk} = (n_{ijk} \pm \lambda)/n$  selon le groupe d'indices avec  $\lambda$  solution unique de :

$$\frac{(n_{111} + \lambda)(n_{221} + \lambda)}{(n_{211} - \lambda)(n_{121} - \lambda)} = \frac{(n_{112} - \lambda)(n_{222} - \lambda)}{(n_{212} - \lambda)(n_{122} + \lambda)}$$

La statistique numérique qui tend vers un  $\chi^2$  à  $(\text{Card I} - 1)(\text{Card J} - 1)(\text{Card K} - 1)$  est  $\chi^2 = \frac{\lambda^2}{n} \sum_{ijk} \frac{1}{\hat{p}_{ijk}}$

Pour une table de dimension  $\text{Card I} \times \text{Card J} \times \text{Card K}$  on obtient en menant un calcul analogue à celui du cas précédent les contraintes suivantes :

$$\frac{p_{rst} p_{ijt}}{p_{ist} p_{rjt}} = \frac{p_{rsk} p_{ijk}}{p_{rsk} p_{rjk}} \quad \text{pour} \quad \begin{cases} r = \text{card I} \\ s = \text{card J} \\ t = \text{card K} \\ i \in (1, r-1) \\ j \in (1, s-1) \\ k \in (1, t-1) \end{cases}$$

La maximisation utilise  $(r-1)(s-1)(t-1)$  multiplicateurs pour ces contraintes et un pour  $\sum_{ijk} p_{ijk} = 1$ .

Comme précédemment le  $\chi^2$  s'exprime en fonctions des multiplicateurs, et plus simplement en fonction des solutions des  $(r-1)(s-1)(t-1)$  équations suivantes :

$$\frac{(n_{srt} + M_{rst})(n_{ijt} + M_{ijt})}{(n_{ist} - M_{ist})(n_{rjt} - M_{rjt})} = \frac{(n_{rsk} - M_{rsk})(n_{ijk} - M_{ijk})}{(n_{isk} + M_{isk})(n_{rjk} - M_{rjk})}$$

$$\chi^2 = \sum_{ijk} M_{ijk}^2 / (n_{ijk} + \epsilon_{ijk} M_{ijk})$$

ou  $\epsilon_{ijk} = 1$  pour  $ijk = rst$

= -1 si un indice différé de  $rst$

= +1 si deux indices diffèrent de  $rst$

= -1 si tous les indices diffèrent de  $rst$

Le nombre de degrés de liberté est égal au nombre de contraintes sur les  $p_{ijk}$  provenant du modèle c'est-à-dire  $(\text{Card I} - 1) (\text{Card J} - 1) (\text{Card K} - 1)$

### III5 Remarque sur l'addition des hypothèses

Résumons les différentes décompositions de l'indépendance totale que nous avons rencontrées :

$$H_0, p_{ijk} = p_i p_j p_k \Rightarrow \chi^2 \text{ card I card J card K} - \text{card I} - \text{card J} - \text{card K} + 2$$

$$I \begin{cases} p_{ik} = p_i p_k & \Rightarrow \chi^2 \text{ card I card K} - \text{card I} - \text{card K} + 1 \\ p_{jk} = p_j p_k & \Rightarrow \chi^2 \text{ card J card K} - \text{card J} - \text{card K} + 1 \\ p_{ijk} = p_{ik} p_{jk} / p_k & \Rightarrow \chi^2 \text{ card K} (\text{card I card J} - \text{card I} - \text{card J} + 1) \end{cases}$$

La somme des degrés de liberté est

$$(\text{card I card J card K} - \text{card I} - \text{card J} - \text{card K} + 2)$$

identique à celle du  $\chi^2$  d'indépendance totale

$$II \begin{cases} p_{ij} = p_i p_j & \Rightarrow \chi^2 \text{ card I card J} - \text{card I} - \text{card J} + 1 \\ p_{ijk} = p_{ij} p_k & \Rightarrow \chi^2 \text{ card I card J card K} - \text{card I card J} - \text{card K} + 1 \end{cases}$$

On obtient toujours la même somme des degrés de liberté :

$$III \begin{cases} p_{ij} = p_i p_j & \Rightarrow \chi^2 (\text{card J} - 1) (\text{card I} - 1) \\ p_{ik} = p_i p_k & \Rightarrow \chi^2 (\text{card I} - 1) (\text{card K} - 1) \\ p_{jk} = p_j p_k & \Rightarrow \chi^2 (\text{card K} - 1) (\text{card J} - 1) \\ \text{non interaction} & \Rightarrow \chi^2 (\text{card I} - 1) (\text{card J} - 1) (\text{card K} - 1) \end{cases}$$

Même somme de degrés de liberté

On a aussi

$$(Card I - 1) (card J - 1) (card K - 1) + (card I - 1) (card K - 1) + (card J - 1) (card K - 1)$$

non interaction + Indép. I x K + indép I x K

=

$$(card I card J - 1) (card K - 1)$$

Indép (IJ) x k

Ces relations entre degrés de liberté ne sont pas sans justifications théoriques, M. Nitra a démontré dans l'article cité en référence qu sous l'hypothèse  $H_0$  d'indépendance totale toutes ces quantités numériques suivant des lois de  $\chi^2$  qui permettent de tester les sous-hypothèses, sont asymptotiquement indépendantes en probabilité et leur somme converge en probabilité, pour un effectif  $n$  tendant vers l'infini, vers la statistique permettant de tester l'hypothèse d'indépendance totale.

Ainsi, bien qu'il n'y ait pas additivité au sens algébrique du terme entre les expressions mathématiques proposées on peut tout de même décomposer une hypothèse et obtenir asymptotiquement les mêmes propriétés qu pour une analyse de la variance classique.

IV - THEORIE DE L'INFORMATION ET TABLES DE CONTINGENCE

IV1 - Définitions, propriétés et exemples

Considérons deux hypothèses probabilistes exclusives et complémentaires  $H_1$  et  $H_2$  et une variable aléatoire  $X$ , posons :

$P_1(x) = P(X = x | H_1)$  probabilité de  $(X = x)$  dans l'hypothèse  $H_1$

$P_2(x) = P(X = x | H_2)$  .....  $H_2$

Les hypothèses ont des probabilités à priori  $P(H_1)$  et  $P(H_2)$  résultant d'expériences antérieures, nous pouvons écrire :

$P(X = x \text{ et } H_1) = P(x) \times P(H_1/x) = P(H_1) \times P_1(x)$

$P(X = x \text{ et } H_2) = P(x) \times P(H_2/x) = P(H_2) \times P_2(x)$

avec  $P(x) = P_1(x) \times P(H_1) + P_2(x) \times P(H_2)$  probabilité moyenne d'observer la valeur  $x$ .

La formule de Bayes, dite formule des probabilités des causes se déduit de ces définitions.

$$P(H_1 | x) = \frac{P(H_1) \times P_1(x)}{P(H_1) \times P_1(x) + P(H_2) \times P_2(x)} = \frac{P(H_1) \times P_1(x)}{P(x)}$$

Elle permet de reconsidérer la probabilité affectée à l'hypothèse  $H_1$  après la réalisation  $x$ . Cette probabilité à posteriori de  $H_1$  est calculée à partir des probabilités à priori des hypothèses possibles (ici  $H_1$  et  $H_2$  ;  $P(H_1) + P(H_2) = 1$ ) et des vraisemblances de l'événement observé dans ces différentes hypothèses.

Pour  $H_2$  nous avons :

$$P(H_2 | x) = \frac{P(H_2) P_2(x)}{P(H_1) P_1(x) + P(H_2) P_2(x)}$$

donc

$$\frac{P(H_1 | x)}{P(H_2 | x)} = \frac{P(H_1)}{P(H_2)} \times \frac{P_1(x)}{P_2(x)}$$

H.S Kullback définit l'information en faveur de l'hypothèse  $H_1$  résultant de la réalisation  $X = x$  par la formule :

$$\text{Inf}(1, 2, x) = \text{Log} \frac{P(H_1 | x)}{P(H_2 | x)} - \text{Log} \frac{P(H_1)}{P(H_2)} = \text{Log} \frac{P_1(x)}{P_2(x)}$$

L'information se présente donc comme une différence entre les logarithmes des probabilités à postériori et des probabilités à priori.

Cette différence traduit la modification apportée aux probabilités des hypothèses par l'observation d'une réalisation aléatoire. Il est à remarquer que l'information  $\text{Inf}(1, 2, x)$  n'est autre que le logarithme du rapport de vraisemblances de l'événement  $x$ . Cette grandeur est assez classique en statistiques mathématiques. L'approche Bayésienne et la théorie de l'information utilisent donc les mêmes outils mathématiques que M. S.S. Wilks qui a développé les tests sur les rapports de vraisemblances.

L'information en faveur de  $H_1$  apportée par l'événement  $X = x$  sera donc positive si  $P_1(x) > P_2(x)$  c'est-à-dire si la valeur  $x$  est plus vraisemblable dans l'hypothèse  $H_1$ , elle sera négative dans le cas contraire. Une information nulle est associée à un événement aussi probable dans les deux hypothèses et une information infinie correspond à un événement impossible dans l'une des deux hypothèses.

La base des logarithmes utilisés est arbitraire, on prend souvent les logarithmes en base 2 pour étudier les informations sur des distributions binomiales, ici on adoptera comme M. S KULLBACK les logarithmes népériens qui se prêtent mieux aux calculs numériques et aux développements limités.

Considérons à présent une suite de réalisations aléatoires  $X = x_1, \dots, X = x_n$  en étendant notre définition il vient :

$$\text{Inf}(1, 2, x_1, \dots, x_n) = \log \frac{P(H_1 | x_1, \dots, x_n)}{P(H_2 | x_1, \dots, x_n)} - \log \frac{P(H_1)}{P(H_2)} = \log \frac{P_1(x_1, \dots, x_n)}{P_2(x_1, \dots, x_n)}$$

Si les réalisations sont indépendantes, les probabilités de la suite de réalisations se transforment en produits de probabilités.

$$\text{Inf}(1, 2, x_1, \dots, x_n) = \log \frac{P_1(x_1) \dots P_1(x_n)}{P_2(x_1) \dots P_2(x_n)} = \sum_{i=1}^n \log \frac{P_1(x_i)}{P_2(x_i)}$$

$\text{Inf}(1, 2, x_1, \dots, x_n) = \sum_{i=1}^n \text{Inf}(1, 2, x_i)$  , lorsque les réalisations sont indépendantes en probabilité.

Nous venons de définir l'information totale résultant d'une suite de réalisations aléatoires. On peut maintenant définir une information moyenne attachée à une réalisation. Pour calculer cette moyenne il faut choisir une répartition pour la variable aléatoire X, c'est-à-dire calculer la moyenne dans l'une des deux hypothèses. Nous noterons  $\text{Inf}(1, 2) = E_{H_1} [\text{Inf}(1, 2, x)]$



C'est l'information moyenne apportée par une réalisation de  $X$  en faveur de l'hypothèse  $H_1$  et contre  $H_2$  lorsque  $H_1$  est vraie. En effet ici l'espérance mathématique est calculée avec les probabilités  $P_1(x)$  définies dans l'hypothèse  $H_1$ .

Deux propriétés importantes découlent de la présence de la fonction logarithme dans la définition.

Tout d'abord on a :

$$\text{Inf}(1,2) = \sum_{x \in \{x_i\}} P_1(x) \log \frac{P_1(x)}{P_2(x)} \geq 0 \quad \forall P_1, P_2 \quad \text{avec égalité pour } P_1 \equiv P_2$$

L'information moyenne calculée dans une certaine hypothèse à l'aide de la répartition correspondant à cette hypothèse et apportée en faveur de cette hypothèse est toujours positive. C'est-à-dire que l'observation de la répartition correspondant à  $H_1$  apporte une information favorable à  $H_1$  contre toute autre hypothèse différente de  $H_1$ .

On a aussi :

$$\sum_{x \in \{x_i\}} P_1(x) \log \frac{P_1(x)}{P_2(x)} \geq \sum_{y \in \{y_j\}} P_1(y) \log \frac{P_1(y)}{P_2(y)}$$

avec  $y_j = \cup \{x_i \mid i \in J\}$ , cette propriété de convexité traduit le fait que lorsque les événements possibles sont moins variés (regroupement d'événements ou  $\sigma$ -algèbre moins fine) on acquiert moins d'information en faveur de l'hypothèse choisie. La multiplicité et la variété des réalisations possibles est donc favorable à la discrimination entre les hypothèses.

Étudions un exemple d'application de ces nouvelles notions.  
 Dans les tables de contingence entre deux caractères I et J on est intéressé par l'hypothèse d'indépendance et on désire la comparer à toute autre hypothèse alternative, soit :

$$H_1 \quad p_{ij} \text{ quelconque}$$

$$H_2 \quad p_{ij} = p_i p_j \text{ indépendance entre les caractères}$$

Une observation de table de contingence correspond à N observations d'une loi multinomiale à Card I x Card J classements possibles. D'où :

$$I_n f(1,2) = N \sum_{ij} \log \frac{p_{ij}}{p_i p_j} \times p_{ij}$$

l'information totale = N x l'information moyenne.

Cette grandeur est purement théorique puisque tous ses constituants résultent d'hypothèses prédéterminées. Il est intéressant d'essayer d'estimer les paramètres que nous avons introduit pour obtenir une grandeur numérique aléatoire à laquelle sera attachée une décision. Les quantités  $p_{ij}$  pourront être déterminées par la méthode du maximum de vraisemblance sur l'observation  $n_{ij}$  du

tableau I x J Soit : 
$$I_n f(1,2) = N \sum_{ij} \frac{n_{ij}}{N} \log \frac{n_{ij}}{N p_i p_j} = \sum_{ij} n_{ij} \log \frac{n_{ij}}{N p_i p_j}$$

La répartition  $p_{ij}$  étant calculée sur l'observation, on va alors s'efforcer de minimiser l'information en faveur de l'hypothèse  $H_1$  en estimant les paramètres  $p_i$  et  $p_j$ . Ainsi on définira la répartition indépendante la plus proche, au sens de l'information, de la répartition observée.

$$\frac{\partial}{\partial p_i} = - \sum_j n_{ij} / p_i + \lambda = 0 \Rightarrow p_i = n_i / \lambda$$

$$\sum p_i = 1 \rightarrow \lambda = N$$

Les équations de minimisation sous contraintes conduisent aux fréquences observées :

$$\hat{p}_i = \frac{n_i}{N} \quad ; \quad \hat{p}_j = \frac{n_j}{N} \quad \text{avec} \quad n_i = \sum_j n_{ij} \quad ; \quad n_j = \sum_i n_{ij}$$

La statistique numérique  $\text{Inf}(1^*, 2^*) = \sum_{ij} n_{ij} \text{Log} \frac{n_{ij} N}{n_i n_j}$  mesure donc une certaine distance entre la répartition observée  $n_{ij}/N$  et la répartition indépendante  $n_i n_j / N^2$ . En remarquant que si  $\frac{n_{ij}}{n_i n_j / N} = 1 + \varepsilon$  avec  $\varepsilon$  petit devant 1 on obtient l'équivalence.

$$\begin{aligned} \text{Inf}(1^*, 2^*) &= \sum_{ij} \frac{n_i n_j}{N} (1 + \varepsilon) \text{Log}(1 + \varepsilon) \approx \sum_{ij} \frac{n_i n_j}{N} (1 + \varepsilon) \left( \varepsilon - \frac{\varepsilon^2}{2} \right) \\ &\approx \sum_{ij} \frac{n_i n_j}{N} \left( \varepsilon + \varepsilon^2 - \frac{\varepsilon^2}{2} \right) = \frac{1}{2} \sum_{ij} \frac{n_i n_j}{N} \varepsilon^2 \end{aligned}$$

$$\text{car} \quad \sum_{ij} \frac{n_i n_j}{N} \varepsilon = \sum_{ij} \left( n_{ij} - \frac{n_i n_j}{N} \right) = 0$$

$$\text{Inf}(1^*, 2^*) \approx \frac{1}{2} \sum_{ij} \frac{n_i n_j}{N} \left( \frac{n_{ij} - n_i n_j / N}{n_i n_j / N} \right)^2 = \frac{1}{2} \sum_{ij} \frac{(n_{ij} - n_i n_j / N)^2}{n_i n_j / N} = \frac{\chi^2}{2}$$

Cette information estimée selon les méthodes habituelles est donc équivalente au  $\chi^2$  d'indépendance entre les caractères I et J. La distribution asymptotiquement (pour  $N \rightarrow \infty$ ) de  $2 \text{Inf}(1^*, 2^*)$  est effectivement un  $\chi^2$  à  $(\text{card I} - 1)(\text{card J} - 1)$  degrés de liberté dans l'hypothèse d'indépendance des caractères. Elle permet donc d'effectuer un test d'indépendance sur la table I x J.

L'information a d'autres propriétés d'additivité très intéressantes qui sont liées aux sous-hypothèses de l'hypothèse à laquelle on compare la répartition  $\frac{n_{ij}}{N}$ .

Effectuons une décomposition de l'information en faveur de la répartition  $n_{ij}/N$  contre la répartition d'indépendance définie à priori  $p_i p_j$ . Nous souhaitons une décomposition faisant apparaître des termes ne dépendant que d'un indice :

soit :

$$\text{Inf}(1,2) = \sum_{ij} n_{ij} \log \frac{n_{ij}}{N p_i p_j} = \sum_i n_i \log \frac{n_i}{N p_i} + \sum_j n_j \log \frac{n_j}{N p_j} + \sum_{ij} n_{ij} \log \frac{n_{ij} N}{n_i n_j}$$

$$\text{On vérifie que : } \sum_{ij} n_{ij} \log \frac{n_{ij}}{N p_i p_j} \geq \sum_i n_i \log \frac{n_i}{N p_i}$$

ce qui correspond à la propriété de convexité.

Les deux premiers termes sont équivalents aux  $\chi^2$  suivants

$$0 \leq \sum_i n_i \log \frac{n_i}{N p_i} \approx \sum_i \frac{(n_i - N p_i)^2}{N p_i} = \chi^2_{\text{Card I} - 1}$$

$$0 \leq \sum_j n_j \log \frac{n_j}{N p_j} \approx \sum_j \frac{(n_j - N p_j)^2}{N p_j} = \chi^2_{\text{Card J} - 1}$$

Le troisième terme a déjà été rencontré. Il correspond à l'écart entre la répartition de fréquence observée et la répartition d'indépendance la plus proche.

Les  $\chi^2$  d'adéquation aux marges ne sont assimilables aux deux premiers termes du développement que de façon asymptotique et seulement si les hypothèses de répartition marginale sont vérifiées.

Lorsque l'on désire tester l'hypothèse générale  $H_{p_{ij} = p_i p_j}$ , on peut effectuer une décomposition de l'information qui permettra grâce à :

$$- \sum_{i,j} n_{ij} \text{Log} \frac{n_{ij}}{N p_i p_j} \approx \chi^2_{\text{Card I} \times \text{Card J} - 1}$$

d'accepter ou de rejeter  $H$

$$- \sum_i n_i \text{Log} \frac{n_i}{N p_i} \quad \text{de retenir ou de refuser la}$$

répartition selon les marges

- idem en j

$$- \sum_{i,j} n_{ij} \text{Log} \frac{n_{ij}}{N p_i p_j} \quad \text{d'accepter ou de rejeter l'hypothèse d'indé}$$

pendance entre les caractères.

Contrairement à la décomposition de la variance, l'étude des composantes additives d'une information totale permet de tester des sous-hypothèses avec les critères habituels, et donc dans les meilleures conditions de risque :

Une telle décomposition peut se résumer dans le tableau suivant

| Composantes dues                                 | 2 x information                                   | degrès de liberté des distributions asymptotiques en $\chi^2$ |
|--|---|---|
| aux lignes $H(I)$<br>$\hat{p} = \frac{n_i}{N}$   | $\sum n_i \text{Log} \frac{n_i}{N p_i}$           | $\text{Card I} - 1$   |
| aux colonnes $H(J)$<br>$\hat{p} = \frac{n_j}{N}$ | $\sum n_j \text{Log} \frac{n_j}{N p_j}$           | $\text{Card J} - 1$   |
| à l'indépendance $H(I \times J)$                 | $\sum n_{ij} \text{Log} \frac{n_{ij} N}{n_i n_j}$ | $(\text{Card I} - 1) \times (\text{Card J} - 1)$              |
| total $2 \text{Inf}(1, 2)$                       | $\sum n_{ij} \text{Log} \frac{n_{ij}}{N p_i p_j}$ | $\text{Card I} \times \text{Card J} - 1$                      |

On remarque que le degré de liberté du  $\chi^2$  équivalent à  $2 \text{ Inf}(1^*, 2_2)$  est égal à la somme des degrés de liberté des informations affectées aux sous-hypothèses. Lorsque  $H$  n'est pas vérifiée  $2 \text{ Inf}(1^*, 2^*)$  est distribuée selon un  $\chi^2$  excentré dont le paramètre d'excentration est égal à  $\text{Inf}(1, 2)$  de.

#### I V2 - APPLICATION AUX CONTINGENCES ENTRE TROIS CARACTERES

Les notations sont les mêmes que précédemment. La plupart des hypothèses et sous-hypothèses qui seront développées ont déjà été rencontrées dans la décomposition de la variance du chapitre II.

1) a) Les hypothèses sont définies à priori :

- généralisation de l'étude précédente

$H_1$   $p_{ijk}$  est quelconque

$H_2$   $p_{ijk} = p_i p_j p_k$  indépendance totale avec des paramètres

fixes à priori.

Les résultats sont analogues à ceux de tables bidimensionnelles:

| Composantes dues                                     | Information   | d.d.l. $\chi^2$ limite                                  |
|--|---|---|
| aux lignes $H_2(I)$<br>$\hat{p}_i = \frac{n_i}{N}$   | $\sum_I n_i \log \frac{n_i}{N p_i}$                       | Card I - 1  |
| aux colonnes $H_2(J)$<br>$\hat{p}_j = \frac{n_j}{N}$ | $\sum_J n_j \log \frac{n_j}{N p_j}$                       | Card J - 1  |
| aux couches $H_2(k)$<br>$\hat{p}_k = \frac{n_k}{N}$  | $\sum_k n_k \log \frac{n_k}{N p_k}$                       | Card K - 1  |
| à l'indépendance<br>$H_2(I \times J \times k)$       | $\sum_{ijk} n_{ijk} \log \frac{n_{ijk} N^2}{n_i n_j n_k}$ | Card I x Card J x Card k - Card I - Card J - Card k + 2 |
| total<br>$\text{Inf}(H_2)$                           | $\sum_{ijk} n_{ijk} \log \frac{n_{ijk}}{N p_i p_j p_k}$   | Card I x Card J x Card k - 1                            |

Les  $\hat{p}$  représentent les estimations du maximum de vraisemblance ou ce qui est équivalent du minimum de l'information pour les paramètres  $p_i, p_j, p_k$  de l'hypothèse  $H_2$ .

On voit que l'adéquation générale à  $p_i, p_j, p_k$  se décompose en trois adéquations marginales et une indépendance.

$$H_2 = H_2(I) \cap H_2(J) \cap H_2(K) \cap H_2(I \times J \times K)$$

Les degrés de liberté se retrouvent facilement connaissant la forme algébrique des  $\chi^2$  équivalents.

- L'indépendance entre un caractère et les deux autres.

L'hypothèse intéressante est alors :

$$H_{p_{ijk}} = p_i p_j p_k \quad \text{hypothèse de base des correspondances ternaires}$$

Elle est équivalente à une indépendance sur une table à deux entrées de dimension  $\text{card I} \times (\text{card J} \times \text{card K})$ . On a en effet :

| Composantes dues   | Information   | degrès de liberté :  |
|--|---|--|
| au caractère I $H(I)$<br>$\hat{p}_i = \frac{n_i}{N}$                     | $\sum_i n_i \log \frac{n_i}{N p_i}$                     | $\text{Card I} - 1$  |
| aux colonnes et couches<br>$\hat{p}_{jk} = \frac{n_{jk}}{N} \quad H(JK)$ | $\sum_{jk} n_{jk} \log \frac{n_{jk}}{N p_{jk}}$         | $\text{Card J} \times \text{Card K} - 1$                       |
| à l'indépendance $H(I \times JK)$  | $\sum_{ijk} n_{ijk} \log \frac{n_{ijk} N}{n_i n_{jk}}$  | $(\text{Card I} - 1) (\text{Card J} \times \text{Card K} - 1)$ |
| Total H  | $\sum_{ijk} n_{ijk} \log \frac{n_{ijk}}{N p_i p_j p_k}$ | $\text{Card I} \times \text{Card J} \times \text{Card K} - 1$  |

On retrouve la décomposition :

$$H = H(I) \cap H(JK) \cap H(I, JK)$$

adéquation générale = adéq. ligne  $\cap$  adéq. colo. x couche  $\cap$  Indép. I et JK

— L'indépendance conditionnelle entre I et J si k.

La répartition correspondant à cette hypothèse est :

$$\exists k \quad p_{ijk} = \frac{p_{ik} p_{jk}}{p_k}$$

Pour une couche du tableau il y a indépendance entre I et J. Le estimateurs des probabilités conditionnelles sont :

$$\frac{\hat{p}_{ik}}{\hat{p}_k} = \frac{n_{ik}}{n_k} \quad \frac{\hat{p}_{jk}}{\hat{p}_k} = \frac{n_{jk}}{n_k}$$

l'adéquation générale se fait entre la densité conditionnelle sur I x J x K

$\frac{p_{ijk}}{p_k} = \frac{p_{ik}}{p_k} \times \frac{p_{jk}}{p_k}$  et la densité observée  $n_{ijk}/n_k$  d'où le tableau suivant

| composantes dues                         | $\lambda$ * Information  | degrès de liberté         |
|--|--|---------------------------|
| aux lignes H(I k)                        | $2 \sum_i h_{ik} \log \frac{n_{ik}}{n_k \left( \frac{p_{ik}}{p_k} \right)}$          | Card I - 1                |
| aux colonnes H(J k)                      | $2 \sum_j n_{jk} \log \frac{n_{jk}}{n_k \left( \frac{p_{jk}}{p_k} \right)}$          | Card J - 1                |
| à l'indépendance condit.<br>H(I x J   k) | $2 \sum_{ij} n_{ijk} \log \frac{n_{ijk}}{\frac{n_{ik} n_{jk}}{n_k}}$                 | (Card I - 1) (Card J - 1) |
| Total H                                  | $2 \sum_{ij} n_{ijk} \log \frac{n_{ijk}}{n_k \frac{p_{ik}}{p_k} \frac{p_{jk}}{p_k}}$ | Card I x Card J - 1       |



Le conditionnement est défini ici par rapport à un classement  $l$ . Dans les informations il n'y a pas de sommation par rapport à cet indice  $k$ . Pour retrouver l'hypothèse étudiée dans la théorie des tests qui peut s'écrire :

$$P_{ijk} = \frac{P_{i.} P_{.k}}{P_{..}} \quad \forall k \in K \quad \text{Indépendance conditionnelle}$$

entre  $I$  et  $J$  si  $K$ , il suffit d'effectuer des sommations sur  $k$ , sur les informations obtenues ci-dessus.

On fait ainsi une moyenne de ces informations pour toutes les profondeurs du tableau et les degrés de liberté des distributions asymptotiques en  $\chi^2$  sont multipliés par  $\text{card } K$ .

En particulier on obtient pour l'indépendance :

$$H(IJ/K) = H(I/K \times J/K) \quad 2 \cdot \text{Inf} = 2 \sum_{ijk} n_{ijk} \log \frac{n_{ijk} n_{..}}{n_{i.} n_{.k}}$$

$$2 \cdot \text{Inf} \rightarrow \sum_{ijk} (n_{ijk} - n_{i.} n_{.k} / n_{..})^2 / (n_{i.} n_{.k} / n_{..}) \quad \text{si } H \text{ est vraie}$$

Grâce aux propriétés de la fonction logarithme les décompositions de l'information relative à une adéquation sur le tableau complet nous ont toujours permis de retrouver des composantes relatives aux adéquations sur les marges et l'hypothèse d'indépendance complémentaire. Les informations calculées ont en plus leur additivité, toutes les propriétés limites optimales pour effectuer correctement les tests d'hypothèses et de sous-hypothèses. Ici il n'est pas nécessaire que l'hypothèse soit vérifiée pour que les informations relatives aux sous-hypothèses aient les distributions limites souhaitées. Il serait intéressant de scinder en composantes additives l'information relative à l'indépendance entre les trois caractères, et d'essayer de retrouver des statistiques permettant de tester les sous-hypothèses définies au chapitre II.

2) Les paramètres des hypothèses sont estimés :

- Indépendance (I x J x K) = Indép I x (JK)  $\wedge$  Indép J x K.

On a bien grâce à la convexité de  $x \log x$

$$N \left\{ = \sum_{ijk} n_{ijk} \log \frac{n_{ijk} N^2}{n_i n_j n_k} \geq \sum_{jk} \left( \sum_i n_{ijk} \right) \log \frac{\sum_i n_{ijk} N^2}{\sum_i n_i n_j n_k} \right.$$

$$\left. \cong \sum_{jk} n_{ijk} \log \frac{n_{ijk} N}{n_j n_k} \right.$$

$$\text{Inf}(I \times J \times K) \geq \text{Inf}(J \times K)$$

$$\sum_{ijk} \left( n_{ijk} \log \frac{n_{ijk} N^2}{n_i n_j n_k} - n_{ijk} \log \frac{n_{ijk} N}{n_j n_k} \right) = \sum_{ijk} n_{ijk} \log \frac{n_{ijk} N}{n_i n_j n_k}$$

La différence entre ces deux informations est bien égale à l'information associée à l'hypothèse d'indépendance entre I et J x K. Il y a une addition parfaite des informations liées aux deux sous-hypothèses de l'indépendance :

$$\text{Inf}(I \times J \times K) = \text{Inf}(I \times JK) + \text{Inf}(J \times K)$$

Et dans l'hypothèse d'indépendance on a l'égalité pour N-infini :

$$\chi^2_{IJK} = \chi^2_{JK} + \chi^2_{I \times JK}$$

Les  $\chi^2$  représentent ici les statistiques qui permettent habituellement de tester les sous-hypothèses (cf. SN ROY). Ce ne sont pas les grandeurs rencontrées au chapitre II qui vérifient l'égalité de parleur définition et pas seulement de façon asymptotique.

Progressons dans la décomposition : le seul terme intéressant est celui qui utilise trois indices soit Inf (I x JK)

- Indépendance(I x JK) = Indép.(I x K)  $\wedge$  Indép.(I x J/K)

De l'indépendance entre un caractère et l'ensemble des deux autres on extrait l'indépendance entre deux caractères. L'hypothèse complémentaire sera une hypothèse d'indépendance conditionnelle.

Rappelons qu'à l'indépendance I x JK nous avons affecté la statistique  $\chi^2_{IJK} - \chi^2_{JK} = \underbrace{\chi^2_{IJ}} + \chi^2_{IK} + \underbrace{\chi^2_{Inter.}}$

en utilisant les éléments de la variance totale étudiée au chapitre II. GN SNEDECOR suggère d'utiliser cette grandeur pour tester l'hypothèse H(I x JK) ; mais nous avons vu que la variance de l'interaction

$$\chi^2_{Int.} = \underbrace{\chi^2_{IJK} - \chi^2_{IJ} - \chi^2_{JK} - \chi^2_{IK}}_{\chi^2 \text{ d'indépendance habituels}}$$

dont s'est servi H. O. LANCASTER, n'avait des propriétés intéressantes que dans l'hypothèse d'indépendance globale. Ces grandeurs ne sont utilisables qu'au voisinage de l'indépendance. J. N. DARROCK a montré qu'il était possible de tester correctement H(I,JK) à l'aide de  $\chi^2_{IJ} + \chi^2_{JK} + \chi^2_{Int.}$  mais en prenant pour  $\chi^2_{Int.}$  le  $\chi^2$  d'interaction défini et calculé par SN. ROY.

Voyons comment la théorie de l'information aborde ces difficultés de décomposition.

$\chi^2_{IJK}$  $\chi^2_{JK}$ 

$$\sum_{ijk} n_{ijk} \text{Log} \frac{n_{ijk} N}{n_{jk} n_i} \geq \sum_{ik} n_{ik} \text{Log} \frac{n_{ik} N}{n_k n_i}$$

$$= \sum_{ik} n_{ik} \text{Log} \frac{n_{ik} N}{n_k n_i} + \sum_{ijk} n_{ijk} \text{Log} \frac{n_{ijk} n_k}{n_{jk} n_{ik}}$$

Le complément à la statistique de l'indépendance I x K est l'information résultant de la comparaison des deux densités  $n_{ijk}$  et  $n_{ik} n_{jk} / n_k$ . C'est l'information affectée à l'indépendance conditionnelle entre i et j pour k fixé

Jusqu'ici il n'y a pas de problème de décomposition. La composante que nous avons trouvée correspond à  $\chi^2_{IJ} + \chi^2_{IK} + \chi^2_{JK}$ . Passons à la dernière étape.

- Indépendance conditionnelle IJ/K =  
Indépendance(I x J)  $\cap$  Non interaction

Si cette addition de sous-hypothèses correspond à une addition des informations on doit obtenir :

$$\text{Inf (non interaction)} = \text{Inf (IJ/K)} - \text{Inf (IJ)}$$

$$= \sum_{ijk} \left( n_{ijk} \text{Log} \frac{n_{ijk} n_k}{n_k n_{jk}} - n_{ijk} \text{Log} \frac{n_{ij} N}{n_i n_j} \right)$$

$$= \sum_{ijk} n_{ijk} \text{Log} \frac{n_{ijk} n_i n_j n_k}{N n_{ij} n_{ik} n_{jk}}$$

Il s'agit d'une comparaison entre la densité observée  $n_{ijk} / N$  et la répartition suivante  $n_{ij} n_{ik} n_{jk} / n_i n_j n_k$

Cette répartition semble correspondre à l'hypothèse :

$$(1) \quad p_{ijk} = \frac{p_{ij} p_{ik} p_{jk}}{p_i p_j p_k}$$

de S.N. ROY en remplaçant les probabilités par les fréquences observées

$$p_{ij} \rightarrow \frac{n_{ij}}{N}, \quad p_{jk} \rightarrow \frac{n_{jk}}{N}, \quad p_{ik} \rightarrow \frac{n_{ik}}{N}, \quad p_i \rightarrow \frac{n_i}{N}, \quad p_j \rightarrow \frac{n_j}{N}, \quad p_k \rightarrow \frac{n_k}{N}$$

essayons d'utiliser la convexité de l'information

$$\begin{aligned} \sum_{ijk} n_{ijk} \log \frac{n_{ijk}}{\frac{n_{ik} n_{jk}}{n_k}} &\geq \sum_{ij} n_{ij} \log \frac{\sum_k n_{ijk}}{\sum_k \frac{n_{ik} n_{jk}}{n_k}} \\ &\geq \sum_{ij} n_{ij} \log \frac{n_{ij}}{\sum_k \frac{n_{ik} n_{jk}}{n_k}} \end{aligned}$$

La densité induite par cette opération de concentration est

$\sum_n \frac{n_{ik} n_{jk}}{n_k N}$  et non  $n_i \cdot n_j / N$ . Lorsqu'il y a égalité entre ces densités la grandeur complémentaire est bien positive et peut être associée à l'interaction. Nous verrons un exemple réel où la quantité complémentaire est négative

Le caractère négatif de la quantité  $\sum n_{ijk} \log \frac{n_{ijk}}{N \left( \frac{n_{ik} n_{jk}}{n_i n_j n_k} \right)}$  n'est

pas en contradiction avec les propriétés générales de l'information. En effet la répartition  $u_{ijk} = \frac{n_{ij} n_{ik} n_{jk}}{n_i n_j n_k}$  n'est pas une répartition de fréquence, nous n'avons pas  $\sum_{ijk} u_{ijk} = 1$  pour tous les tableaux.

Par exemple pour le tableau 2 x 2 x 2.

$$\left. \begin{aligned} n_{111} = n_{222} = 1 \\ n_{ijk} = 0 \text{ ailleurs} \end{aligned} \right\} \text{ On obtient } \begin{cases} u_{111} = 1 \\ u_{222} = 1 \\ u_{ijk} = 0 \text{ ailleurs} \end{cases} \quad \sum_{ijk} u_{ijk} = 2$$

Remarquons que les conditions :

$$\forall ij \quad \sum_k \frac{n_{ik} n_{jk}}{n_k} = \frac{n_i n_j}{N}$$

correspondent aux contraintes surabondante

qui ont fait rejeter la densité (1) par SN. ROY. En effet ces égalités restreignent la généralité du tableau. Elles définissent des conditions sur les marges, ce qui n'est pas admissible pour une hypothèse de non-interaction. J.N. DARROCK a vérifié que pour de nombreux tableaux ces égalités étaient presque vérifiées.

Il a développé des tests s'appuyant sur la densité  $\frac{n_{ij} n_{ik} n_{jk}}{n_i n_j n_k}$  pour des tables qu'il nomme parfaites.

Au voisinage de l'indépendance nous avons

$$\sum_k \frac{n_{ik} n_{jk}}{n_k} \simeq \sum_k \left( \frac{n_i n_k}{N} \times \frac{n_j n_k}{N} \right) / n_k = \sum_k \frac{n_i n_j}{N^2} n_k = \frac{n_i n_j}{N}$$

Pour les tableaux presque indépendants, correspondant aux tables parfaites de J.N. DARROCK, l'information complémentaire peut servir à tester la non interaction, tout comme la variance résiduelle  $\chi^2_{Jnt}$  ou le  $\chi^2$  suivant

$$\sum_{ijk} \left( n_{ijk} - (n_{i.} n_{.j} n_{.k} / n_i n_j n_k) \right)^2 / (n_{i.} n_{.j} n_{.k} / n_i n_j n_k)$$

- Nouvelle définition de la non-interaction.

En utilisant la convexité de l'information on peut décomposer l'hypothèse d'indépendance conditionnelle en deux nouvelles sous-hypothèses, et définir de nouvelles informations partielles toujours positives et qui tendront vers des  $\chi^2$  pour un effectif N très grand.

Nous avons :

$$\sum_{iik} n_{iik} \text{Log} \frac{n_{iik}}{\frac{n_{ik} n_{ij}}{n_k}} \geq \sum_{ij} n_{ij} \text{Log} \frac{n_{ij}}{\sum_k \frac{n_{ik} n_{jk}}{n_k}} = \text{Inf} (IJ \text{ par } k)$$

Cette dernière information permet de comparer la répartition marginale  $n_{ij}$  à la répartition sur cette même marge I x J induite par les deux autres marges et par l'hypothèse d'indépendance conditionnelle relativement à K. Nous verrons que cette densité constitue en quelque sorte une prévision de l'association I x J.

L'information Inf (IJ par K) est toujours positive.

$$\frac{\hat{n}_{ij}}{N} = \sum_k \frac{n_{ik} n_{jk}}{n_k N}$$

définit bien une densité de fréquence

(on a même :  $\sum_i \hat{n}_{ij} = n_j$  ;  $\sum_j \hat{n}_{ij} = n_i$  )

La distribution limite est un  $\chi^2$  à (card I-1) (Card J-1) degrés de liberté.

La sous-hypothèse liée à cette information est appelée (I x J) interaction par M. S KULLBACK. L'hypothèse complémentaire pour retrouver l'indépendance conditionnelle est une interaction entre les trois caractères. Son expression n'est pas symétrique selon les trois caractères. Cette (IJ x K) interaction se manifeste par l'information :

$$\text{Inf} (IJ \times k \text{ Inter.}) = \sum_{ijk} n_{ijk} \text{Log} \frac{n_{ijk}}{\frac{n_{i.} n_{.j} n_{.k}}{\hat{n}_{ij} n_{.k}}}$$

Sa distribution limite est un  $\chi^2$  à (card I-1) (card J-1) (card K-1) degrés de liberté, et dans l'hypothèse d'indépendance totale ( $\hat{n}_{ij} = n_{i.} n_{.j} / N$ )

elle s'identifie à l'information complémentaire du paragraphe précédent :

$$\text{Inf} + (IJ/K) - \text{Inf} (I \times J).$$



V - COMPARAISON DES TROIS APPROCHES

VI Equivalence au voisinage de l'indépendance

Lorsqu'on aborde l'étude d'une table de contingence sans avoir de renseignement sur sa structure, on s'intéresse tout d'abord à l'indépendance entre les caractères. On exprime les paramètres du modèle à l'aide d'estimateurs résultants de la table de contingence. C'est pourquoi la décomposition de la fonction écart a été effectuée au voisinage de cette hypothèse. C'est-à-dire que l'écart était calculé entre l'observation  $f_{IJK}$  et la densité estimée  $f_I f_J f_K$ ; la métrique utilisée correspondait à cette même densité indépendante. Outre l'importance de cette hypothèse pour juger des associations entre caractères, on doit remarquer qu'elle est la seule à conduire à des décompositions très interprétables puisque les écarts à la densité de référence s'expriment par des produits de facteurs sur un caractère, de moyenne nulle et non corrélés.

Pour comparer les trois études de tables de contingence nous nous placerons donc dans cette hypothèse générale. Cela suppose des distributions de probabilité bien définies pour toutes les grandeurs calculées. Cette restriction n'est pas très gênante pour la première étape de l'analyse d'une contingence puisqu'il ne s'agit alors que de préciser la structure du tableau.

Nous travaillerons uniquement sur des fréquences observées

$f_{IJK}$ ,  $f_{IJ}$ ,  $f_{JK}$ ,  $f_{IK}$ ,  $f_I$ ,  $f_J$ ,  $f_K$  et nous poserons :

$$\begin{aligned} f_{IJK} &= f_I f_J f_K (1 + \epsilon) \\ f_{IJ} &= f_I f_J (1 + \epsilon_{IJ}) \\ f_{JK} &= f_J f_K (1 + \epsilon_{JK}) \\ f_{IK} &= f_I f_K (1 + \epsilon_{IK}) \end{aligned} \quad \left. \begin{array}{l} \epsilon \text{ représente la fonction écart} \\ \text{définissent les écarts entre} \\ \text{les fréquences marginales et} \\ \text{les densités estimées} \end{array} \right\}$$

L'interaction que nous avons définie comme l'écart complémentaire est égale à  $(\varepsilon - \varepsilon_{II} - \varepsilon_{IK} - \varepsilon_{JK})$

Dans l'hypothèse d'indépendance entre les trois caractères les fréquences observées suivent des lois connues qui s'expriment en fonction des probabilités réelles des modalités  $i, j$  et  $k$ . Lorsque l'effectif  $N$  du tableau est important et que chaque association  $(ijk)$  a été rencontrée une dizaine de fois ces lois sont assimilables à des lois de Gauss-Laplace.

$$f_{ijk} \rightarrow \mathcal{N}\left(p_i p_j p_k, \frac{p_i p_j p_k (1 - p_i p_j p_k)}{N}\right) = \mathcal{N}(m, \sigma^2)$$

Dans tous les tableaux suffisamment importants et bien équilibrés on aura:  $1 - p_i p_j p_k \approx 1$ .

On peut montrer que la densité  $f_{IJK}$  produit des densités marginales a une variance qui décroît beaucoup plus vite que celle de  $f_{IJK}$  pour  $N$  augmentant  $E(f_{IJK}) = p_i p_j p_k$   $\text{Var}(f_{IJK}) \approx \sigma^2 (p_i p_j + p_i p_k + p_j p_k - 3 p_i p_j p_k)$

On a donc l'équivalence suivante :

$$\mathcal{N}(p_i p_j p_k, p_i p_j p_k (1 - p_i p_j p_k) / N) \approx \mathcal{N}(f_i f_j f_k, f_i f_j f_k / N)$$

La fonction écart  $\varepsilon$  est une variable aléatoire puisqu'elle est définie à partir des fréquences, elle possède les propriétés suivantes :

$$\varepsilon = \frac{f_{ijk} - f_i f_j f_k}{f_i f_j f_k} = \frac{f_{ijk} - f_i f_j f_k}{\sqrt{f_i f_j f_k}} \times \frac{1}{\sqrt{N f_i f_j f_k}} \approx \mathcal{N}(0,1) \times \frac{C^*}{\sqrt{N}}$$

Elle est proportionnelle à une loi normale centrée, le coefficient de proportionnalité est en  $1/\sqrt{N}$ .

$\epsilon$  Converge donc en probabilité vers zéro pour N augmentant indéfiniment

Pour les autres écarts nous obtenons les mêmes propriétés,  $f_{IJ}$  étant un estimateur plus convergent de  $p_I p_J$  que l'observation marginale  $f_{IJ}$ , nous pouvons opérer les mêmes substitutions. Tous les écarts sont des variables aléatoires normales centrées, infiniment petites d'ordre 1/2 en N.

La conclusion de ces remarques sur les lois de probabilité est que certaines approximations pourront avoir lieu dans les développements limités. Ainsi on pourra négliger les  $\epsilon$  devant 1, et les carrés ou produits d' $\epsilon$  devant les  $\epsilon$  eux-mêmes.

— Les  $\chi^2$  et les variances des écarts

$$\text{Nous avons : } \chi^2 = N \sum_{iik} \left( \frac{f_{iik} - f_i f_j f_k}{f_i f_j f_k} \right)^2 f_i f_j f_k = N \| \epsilon \|^2_{f_i f_j f_k}$$

Le  $\chi^2$  d'indépendance globale nous avait permis de définir notre fonction écart.

$$\chi_{ij}^2 = N \sum_{iij} \left( \frac{f_{iij} - f_i f_j}{f_i f_j} \right)^2 f_i f_j = N \| \epsilon_{ij} \|^2_{f_i f_j} = N \| \epsilon_{ij} \|^2_{f_i f_j f_k}$$

Pour toutes les faces du tableau il y a équivalence entre le  $\chi^2$  de contingence et la variance partielle. Le  $\chi^2$  d'interaction défini par la décomposition était :

$$\begin{aligned} \chi_I^2 &= \chi^2 - \chi_{IJ}^2 - \chi_{IK}^2 - \chi_{JK}^2 = N \left( \| \epsilon \|^2 - \| \epsilon_{IJ} \|^2 - \| \epsilon_{IK} \|^2 - \| \epsilon_{JK} \|^2 \right) \\ &= N \left( \| \epsilon - \epsilon_{IJ} - \epsilon_{IK} - \epsilon_{JK} \|^2 \right) = N \text{Var}(\text{Interaction}) \end{aligned}$$

En vertu des propriétés d'orthogonalité des sous-espaces de projection.

Le  $\chi^2$  d'interaction obtenu par S.N. ROY ne s'exprime pas en fonction des densités observées. Il faudra donc comparer les deux approches à l'aide des  $\chi^2$  correspondant aux regroupements de sous-hypothèses (Indépendance (IJ et K ou (IJ si K)).

L'interaction obtenue par décomposition peut se rapprocher du  $\chi^2$  permettant de tester la non interaction dans les tables parfaites de JN DARRC en effet :

$$\begin{aligned} \chi^2 &= N \sum_{ijk} (f_{ijk} - f_{ij} \cdot f_{jk} \cdot f_{ik} / f_i f_j f_k)^2 / (f_{ij} f_{jk} f_{ik} / f_i f_j f_k) \\ &= N \sum_{ijk} f_i f_j f_k \left[ \frac{(1 + \varepsilon_{ijk}) - (1 + \varepsilon_{ij})(1 + \varepsilon_{jk})(1 + \varepsilon_{ik})}{(1 + \varepsilon_{ij})(1 + \varepsilon_{jk})(1 + \varepsilon_{ik})} \right]^2 \\ &= N \left\| \left[ 1 + \varepsilon_{ijk} - (1 + \varepsilon_{ij} + \varepsilon_{jk} + \varepsilon_{ik} + \mathcal{O}\left(\frac{1}{N}\right)) \right] / \left( 1 + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right) \right) \right\|_{f_i f_j f_k}^2 \\ &\simeq N \left\| \varepsilon_{ijk} - \varepsilon_{ij} - \varepsilon_{jk} - \varepsilon_{ik} \right\|_{f_i f_j f_k}^2 = \chi_{II}^2 \end{aligned}$$

Pour l'indépendance conditionnelle entre deux caractères nous disposons de :

$$\chi_{II}^2 + \chi_{II}^2 = \chi^2 - \chi_{IK}^2 - \chi_{JK}^2$$

$$\text{et de } \chi_{(IJ/K)}^2 = N \sum_{ijk} (f_{ijk} - f_{ik} f_{jk} / f_k)^2 / (f_{ik} f_{jk} / f_k)$$

$$\text{Or } \chi_{II}^2 + \chi_{II}^2 = N \left\| \varepsilon_{ijk} - \varepsilon_{ik} - \varepsilon_{jk} \right\|^2$$

$$\left\{ \begin{aligned} N \sum_{ijk} (f_{ijk} - f_{ik} f_{jk} / f_k)^2 / (f_{ik} f_{jk} / f_k) &= N \sum_{ijk} f_i f_j f_k \frac{(1 + \varepsilon_{ijk}) - (1 + \varepsilon_{ik})(1 + \varepsilon_{jk})}{(1 + \varepsilon_{ik})(1 + \varepsilon_{jk})} \\ \chi_{(IJ/K)}^2 &= N \left\| (\varepsilon_{ijk} - \varepsilon_{ik} - \varepsilon_{jk}) / (1 + \varepsilon_{ik} + \varepsilon_{jk} + \varepsilon_{ik} \varepsilon_{jk}) \right\|^2 \\ &\simeq N \left\| \varepsilon_{ijk} - \varepsilon_{ik} - \varepsilon_{jk} \right\|^2 \end{aligned} \right.$$

$$\text{pour } N \rightarrow \infty \quad \chi_{(II/K)}^2 \simeq \chi_{II}^2 + \chi_{II}^2$$

La seconde sous-hypothèse résultant de regroupement est l'indépendance entre un caractère et les deux autres.

$$\begin{aligned}
 \chi^2_{(I, j, k)} &= N \sum_{ijk} (f_{ijk} - f_{ij} f_k)^2 / f_{ij} f_k \\
 &= N \sum_{ijk} f_i f_j f_k (1 + \epsilon) - (1 + \epsilon_{ij})^2 / (1 + \epsilon_{ij}) \\
 &\approx N \| \epsilon - \epsilon_{ij} \|^2 = \chi^2 - \chi^2_{I, j} \\
 &= \chi^2_I + \chi^2_{I, k} + \chi^2_{j, k}
 \end{aligned}$$

Le regroupement :

$$\chi^2_I + \chi^2_{I, k} + \chi^2_{j, k} = \chi^2 - \chi^2_{I, j}$$

est bien un indicateur de validité pour la sous-hypothèse testée par  $\chi^2_{(I, j, k)}$

Au voisinage de l'indépendance totale et pour un effectif assez important, les variances (ou normes) qui apparaissent dans la décomposition de la variance de la fonction écart sont équivalentes au  $\chi^2$  permettant de tester les sous-hypothèses de l'indépendance totale.

Dans ces conditions la décomposition des  $\chi^2$  de contingence se justifie et elle permet de mettre en évidence les sous-hypothèses qui pourront être acceptées ou les associations qui méritent une explication complémentaire (autre modèle, analyse factorielle...).

— L'équivalence entre information et  $\chi^2$ .

1  
2  
3

Par transitivité nous aurons une équivalence générale entre les trois méthodes de calcul.

L'équivalence au voisinage de l'indépendance entre les informations et les  $\chi^2$  peut se démontrer pour toutes les densités équivalentes. Nous en avons un exemple dans le chapitre sur l'information, pour une table de contingence à deux entrées.

$$2 \text{ Inf} = 2N \sum p \log \frac{p}{\hat{p}} \quad ; \quad \chi^2 = N \sum \left( \frac{p - \hat{p}}{\hat{p}} \right)^2 \hat{p}$$

Posons des conditions d'équivalence sur les densités

$$p \rightarrow \hat{p} \quad \text{pour } N \rightarrow \infty \quad ; \quad p = \hat{p}(1+e) \quad ; \quad e \rightarrow 0$$

Elles correspondent aux cas déjà rencontrés, e ayant successivement pour partie principale,  $\epsilon$ ,  $\epsilon_{11}$ ,  $\epsilon_{1k}$ ,  $\epsilon_{1k}$ ,  $\epsilon - \epsilon_{11} - \epsilon_{1k} - \epsilon_{1k}$ ,  $\epsilon - \epsilon_{1k} - \epsilon_{1k}$ , et  $\epsilon - \epsilon_{11}$ .

$$2 \text{ Inf} = 2N \sum \hat{p}(1+e) \log(1+e) = 2N \sum \hat{p}(1+e) \left( e - \frac{e^2}{2} + \dots \right)$$

$$\approx 2N \sum \hat{p}e + N \sum e^2 \hat{p} = N \sum \hat{p} \left( \frac{p - \hat{p}}{\hat{p}} \right)^2$$

Or  $\sum \hat{p} = 1$ ,  $\sum p = 1$

$$\sum p = \sum \hat{p} + \sum \hat{p}e \Rightarrow \sum \hat{p}e = 0$$

Ceci est vrai pour toute densité, la répartition  $\frac{n_{11} n_{1k} n_{2k}}{n_{1.} n_{.1} n_{.k}}$

qui conduit à une interaction négative (cf Chap IV, 2ème partie) ne vérifie pas cette condition.

$$2 \text{ Inf} \approx N \sum e^2 \hat{p}$$

$$\chi^2 = N \sum \left( \frac{p - \hat{p}}{\hat{p}} \right)^2 \hat{p} = N \sum e^2 \hat{p}$$

B

B (

Nous avons ainsi obtenu une équivalence entre information et  $\chi^2$  sous la seule condition que l'écart entre les densités soit un infiniment petit.

Ceci peut être vrai en dehors de l'indépendance totale.

En particulier cette condition est vérifiée si  $p \rightarrow \hat{p}$  c'est-à-dire si l'hypothèse liée à  $\hat{p}$  est vérifiée puisqu'alors  $E(p) = E(\hat{p})$ ; les densités de fréquence observées et estimées ont même espérance et elles convergent vers la densité définie par l'hypothèse.

L'information a donc les mêmes propriétés optimales que les  $\chi^2$  pour effectuer des tests d'hypothèses ou de sous-hypothèses dans les tables de contingence. Pour illustrer à la fois les ressemblances et les différences qui peuvent exister entre ces méthodes nous avons choisi deux exemples numériques extrêmes. Le premier tableau est proche de l'indépendance et toutes les équivalences sont admissibles, le second en est très éloigné et certains tests de sous-hypothèses ne sont plus valables.

## V2 - Deux exemples numériques

- Les productions de quatre industriels A, B, C et D sont soumises à deux tests de qualité  $T_1$  et  $T_2$ , qui donnent lieu à des acceptations ou des refus pour les échantillons présentés. Le test, le résultat et l'origine du produit jouent le rôle des trois caractères mis en correspondance, il y a au total  $2 \times 2 \times 4 = 16$  cellules dans le tableau de contingence.

Les données numériques sont les suivantes :

|   | T <sub>1</sub> |       | T <sub>2</sub> |       |
|---|----------------|-------|----------------|-------|
|   | Accept.        | Refus | Accept.        | Refus |
| A | 112            | 32    | 84             | 24    |
| B | 76             | 20    | 86             | 10    |
| C | 87             | 9     | 58             | 14    |
| D | 41             | 7     | 40             | 8     |

Les calculs de variance,  $\chi^2$  et information ont été fait pour le hypothèses et sous-hypothèses résultant de la décomposition de la fonction écart

|                       | Décomposition de la variance. | Test de $\chi^2$ | Information |
|-----------------------|-------------------------------|------------------|-------------|
| Indépendance totale   |                               |                  |             |
| 10 degrés de liberté  | 16,82                         | 16,82            | 17,27       |
| Indépendance          |                               |                  |             |
| Producteur x Test     | 3,51                          | 3,51             | 3,51        |
| 4 d.d.l.              |                               |                  |             |
| Indépendance          |                               |                  |             |
| Producteur x Résultat | 6,28                          | 6,28             | 6,17        |
| 4 d. d. l.            |                               |                  |             |



|                              |       |        |       |
|------------------------------|-------|--------|-------|
| Indépendance                 |       |        |       |
| Test x Résultat              | 0,02  | 0,02   | 0,02  |
| 1. d. d. l.                  |       |        |       |
|                              |       |        |       |
| Intéraction                  | 7,01  | (7,57) | 7,57  |
| 3. d. d. l.                  |       |        |       |
|                              |       |        |       |
| Indépendance conditionnelle  |       |        |       |
| Producteur x test            | 10,52 | 11,04  | 11,08 |
| 6. d. d. l.                  |       |        |       |
|                              |       |        |       |
| Indépendance conditionnelle  |       |        |       |
| Producteur x Résultat        | 13,29 | 12,81  | 13,74 |
| 6. d. d. l.                  |       |        |       |
|                              |       |        |       |
| Indépendance conditionnelle  |       |        |       |
| Test x Résultat              | 7,03  | 7,56   | 7,59  |
| 4. d. d. l.                  |       |        |       |
|                              |       |        |       |
| Indépendance                 |       |        |       |
| (Producteur x Test) Résultat | 13,31 | 12,84  | 13,76 |
| 7. d. d. l.                  |       |        |       |
|                              |       |        |       |
| Indépendance                 |       |        |       |
| (Producteur x Résultat) Test | 10,55 | 11,05  | 11,10 |
| 7. d. d. l.                  |       |        |       |
|                              |       |        |       |
| Indépendance                 |       |        |       |
| (Test x Résultat) Producteur | 16,80 | 16,76  | 17,25 |
| 9. d. d. l.                  |       |        |       |

Aucune de ces valeurs numériques n'indique une liaison significative entre les caractères. Elles sont toutes inférieures aux seuils de 5 % de tables de  $\chi^2$  ayant le même nombre de degrés de liberté, auxquels il faut les comparer.

Les données proposées ne permettent pas de distinguer un test pour sa sélectivité, ou un producteur en fonction des résultats de ses produits

L'interaction donnée entre parenthèses dans la colonne des test du  $\chi^2$  correspond à la formule de Darroch pour les tables parfaites.

Il est à remarquer que les différences observées entre les résultats des trois modes de calcul n'atteignent jamais 10 % de la valeur la plus faible.

Ces différences n'affectent en rien les conclusions des tests autre résultat intéressant, elles ne sont pas toujours en faveur du même mode de calcul. Nous n'obtenons pas de mode de calcul par défaut ou par excès d'une certaine quantité théorique. Nous aurions plutôt des estimateurs aléatoires de cette quantité. Les approximations qui permettent de rapprocher les trois méthodes portent en effet sur des variables aléatoires, et il serait normal de retrouver ce caractère sur les différences numériques.

— Taux de mortalité de mouches mâles ou femelles en fonction d'une quantité d'insecticide absorbée.

Le sexe, l'état dans lequel se trouve la mouche après l'expérience, constituent deux caractères à deux modalités.

Les niveaux d'absorption de l'insecticide sont au nombre de douze. Le tableau est le suivant :

| Niveau<br>d'insecticide | Mâles    |        | Femelles |        |
|-------------------------|----------|--------|----------|--------|
|                         | Vivantes | mortes | Vivantes | mortes |
| 1                       | 17       | 40     | 46       | 6      |
| 2                       | 14       | 44     | 44       | 5      |
| 3                       | 19       | 42     | 48       | 5      |
| 4                       | 21       | 33     | 41       | 4      |
| 5                       | 9        | 39     | 68       | 8      |
| 6                       | 21       | 38     | 70       | 5      |
| 7                       | 19       | 40     | 56       | 4      |
| 8                       | 15       | 32     | 51       | 8      |
| 9                       | 20       | 35     | 73       | 9      |
| 10                      | 15       | 29     | 78       | 5      |
| 11                      | 12       | 19     | 69       | 2      |
| 12                      | 12       | 29     | 75       | 3      |

Ce sont toujours les mêmes hypothèses et sous-hypothèses qui font l'objet de calculs :

| Indépendance | Décomposition<br>de la variance | Test du<br>$\chi^2$ | Information |
|--------------|---------------------------------|---------------------|-------------|
| Totale       |                                 |                     |             |
| 3.d.d.l      | ① 621,14                        | 621,14              | 650,62      |

Indép.  
 Niveau x Mortalité  
 11. d.d.l.

621,16  
 (2) 28,76

28,76  
 29,46

Indép.  
 Niveau x Sexe  
 11 d.d.l.

36,54  
 36,54  
 36,88

Indép.  
 Sexe Mortalité  
 1 d.d.l.

551,36  
 551,36  
 593,58

Intéraction  
 11.d.d.l.

4,48  
 (26,99)

- 9,30  
 92  
 chp 95

Indép. Cond.  
 Niveau Mortalité  
 22. d.d.l.

33,24  
19,03  
20,16

Indép. Cond.  
 Niveau Sexe  
 22 d.d.l.

41,02  
27,56  
27,58

Indép. Cond.  
 Mortalité Sexe  
 12.d.d.l.

555,84  
 548,98  
 584,28

Indép.  
 (Niveau x Mortalité) Sexe  
 23 d.d.l.

621,16  
 - 28,76  
 -----  
 592,40  
 -112- 38

(1) - (2)

592,38  
 568,69  
 621,16

|                           |        |        |        |
|---------------------------|--------|--------|--------|
| Indép.                    |        |        |        |
| (Niveau x Sexe) Mortalité | 584,60 | 563,00 | 613,74 |
| 23 d.d.l.                 |        |        |        |
| Indép.                    |        |        |        |
| (Sexe x Mortalité) Niveau | 69,78  | 56,46  | 57,04  |
| 23 d.d.l.                 |        |        |        |

Les résultats soulignés d'un trait ne sont pas significatifs au seuil de 1 %, les résultats soulignés de deux traits ne sont pas significatifs à 5 %.

Dans cet exemple nous nous trouvons très loin de l'indépendance à cause de la forte liaison qui existe entre le sexe et le taux de mortalité de ces mouches.

|          | Vivantes | Mortes |     |
|----------|----------|--------|-----|
| Mâles    | 194      | 420    | 614 |
| Femelles | 719      | 64     | 783 |
|          | 913      | 484    |     |

Contingence sexe mortalité

Les mouches femelles ont une résistance notablement supérieure aux mâles.

Ce non respect des conditions nécessaires à l'équivalence des modes de calcul n'a pas de gros effets sur les indicateurs liés aux hypothèses d'indépendance simple. Là encore les écarts n'atteignent pas dix pour cent. Pour les hypothèses plus fines, telles que l'interaction ou les indépendances conditionnelles les écarts sont beaucoup plus importants et ils peuvent modifier la décision d'acceptation ou de rejet de ces hypothèses.

De toutes les hypothèses d'indépendance entre caractères ou regroupements de caractères, seule la dernière (sexe x mortalité) x Niveau donne une variance de 20 % supérieur au  $\chi^2$  et à l'information.

2210-1

Ceci s'explique par la manière de calculer cette variance, on l'obtient en effet par différence de deux variances très fortes (Indép; totale et Indép. sexe mortalité) et les écarts de 5 % que nous avons observé entre ces quantités et les informations, donnent par différence des écarts d'un ordre de grandeur supérieur.

Cet exemple confirme que la décomposition de la variance peut encore jouer un rôle d'indicateur d'association entre caractères pour des tables très éloignés de l'indépendance totale.

Après ces comparaisons nous pouvons dresser une table des appo  
respectifs des trois méthodes d'étude des tables de contingence ternaire.

Analyse de la  
Fonction Ecart

Tests et Information

Choix des contingences significatives qu  
méritent une décomposition.

Essai d'interprétation des  
associations significatives  
à l'aide de facteurs non  
correllés.

Simplification des calculs de  
variances partielles par  
l'utilisation d'une seule  
pondération

Procédures optimales pour tester des  
sous-hypothèses et simplifier la  
structure du tableau

Définition de nouvelles sous-hypothèse  
et de nouvelles densités à analyser pa  
décomposition.

Explication de la dépendance  
entre les  $\chi^2$  ou les informations  
en utilisant des sous-espaces  
indépendants

## VI - LE CAS PARTICULIER DE L'INDEPENDANCE CONDITIONNELLE

### VI 1 - Un modèle de prévision pour une contingence binaire

Dans certains tableaux de contingence on peut observer une dépendance très forte entre deux caractères et être tenté d'expliquer ce phénomène par la prise en compte d'une troisième variable K. Ainsi les types de dépenses et les modes d'habitation sont certainement très liés si on observe cette contingence sur une population variée. Pour un ensemble plus homogène, même catégorie socio professionnelle ou même niveau de revenu, on peut espérer une certaine indépendance entre ces deux variables.

Pour obtenir une explication d'une contingence (IJ) on est amené à étudier une table à trois entrées (IJK) pour vérifier que pour chaque modalité k de la variable explicative K il y a indépendance entre I et J.

Cette structure nous est connue, elle correspond à l'indépendance conditionnelle entre I et J si K.

Les tableaux fournissant une explication devront donc avoir les propriétés suivantes:

Une des contingences marginales sera significative. Cette même contingence conditionnée par la troisième variable cessera d'être significative.



Le rejet de l'hypothèse d'indépendance et l'acceptation de l'indépendance conditionnelle auront mis en évidence une possible cause d'une association significative.

Cette façon d'aborder l'étude d'une contingence ne pose plus de problèmes théoriques puisque nous disposons de tous les tests nécessaires. L'obstacle principal à l'étude d'une indépendance conditionnelle est certainement le recueil des données. Dans ce domaine de l'analyse de données, le travail porte souvent sur des chiffres qui n'ont pas été recueillis pour être soumis à ce type d'étude. Les statistiques courantes ne se présentent pas sous forme de tableaux à trois entrées, et il est difficile d'obtenir une ventilation selon les modalités de la variable que l'on suppose explicative. Il serait donc très intéressant d'approcher la structure d'indépendance conditionnelle en n'utilisant que les marges bidimensionnelles du tableau (IJK).

Ainsi au lieu de vérifier que chaque couche donne une répartition indépendante, on va vérifier que les marges qui font intervenir la profondeur du tableau sont compatibles avec une telle structure.

$$P_{ij} = P_{ik} P_{jk} / P_k$$

Nous supposons  $P_{ijk} = P_{ik} P_{jk} / P_k$

donc  $P_{ij} = \sum_k P_{ik} P_{jk} / P_k$  relation qui ne fait intervenir que les marges du tableau.

En utilisant cette hypothèse et les comptages marginaux on peut effectuer une prévision de la répartition sur  $I \times J$  et donc de la fréquence observée

Pour obtenir cette prévision on utilise les estimateurs du maximum de vraisemblance pour  $p_{IK}$ ,  $p_{JK}$  et  $p_K$

Le nombre d'observations prévu dans la case  $ij$  est donc :

$$\hat{n}_{ij} = \sum_k n_{ik} n_{jk} / n_k$$

Cette méthode de prévision est très intéressante car elle s'appuie sur l'hypothèse d'indépendance conditionnelle qui fournit une explication de la contingence  $I \times J$  et de plus elle n'utilise que des tableaux à un ou deux indices. C'est donc avec des tableaux que l'on rencontre couramment dans les résultats statistiques qu'il sera possible d'expliquer l'association  $IJ$  par la variable  $K$ .

A partir de  $n_{IK}$  et  $n_{JK}$  on calcule  $n_K$  et le tableau induit ; ensuite il s'agit de comparer cette prévision à l'observation réelle. Les trois méthodes rencontrées proposent une solution.

La comparaison des deux densités  $f_{ij} = \frac{n_{ij}}{N}$  et  $\hat{f}_{ij} = \frac{\hat{n}_{ij}}{N}$  fournit une information :

$$2 \text{ Inf} = 2N \sum_{ij} f_{ij} \text{ Log } \frac{f_{ij}}{\hat{f}_{ij}}$$

qui tend vers un  $\chi^2$  à  $(\text{card } I - 1) (\text{card } J - 1)$

degrès de liberté.

S. KULLBACK appelle cette information l'interaction  $I \times J$ . C'est effectivement le complément d'information apporté par la comparaison entre l'observation et la prévision liée à une hypothèse.

L'adéquation entre  $n_{IJ}$  et  $\hat{n}_{IJ}$  peut se faire en construisant un

$$\chi^2 = \sum \frac{(\text{Observé} - \text{Théorique})^2}{\text{Théorique}} = \sum_{ij} \frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}}$$

Ce  $\chi^2$  est équivalent à l'information précédente, la répartition  $\hat{n}_{ij}$  est soumise à des conditions de marge :

$$\sum_j \hat{n}_{ji} = n_i = \sum_j n_{ji} \quad \forall i \in I \quad \text{Card I conditions}$$

$$\sum_i \hat{n}_{ij} = n_j = \sum_i n_{ij} \quad \forall j \in J \quad \text{Card J conditions}$$

De cette façon on compte deux fois la condition  $\sum_{ij} n_{ij} = N$

d'où finalement le nombre de degrés de liberté ( $n^b$  de cases -  $n^b$  de conditions indépendantes:  $\text{card I} \times \text{card J} - \text{card I} - \text{card J} + 1 = (\text{card I} - 1) (\text{card J} - 1)$ )

Ces deux techniques permettent d'effectuer des tests et de prendre une décision quant à l'hypothèse supposée et au pouvoir explicatif de la variable K.

L'analyse factorielle peut apporter sa contribution à la comparaison des deux densités en mettant en évidence les facteurs résiduels qui expliquent la différence ( $f_{IJ} - \hat{f}_{IJ}$ ).

On cherchera des facteurs orthonormés sur I et J dans les métriques  $f_I$  et  $f_J$ , la décomposition portera donc sur :

$$\frac{f_{ij} - \hat{f}_{ij}}{f_i f_j} = \sum_{\alpha} \lambda_{\alpha} \phi_{\alpha}^i \phi_{\alpha}^j$$

Si les facteurs  $\phi_{\alpha}$  révèlent des associations interprétables, la variable K ne pourra pas être considérée comme la seule explication de la contingence (IJ).

La variance de cet écart résiduel, pour la densité produit, diffère légèrement du  $\chi^2$  permettant le test d'adéquation et de l'information équivalente

$$N \sum_{ij} \left( \frac{f_{ij} - \hat{f}_{ij}}{f_i f_j} \right)^2 f_i f_j = N \sum_{\alpha} \lambda_{\alpha}^2 \neq N \sum_{ij} \left( \frac{f_{ij} - \hat{f}_{ij}}{\hat{f}_{ij}} \right)^2 \hat{f}_{ij} = \chi^2$$

La différence provient de l'utilisation de la densité  $f_I \times f_J$  correspondant à l'indépendance entre I et J au lieu de  $\hat{f}_{IJ}$  associée à l'indépendance conditionnelle; La variance de cette nouvelle fonction d'écart ne permet pas d'effectuer un test correct. Elle traduira simplement l'ampleur des écarts entre la prévision et l'observation. On pourra avec profit la comparer au  $\chi^2$  de contingence entre I et J pour vérifier que  $f_{IJ}$  est plus proche de  $\hat{f}_{IJ}$  que de  $f_I f_J$ .

## VI2 - Analyse factorielle conditionnelle

L'opération de prévision que nous venons de définir et qui consiste à construire une dépendance entre I et J, à l'aide des dépendances observées entre I et K et J et K, pour ensuite la comparer à la contingence réellement observée, s'illustre et s'interprète en revenant aux décompositions factorielles du premier chapitre.

Nous disposons de la décomposition de l'écart total

$$\frac{f_{IJK}}{f_I f_J f_K} - 1 = \sum_{p,m,n} \lambda_{p,m,n} \phi_p^I \phi_m^J \phi_n^K \quad \text{avec } \phi_0 = 1$$

qui se scindait en quatre parties dont trois fournissaient les écarts marginaux

$$\frac{f_{IJ}}{f_I f_J} - 1 = \sum_{p,m} \lambda_{p,m,0} \phi_p^I \phi_m^J$$

$$\frac{f_{IK}}{f_I f_K} - 1 = \sum_{p,n} \lambda_{p,0,n} \phi_p^I \phi_n^K$$

$$\frac{f_{JK}}{f_J f_K} - 1 = \sum_{m,n} \lambda_{0,m,n} \phi_m^J \phi_n^K$$

Nous voulons construire la densité :

$$\hat{f}_{IJ} = \sum_K \frac{f_{IK} f_{JK}}{f_K}$$

ou ce qui sera équivalent pour les calculs suivants, l'écart :

$$\begin{aligned} \frac{\hat{f}_{IJ}}{f_I f_J} - 1 &= \sum_K \frac{f_{IK} f_{JK}}{f_I f_J f_K} - 1 = \sum_K \left( \frac{f_{IK}}{f_I f_K} - 1 \right) \left( \frac{f_{JK}}{f_J f_K} - 1 \right) f_K \\ &= \sum_K \left( \sum_{p,n} \lambda_{p,0,n} \phi_p^I \phi_n^K \right) \left( \sum_{m,n'} \lambda_{0,m,n'} \phi_m^J \phi_{n'}^K \right) f_K \\ &= \sum_{p,m} \phi_p^I \phi_m^J \sum_{n,n'} \lambda_{p,0,n} \lambda_{0,m,n'} \left( \sum_K \phi_n^K \phi_{n'}^K f_K \right) \end{aligned}$$

et en vertu de l'orthonormalité des facteurs de la décomposition

$$\frac{\hat{f}_{IJ}}{f_I f_J} - 1 = \sum_{\substack{p,m,n \\ \neq 0}} \lambda_{p,0,n} \lambda_{0,m,n} \phi_p^I \phi_m^J$$

Lorsque certains facteurs sur I ou J sont absents des décompositions des écarts sur IK et JK les coefficients  $\lambda$  sont nuls.

La densité induite sur IJ par l'opération de prévision est donc constituée de couples de facteurs  $(p,m)$ , ayant la propriété d'apparaître dans les contingences IK et JK, associés à un même facteur sur K. C'est en effet la condition nécessaire pour que le coefficient  $\lambda_{pon} \lambda_{omn}$  ne soit pas nul. La dépendance prévisible sur I x J a été construite à l'aide du sous-ensemble de la base des fonctions sur k défini par

$$n^* = \{ n \mid \langle \Phi_n^k, f_{IK} \rangle \neq 0, \text{ et } \langle \Phi_n^k, f_{JK} \rangle \neq 0$$

L'information transmise par les marges  $f_{IK}$  et  $f_{JK}$  se fait par l'intermédiaire des facteurs sur K, le coefficient d'une association  $(p,m)$  est  $\sum_n \lambda_{pon} \lambda_{omn}$ . Par une transition de I vers K on associe à  $\mathcal{E}_I$  un ensemble d'indices n tels que :  $\lambda_{pon} \neq 0$  et par une nouvelle transition de K vers J, se trouve associé aux indices m tels que :  $\lambda_{omn} \neq 0$  pour au moins un n de l'ensemble précédent. Ces transitions associent les facteurs sur deux caractères différents, apparaissant dans un même triplet dans la décomposition avec des facteurs répétitifs.

Cette méthode ne faisant apparaître que des contingences binaires on peut utiliser les décompositions de l'analyse factorielle des correspondances

$$\frac{f_{IK}}{f_I f_K} - 1 = \sum_{\alpha} \lambda_{\alpha} \Phi_{\alpha}^I \Phi_{\alpha}^K \quad ; \quad \frac{f_{JK}}{f_J f_K} - 1 = \sum_{\beta} \lambda_{\beta} \Phi_{\beta}^J \Phi_{\beta}^K$$

$$\left\{ \begin{aligned} \frac{\hat{f}_{IJ}}{f_I f_J} - 1 &= \sum_{\alpha, \beta} \lambda_{\alpha} \lambda_{\beta} \Phi_{\alpha}^I \Phi_{\beta}^J \text{ corr}(\Phi_{\alpha}^K, \Phi_{\beta}^K) \quad \text{ou } \omega_i \\ &= \sum_n \frac{(b_{In} - 1)(b_{Jn} - 1)}{b_{In} b_{Jn}} b_n \\ &= \sum_{\alpha, \beta} \lambda_{\alpha} \lambda_{\beta} \Phi_{\alpha}^I \Phi_{\beta}^J \sum_n b_n \Phi_{\alpha}^K \Phi_{\beta}^K \end{aligned} \right.$$

un couple  $(\alpha, \beta)$  apparait avec un coefficient dépendant de la corrélation entre les facteurs associés sur K.

L'écart résiduel qu'il est souhaitable de décomposer devient :

$$\frac{f_{I_2} - \hat{f}_{I_2}}{f_I f_2} = \left( \frac{f_{I_2}}{f_I f_2} - 1 \right) - \left( \frac{\hat{f}_{I_2}}{f_I f_2} - 1 \right) = \sum_{\rho, m} \left( \lambda_{\rho m 0} - \sum_n \lambda_{\rho 0 n} \lambda_{0 m n} \right) \phi_e^I \phi$$

Le coefficient d'amplitude lié au couple  $(\rho, m)$ ,  $(\lambda_{\rho m 0} - \sum_n \lambda_{\rho 0 n} \lambda_{0 m n})$  est donc le coefficient initial :  $\lambda_{\rho m}$  "diminué" de l'importance prévisible, par l'intermédiaire de  $\{ \phi_n^k \mid n \in N \}$  de l'association  $(\rho, m)$ .

Cette forme algébrique du coefficient est à rapprocher de celle du coefficient de corrélation partielle de deux populations normales.

L'analogie de forme se complète par l'identité des objectifs de ces deux méthodes, il s'agit en effet de tenir compte de l'information apportée par une troisième variable pour expliquer la liaison entre les deux variables étudiées.

Nous appellerons analyse factorielle conditionnelle l'opération qui consiste à décomposer l'écart, entre l'observation  $f_{IJ}$  et la prévision  $\hat{f}_{IJ}$ , rapporté à la densité  $f_I f_J$ .

Bien qu'il s'agisse d'une nouvelle décomposition, cette opération pourra mettre en évidence les facteurs d'association ne provenant pas de la variable k et s'ils restent importants,  $(\sum_n \lambda_{\rho 0 n} \lambda_{0 m n} \ll \lambda_{\rho m})$  les résidus des facteurs rencontrés dans les contingences IK et JK.

Remarquons que les variances des nouvelles répartition sont liées par les inégalités suivantes :

$$\begin{aligned} \frac{\hat{\chi}_{IJ}^2}{N} &= \sum_{IJ} \left( \frac{\hat{f}_{IJ}}{f_I f_J} - 1 \right)^2 f_I f_J = \sum_{IJ} \left( \sum_k \left( \frac{f_{Ik}}{f_I f_k} - 1 \right) \left( \frac{f_{Jk}}{f_J f_k} - 1 \right) f_k \right)^2 f_I f_J \\ &= \sum_{IJ} f_I f_J (\langle h_I, h_J \rangle)^2 \leq \sum_{IJ} f_I f_J \|h_I\|^2 \times \|h_J\|^2 \end{aligned}$$

$$\frac{\hat{\chi}_{IJ}^2}{N} \leq \sum_{IJ} f_I f_J \sum_k \left( \frac{f_{Ik}}{f_I f_k} - 1 \right)^2 f_k \times \sum_k \left( \frac{f_{Jk}}{f_J f_k} - 1 \right)^2 f_k$$

$$\Rightarrow \frac{\chi_{IJ}^2}{N} = \frac{\hat{\chi}_{IJ}^2}{N} \leq \frac{\chi_{Ik}^2}{N} \times \frac{\chi_{Jk}^2}{N}$$

Ainsi après l'étude des contingences IK et JK on obtient par simple multiplication un ordre de grandeur de la variance de la densité induite. La comparaison de  $\chi_{IJ}^2/N$  et de  $\chi_{Ik}^2 \times \chi_{Jk}^2/N^2$  permet tout de suite de savoir si la comparaison de  $f_{IJ}$  et  $\hat{f}_{IJ}$  pourra donner des résultats intéressants

Pour la variance résiduelle on a en effet :

$$\begin{aligned} \frac{\chi_R^2}{N} &= \sum_{IJ} \left( \frac{f_{IJ} - \hat{f}_{IJ}}{f_I f_J} \right)^2 f_I f_J = \sum_{IJ} \left( \left( \frac{f_{IJ}}{f_I f_J} - 1 \right) - \left( \frac{\hat{f}_{IJ}}{f_I f_J} - 1 \right) \right)^2 f_I f_J \\ \frac{\chi_R^2}{N} &= \|g_1 - g_2\|^2 \geq (\|g_1\| - \|g_2\|)^2 \\ \sqrt{\frac{\chi_R^2}{N}} &\geq \left| \sqrt{\frac{\chi_{IJ}^2}{N}} - \sqrt{\frac{\hat{\chi}_{IJ}^2}{N}} \right| \end{aligned}$$

Cette inégalité permet de prévoir la plus faible variance résiduelle que l'on peut espérer.

» d



$$\chi^2_{IJ} \leq \frac{(\chi^2_{IK} \chi^2_{JK})}{N} = \text{quelq. } \chi^2_{IJ}$$

Ainsi, si la quantité  $\frac{\chi^2_{IK} \times \chi^2_{JK}}{N}$  ne représente que

quelque pour cent de  $\chi^2_{IJ}$  il ne sera pas nécessaire de chercher dans la variable K, une explication de l'association I x J.

### VI3 - Extension à plus de trois caractères

Lorsqu'une seule variable ne s'avère pas suffisante pour expliquer une contingence on peut envisager l'étude de tableaux de données plus riches qui feraient intervenir plusieurs variables explicatives.

Dans un tableau de contingence entre quatre caractères on aura donc vérifié que les marges ternaires  $f_{ijk}$  ne correspondent pas à l'indépendance conditionnelle entre i et j si k. Mais ces marges peuvent être la somme de deux indépendances définies par le quatrième indice.

Nous aurions alors :

$$P_{iik} = \frac{P_{ik} P_{ik}}{P_k} \quad \forall k ; \quad \exists k : P_{iik} \neq \frac{P_{ik} P_{ik}}{P_k}$$

Le conditionnement devrait donc se faire par rapport au produit de caractères K x L. Cette hypothèse est analogue à l'indépendance conditionnelle entre deux variables normales dans une population multidimensionnelle.

L'étude théorique de cette hypothèse ne rencontre pas d'obstacle mais pratiquement il sera très difficile de recueillir et de manipuler les tableaux IKL et JKL.

Il faut donc ajouter des hypothèses sur les contingences KL o IKL et JKL. L'hypothèse la moins contraignante ne permettant d'utiliser que de densités à deux indices est l'indépendance conditionnelle entre K et L d'où :

$$\forall i,j \quad P_{iKl} = \frac{P_{iK} P_{lP}}{P_i} \quad , \quad P_{iKl} = \frac{P_{iK} P_{lP}}{P_i}$$

Alors :

$$P_{iKlP} = \frac{P_{iK} P_{lP} P_{iK} P_{lP}}{P_i P_j P_{Kl}} \quad \text{et} \quad \hat{P}_{ij} = \sum_{Kl} \frac{P_{iK} P_{lP} P_{iK} P_{lP}}{P_i P_j P_{Kl}}$$

Une prévision est donc possible en ne considérant que des coup de caractères et des tableaux à deux entrées.

Une hypothèse un peu plus forte permet de simplifier le calcul de la densité induite.

Supposons que les caractères K et L soient indépendants pour la population étudiée. Les hypothèses

$$P_{iKl} = \frac{P_{iK} P_{lP}}{P_i} \quad \text{et} \quad P_{iKlP} = \frac{P_{iK} P_{lP}}{P_i} \quad \forall i,j$$

restent valables e

nous avons en plus :  $p_{kl} = p_k p_l$  d'où la densité induite :

$$\hat{P}_{ij} = \sum_K \frac{P_{iK} P_{jP}}{P_K} \times \sum_P \frac{P_{iP} P_{jP}}{P_P} \times \frac{1}{P_i P_j}$$

Ce qui peut se résumer ainsi en mettant en indice haut l'ensembl ou les ensembles indépendants permettant la prévision

$$\hat{P}_{ij}^{KL} = \hat{P}_{ij}^k \times \hat{P}_{ij}^L \times \frac{1}{P_i P_j}$$

Remarquons que si L se réduit à un élément :

$$\left( \hat{P}_{ij}^L = \frac{P_{i1} P_{j1}}{P_1} = P_i P_j \right) \quad \text{et nous retrouvons} \quad \hat{P}_{ij}^{KL} = \hat{P}_{ij}^k$$

Comme dans les études de correspondances ternaires, lorsque deux ensembles sont indépendants on peut, sans rien perdre de l'information fournie par les données remplacer le produit des ensembles KL (toutes les associations kl) par la somme ou juxtaposition K + L (les card K + card L modalités k ou l). La prévision s'effectue en utilisant les deux tableaux  $(p_{IK}, p_{IL})$  et  $(p_{JK}, p_{JL})$  de dimension card I ou card J par (card K + card L) et non plus  $p_{IKL}$  et  $p_{JKL}$ .

Ce modèle de prévision peut s'utiliser de façon itérative. Imaginons que la densité estimée pour  $\hat{P}_{ij}^k$  ou pour  $\hat{P}_{ij}^L$  ne soit pas jugée suffisante pour approcher l'observation, c'est-à-dire que le caractère K ou le caractère L ne peut pas expliquer l'association IJ. Alors sous réserve d'avoir vérifié l'indépendance KL, on peut, par simple multiplication des prévisions, calculer  $\hat{P}_{ij}^{KL}$ . Et essayer ainsi de se rapprocher de l'observation en utilisant les deux caractères K et L.

La formule de prévision peut se généraliser à un ensemble  $K = \{k_1, \dots, k_n\}$  de caractères explicatifs indépendants :

$$\left\{ \left( \frac{\hat{P}_{ij}^K}{P_i \cdot P_j} \right) = \prod_K \left( \frac{\hat{P}_{ij}^k}{P_i P_j} \right) \right.$$

En n'introduisant dans  $\mathbb{K}$  que des caractères  $K$  donnant une variance résiduelle  $\chi_R^2$  plus faible que la variance de l'observation  $\chi_{I_1}^2$  on peut s'attendre lorsque l'effectif de  $\mathbb{K}$  augmente, à obtenir une estimation de plus en plus précise de la densité observée.

## VII - EXEMPLES DE MODELES DE PREVISION ET D'ANALYSES CONDITIONNELLES

Les données des deux premiers exemples proviennent de la publication du Service d'Etude Technique des Routes et Autoroutes intitulée : Accident corporels de la circulation routière 1969. L'étude est une tentative d'explication de certaines variations de gravité des accidents, par le lieu où ils se sont produits.

L'observation élémentaire des tables de contingence est donc la victime d'un accident, les trois caractères étudiés sont pour le premier exemple

La gravité - Les trois modalités sont les tués (TUS sur les graphiques), les blessés graves (BLG) et les blessés légers (BLL).

Le lieu - Quatre modalités traduisent le degrés d'urbanisation Hors agglomération (HAG) ; Petite agglomération (moins de 2000 habitants) (PAG) Moyenne agglomération (de 2000 à 5000 habitants) (MAG) ; et au-dessus de 5000 habitants les grosses agglomérations (GAG).

L'époque - Ses douze modalités sont les mois de l'année 1969.

Dans cette contingence construite avec près de 326000 observations il s'agira donc d'expliquer la part de la variable lieu dans les fluctuations de gravité qui apparaissent entre les différents mois.

La contingence à expliquer est donc la contingence (gravité x époque). Nous en avons effectuée une décomposition factorielle pour traduire par des associations (niveau de gravité, mois) la variance significative qui ressort du tableau de données.

La variance de la table de fréquence est égale à :

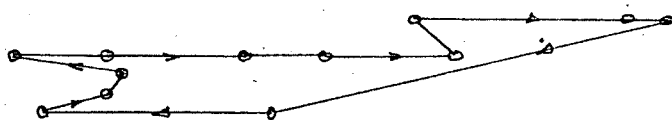
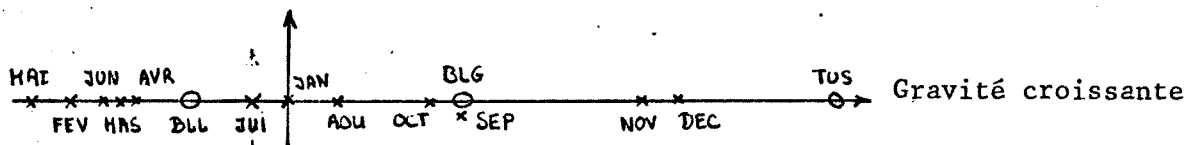
$$\frac{\chi^2}{N} = 4,26 \cdot 10^{-4}$$

Le  $\chi^2$  de contingence est donc  $4,26 \cdot 10^{-4} \times 326 \cdot 10^3 = 138$  pour  $(12-1)(3-1) = 22$  degrés de liberté. Il est hautement significatif.

De cette table de contingence on peut extraire deux facteurs, ils traduisent respectivement 87 % et 13 % de la variance totale. Il sera suffisant de prendre en compte le premier facteur, en effet le second est associé à une variance  $\chi_2^2 = \frac{13}{100} \times 138 = 17$  qui peut être considérée comme une perturbation aléatoire.

$$4,26 \times 0,87 =$$

Le premier axe est une échelle de gravité, mais le facteur sur les mois ne donne pas une classification bien ordonnée.



Ordre chronologique des mois de l'année étudiée.

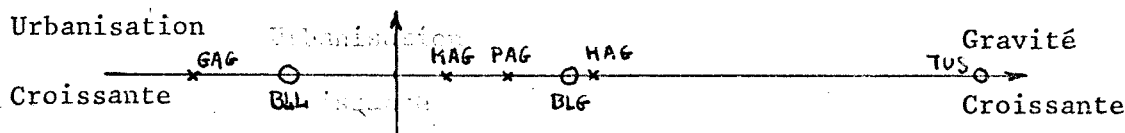
Pour comprendre l'échelle de gravité que l'on vient d'obtenir on va s'intéresser aux deux liaisons gravité-lieu et Lieu-Epoque.

La contingence, Gravité, Lieu montre une très forte liaison pour un  $\chi^2$  à 6 degrés de liberté nous obtenons la valeur numérique 8900.

La variance du tableau de fréquences est égale à  $2,72 \cdot 10^{-2}$ .

Le premier axe est constitué par l'échelle de gravité croissant de la contingence précédente.

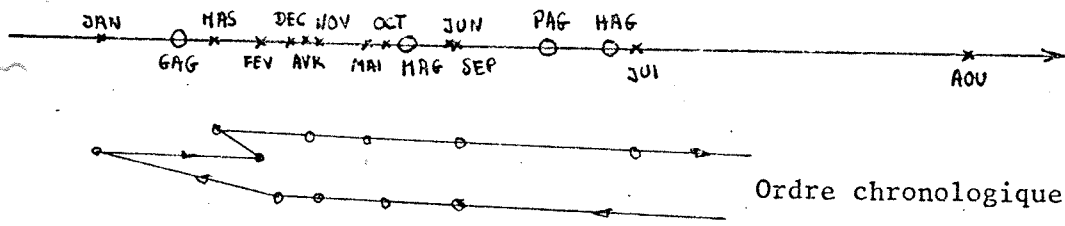
Les associations (GAG, BLL) et (HAG, TUS) sont évidentes.



La troisième marge du tableau tridimensionnel est obtenue en cumulant les différentes gravités à une intersection (mois, lieu). On a donc pu le nombre total de victimes d'accident de la route correspondant à ces deux modalités.

La trace du tableau de fréquence est de  $1,06 \cdot 10^{-2}$  et le  $\chi^2$  égal à 3460 pour 33 degrés de liberté, on doit trouver dans ce tableau de très fortes associations. En effet, sur le premier axe l'échelle d'urbanisation croissante de la contingence précédente permet de séparer les quatre mois de vacances ou de grands week-ends AOU, JUI, JUN et SEP, qui sont associés aux modalités HAG (hors agglomération) et PAG (petite agglomération) de la variable lieu.

lieu  
epoque



Connaissant le lieu où se produisent les accidents à certaines époques de l'année, et le niveau de gravité associé à ce lieu il doit être possible à l'aide des deux contingences précédentes d'effectuer une prévision sur la contingence Gravité-Epoque.

Si les automobilistes du mois d'août ont principalement leur accident hors agglomération, et que cette modalité de la variable lieu se révèle la plus dangereuse, on peut prévoir que le mois d'août sera l'un des plus dangereux de l'année. Le risque encouru par l'automobiliste du mois d'août pourra s'expliquer par l'endroit où il roule. L'endroit n'étant, bien sûr, qu'une variable résultante recouvrant des causes plus importantes mais moins abordables (vitesse, fatigue, durée du trajet...).

Ce transport d'information de deux marges à une troisième correspond à l'hypothèse d'indépendance conditionnelle entre la gravité et l'époque.

Si la contingence observée entre la gravité et l'époque n'était due qu'au lieu où surviennent des accidents nous pourrions estimer les probabilités des associations Gravité-Epoque à l'aide des estimateurs des probabilités, (Epoque-Lieu) et (Lieu Gravité).



Les deux décompositions marginales servant à la prévision ont donné :

I = {Epoque}                      J = {Gravité}                      K = {Lieu}

$$\frac{f_{IK}}{f_I f_K} - 1 = \alpha \phi^I \phi^K + \varepsilon_{IK}$$

$\phi^I$  traduisait la dichotomie Vacances Hiver

$\phi^K$  Les niveaux d'urbanisation,  $\varepsilon_{IK}$  peut être considéré comme une perturbation aléatoire.

$$\frac{f_{JK}}{f_J f_K} - 1 = \beta \phi^J \phi^K + \varepsilon_{JK}$$

Nous conservons le même  $\phi^K$  puisque les échelles d'urbanisation étaient très semblables.

$\phi^J$  donne les niveaux de gravité des accidents.

$\varepsilon_{JK}$  est considéré comme aléatoire.

Si l'association (I x J) est causée par le caractère K la prévision est :

$$\hat{f}_{IJ} = \sum_K \frac{f_{IK} f_{JK}}{f_K}$$

elle donne l'écart suivant

$$\begin{aligned} \frac{\hat{f}_{IJ}}{f_I f_J} - 1 &= \sum_K \left( \frac{f_{IK}}{f_I f_K} - 1 \right) \left( \frac{f_{JK}}{f_J f_K} - 1 \right) f_K \\ &= \sum_K (\alpha \phi^I \phi^K + \varepsilon) (\beta \phi^J \phi^K + \varepsilon') f_K \quad ; \quad \sum_K \phi^K^2 f_K = 1 \\ \frac{\hat{f}_{IJ}}{f_I f_J} - 1 &= \alpha \beta \phi^I \phi^J + \varepsilon'' \end{aligned}$$

(L, K) → (J, T)  
 (L, K)

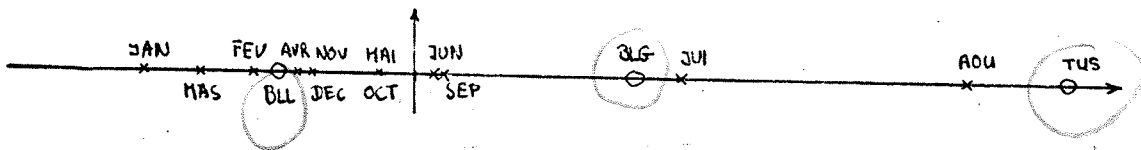
On associe les deux facteurs  $\phi^I$  et  $\phi^J$  qui correspondaient à la même échelle d'urbanisation  $\phi^K$ . Les fluctuations résiduelles  $\varepsilon$  et  $\varepsilon'$  créent des variances très faibles et non significatives,  $\varepsilon''$  aura certainement un effet du même ordre.

Le tableau de fréquences induit en effet une variance égale à  $2,77 \cdot 10^{-4} \approx 1,06 \cdot 10^{-2} \times 2,72 \cdot 10^{-2}$ , la variance associée au seul facteur significatif ( $\alpha, \beta$ ) est assimilable au produit des variances des tableaux de fréquences servant à la prévision.

Le  $\chi^2$  provenant de ce tableau est égal à 91 pour 22 degrés de liberté, on a donc mis en évidence des variations de gravité comparables à celle de la première marge (époque-gravité) dont le  $\chi^2$  était de 138.

Le calcul de  $\hat{n}_{ij} = \sum_k n_{ik} n_{jk} / n_k$  et la décomposition factorielle permettent bien de retrouver le facteur sur les mois de contingence (époque x lieu), et justifient donc les approximations effectuées.

*pour les mois par ailleurs mes (qualifier par 130)*



Cette contingence induite est à comparer à la face époque-gravité. Le facteur sur la gravité est sensiblement le même que la première décomposition par contre le classement des mois selon cette échelle de gravité est notablement différent.

Nous avons un résultat qui peut se résumer ainsi :

$$\frac{f_{I1}}{f_I f_1} - 1 \approx \lambda \phi_0^I \phi_0^J + \varepsilon_{I1} \quad \boxed{\lambda^2 = 4,26 \cdot 10^{-4} \times \frac{87}{100} \approx 3,7 \cdot 10^{-4}}$$

et nous avons trouvé la prévision suivante:

avec  $\frac{f_{I1}}{f_I f_1} - 1 = \alpha \beta \phi_1^I \phi_1^J + \varepsilon_{I1}'' \quad (\alpha \beta)^2 = 2,7 \cdot 10^{-4}$

$$\phi_1^I \neq \phi_0^I \quad \phi_1^J = \phi_0^J$$

La différence entre les variances et les classements des mois confirme que le lieu de l'accident n'est pas un élément de prévision suffisant pour l'association entre les mois et les niveaux de gravité.

Pour juger de l'importance et de la nature des autres facteurs agissant on peut comparer l'observation (époque-gravité) à la prévision, c'est-à-dire les classements  $\Phi_0^I$  et  $\Phi_1^I$ . Pour tenir compte de l'influence quantitative des causes résiduelles il vaudra mieux étudier  $(\lambda \Phi_0^I - \alpha\beta \Phi_1^I) = \text{Influence totale} - \text{Influence du lieu}$ .

Cette différence est aussi le résultat de la décomposition factorielle de  $f_{IJ} - \hat{f}_{IJ}$  ; en effet :

$$\begin{aligned} \frac{f_{IJ} - \hat{f}_{IJ}}{f_I f_J} &= \lambda \Phi_0^I \Phi_0^J - \alpha\beta \Phi_1^I \Phi_1^J + \varepsilon_{IJ} \\ &= \Phi_0^J (\lambda \Phi_0^I - \alpha\beta \Phi_1^I) + \varepsilon_{IJ} \\ &= \gamma \Phi_2^I \Phi_0^J + \varepsilon_{IJ} \end{aligned}$$

$\gamma$  pouvant prendre des valeurs comprises entre  $|\lambda - \alpha\beta|$  et  $(\lambda + \alpha\beta)$   
Selon la corrélation qui existe entre les deux facteurs sur les mois.

Pour savoir si le lieu explique une partie de la variance (époque-gravité) ou si au contraire le classement qu'il permet de prévoir sur les mois est opposé à celui défini par l'observation, on étudie la table

$$n_{IJ} - \hat{n}_{IJ} = n_{IJ} - \sum_k \frac{n_{Ik} n_{Jk}}{n_k}$$

Cette table est de somme nulle, mais la décomposition se fait sans problème par rapport à une densité constamment nulle.

Le tableau résiduel des différences de fréquences à une variance égale à  $6,98 \cdot 10^{-4}$ , ce qui donne un  $\chi^2$  ou une information proche de 230 pour 22 degrés de liberté.  $\chi^2 = 230$

Cette variance est bien plus forte que celle du premier tableau étudié. Les bornes extrêmes qu'elle peut atteindre sont :

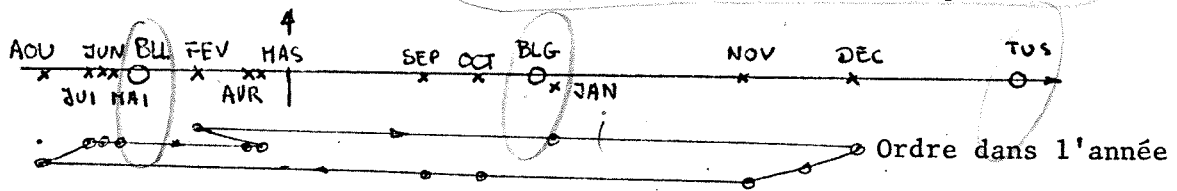
$$(\lambda + \alpha\beta)^2 = (3,7 + 2,77 + 2 \times 3,16) \cdot 10^{-4} = 12,77 \cdot 10^{-4}$$

*(Handwritten:  $\chi^2$  and  $\alpha\beta$  under the terms)*

$$(\lambda - \alpha\beta)^2 = (3,7 - 2,7 - 2 \times 3,16) \cdot 10^{-4} = 0,1 \cdot 10^{-4}$$

Selon que les classements sur les mois sont en opposition ou en accord parfait dans l'observation et dans la prévision.

Le graphique résultant de la décomposition de  $f_{12} - \hat{f}_{12}$ , donne un classement des mois aussi homogène que celui induit par le lieu



Ce facteur résiduel  $\gamma \Phi_2^I$  oppose les mois d'hiver et les mois d'été. Ce classement correspond à l'ensoleillement. (Le coefficient de corrélation des rangs de Sheppard  $r_s = 0,895$  est significatif à 1%), la cause résiduelle des variations de gravité semble donc être les conditions atmosphériques.

*avec que*  
*avec*  
*l'ensemble*

Le niveau de gravité que l'on peut attribuer à un mois résulte donc de deux effets opposés :

Le lieu de circulation qui rend les mois de vacances et donc d'été plus dangereux.

Les conditions atmosphériques de circulation qui compensent largement l'effet précédent (variance  $7 \cdot 10^{-4}$  contre  $2,7 \cdot 10^{-4}$ ) et qui font que dans la contingence observée gravité\*époque, novembre et décembre sont les mois les plus dangereux alors qu'août se trouve à la moyenne générale des mois de l'année.

Nous pouvons écrire en résumé :

$$\lambda \phi_0^I = \alpha\beta \phi_1^I + \gamma \phi_2^I$$

effet général
effet du lieu
effet des conditions atmosphériques

$$(\alpha\beta)^2 = 2,7 \cdot 10^{-4} \ll \gamma^2 = 7 \cdot 10^{-4}$$

?

$$\text{corrélation } (\phi_1^I, \phi_2^I) = \frac{\lambda^2 - \alpha^2\beta^2 - \gamma^2}{2\alpha\beta\gamma} = -0,7$$

Les effets des deux caractères explicatifs sont presque opposés

Le rapport  $\gamma/\alpha\beta = \sqrt{2,5} = 1,6$  peut servir à chiffrer l'influence relative générale des deux effets. Cette étude des variations de gravité amène à penser que les variations dues aux conditions atmosphériques sont de 60 % supérieures à celles dues au lieu de l'accident.

Le second exemple concerne la contingence ternaire :

(Région x lieu x gravité)

Le premier caractère à 22 modalités (la Corse est considérée comme une région).

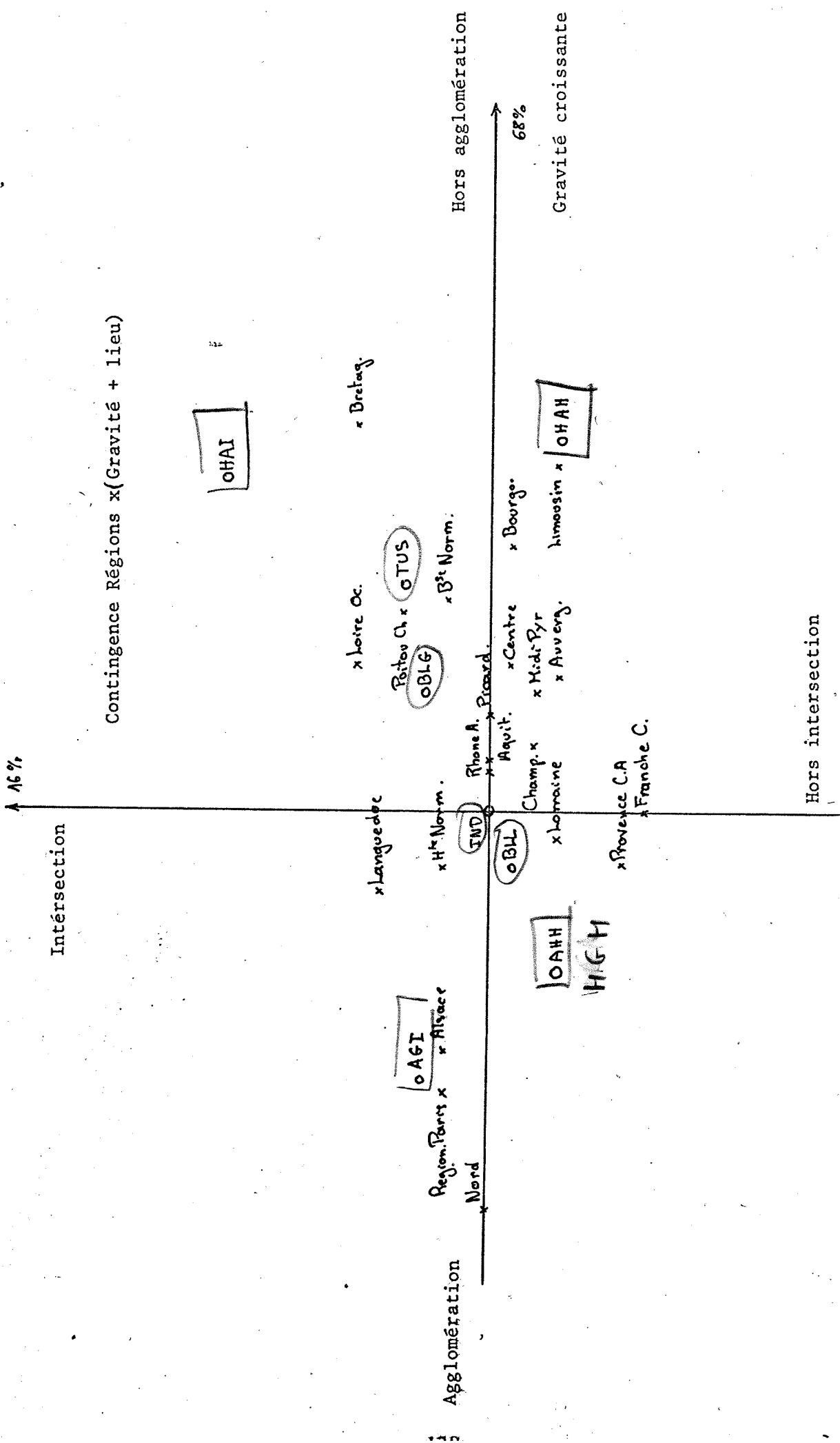
Le lieu a une autre définition que dans l'exemple précédent. Se quatre modalités sont maintenant :

|       |                                    |
|-------|------------------------------------|
| - AGI | agglomération et intersection      |
| HGH   | - - - - - et hors intersection     |
| - HAI | hors agglomération et intersection |
| - HAH | - - - - - et hors intersection     |

Pour la gravité, il y a une nouvelle modalité (IND), les personnes indemmes. La population étudiée n'est plus celle des victimes mais celle des personnes impliquées (611600).

Pour préciser les difficultés que présente une telle contingence où tous les caractères sont dépendants nous avons effectué deux analyses de corrélations ternaires.

La première simplifie le tableau de données. Elle décompose la juxtaposition des deux marges où intervient le caractère "région".



Dans cette contingence Région x (gravité + Lieu) la contribution à la variance totale du caractère gravité (15 %) est bien inférieure à celle du lieu (85 %). La marge Région x Gravité est absorbée par la marge Région x Lieu. On peut vérifier cette propriété en comparant le troisième graphique où la contingence Région-Lieu donne exactement les mêmes classements pour les régions sur les deux axes.

La seconde utilise l'ensemble produit (Gravité x Lieu), il y a maintenant 16 modalités. A un lieu est associé quatre niveaux de gravité. La décomposition donne toujours les mêmes classements pour les régions, les points associés aux modalités du lieu sont simplement décomposés en quatre points voisins correspondant aux quatre niveaux de gravité. Cette contingence utilise toutes les données du tableau ternaire mais sa décomposition n'explique pas directement l'effet du lieu sur l'indice de gravité affecté à chaque région.

Les trois graphiques suivants concernent les contingences marginales de la table tridimensionnelle.

Le tableau Région-gravité donne un premier facteur très significatif sur les régions, il semble opposer les régions méridionales et les régions septentrionales, les premières étant les moins dangereuses.

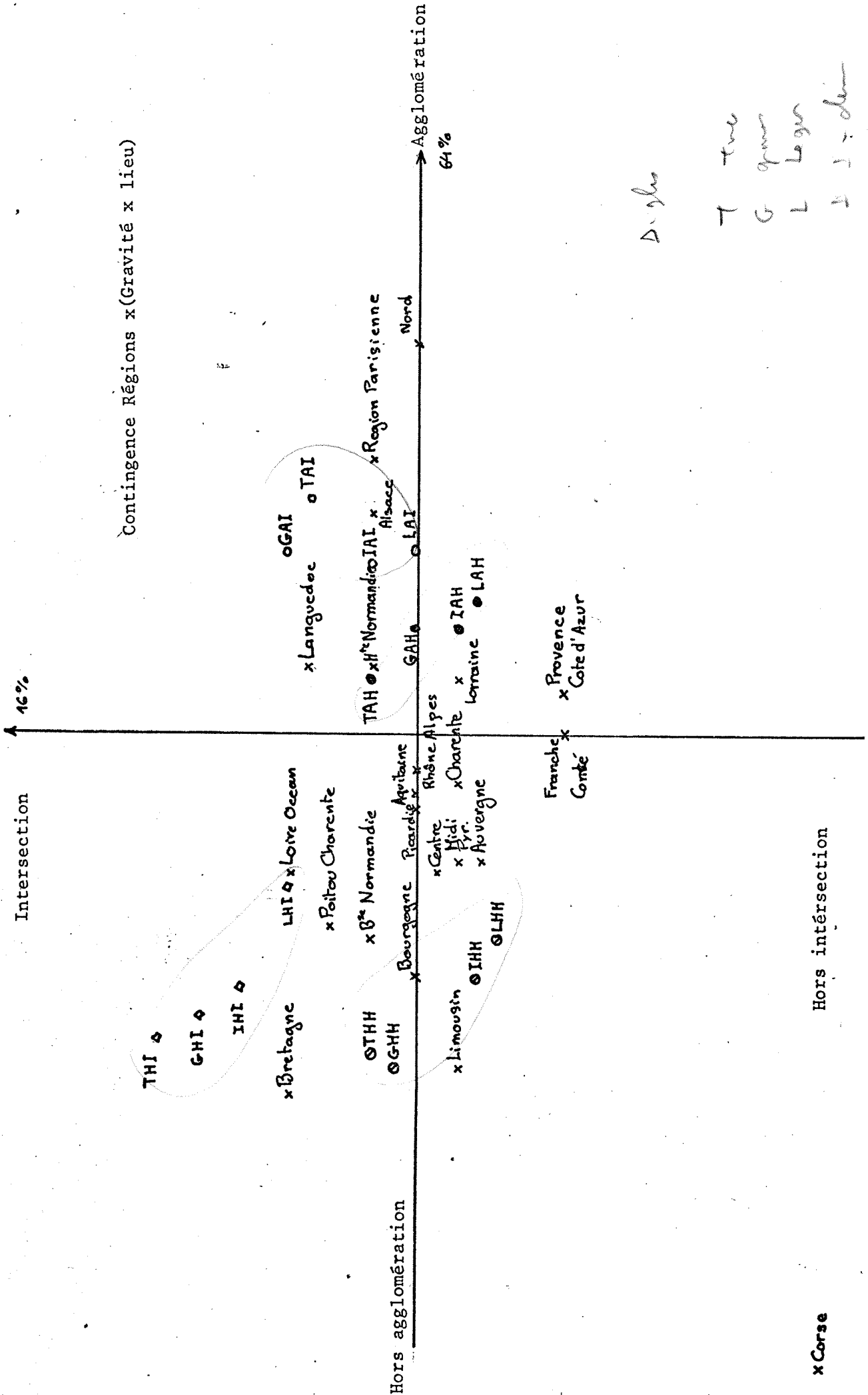
La variance de la table de fréquence est égale à  $8,3 \cdot 10^{-3}$ .

La variance du tableau (Région x Lieu) est égale à  $4,6 \cdot 10^{-2}$  donc cinq fois plus forte que la précédente. Cette disproportion apparaissait dans les correspondances ternaires. Les variations géographiques entre les régions sont bien plus importantes que les variations de gravité des accidents.

Les quatre points représentant les modalités du lieu sont bien séparés, le premier axe oppose les régions urbanisées aux régions non urbanisées, le second porte sur la présence ou l'absence d'intersection routière.

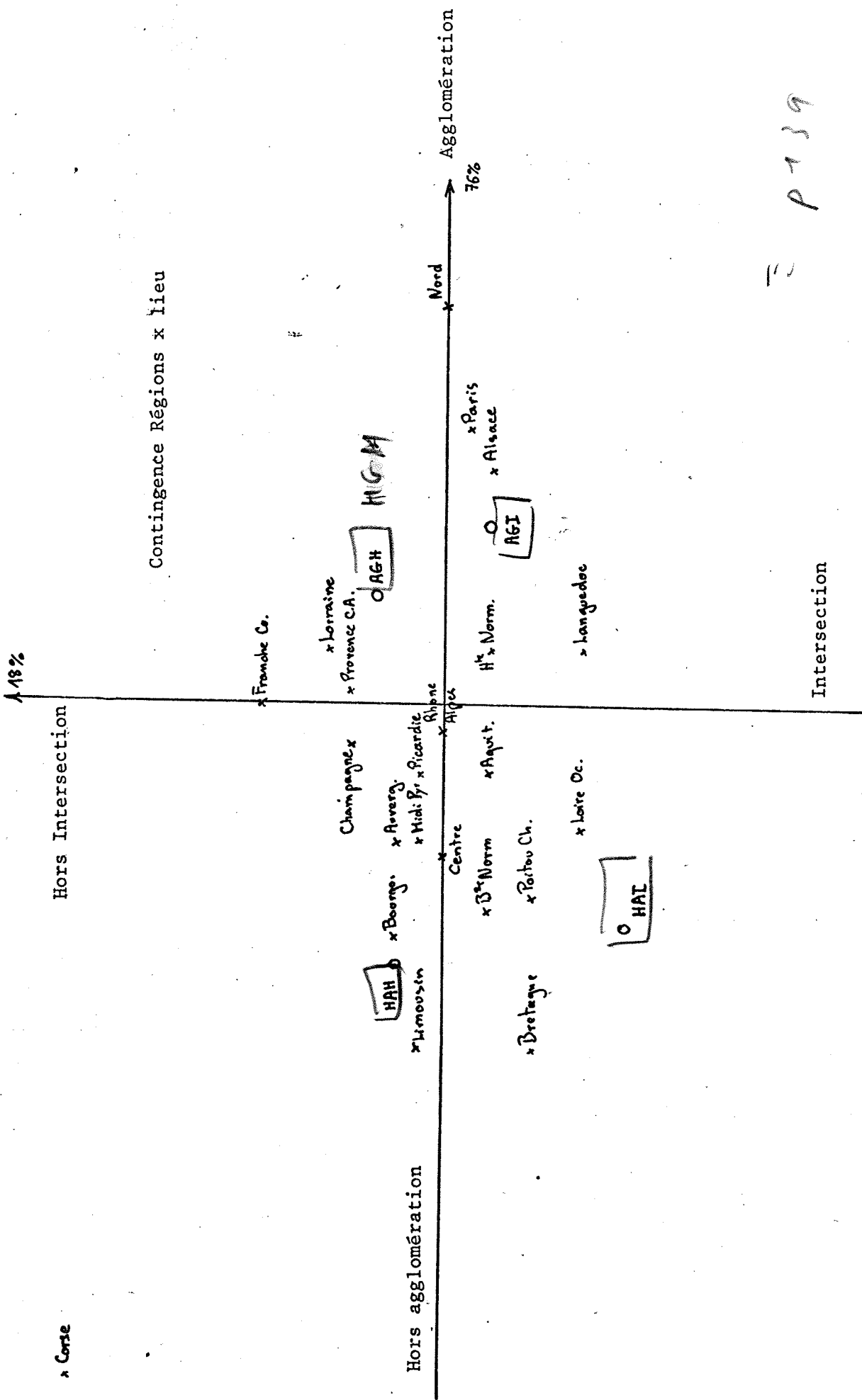


Contingence Régions x (Gravité x lieu)



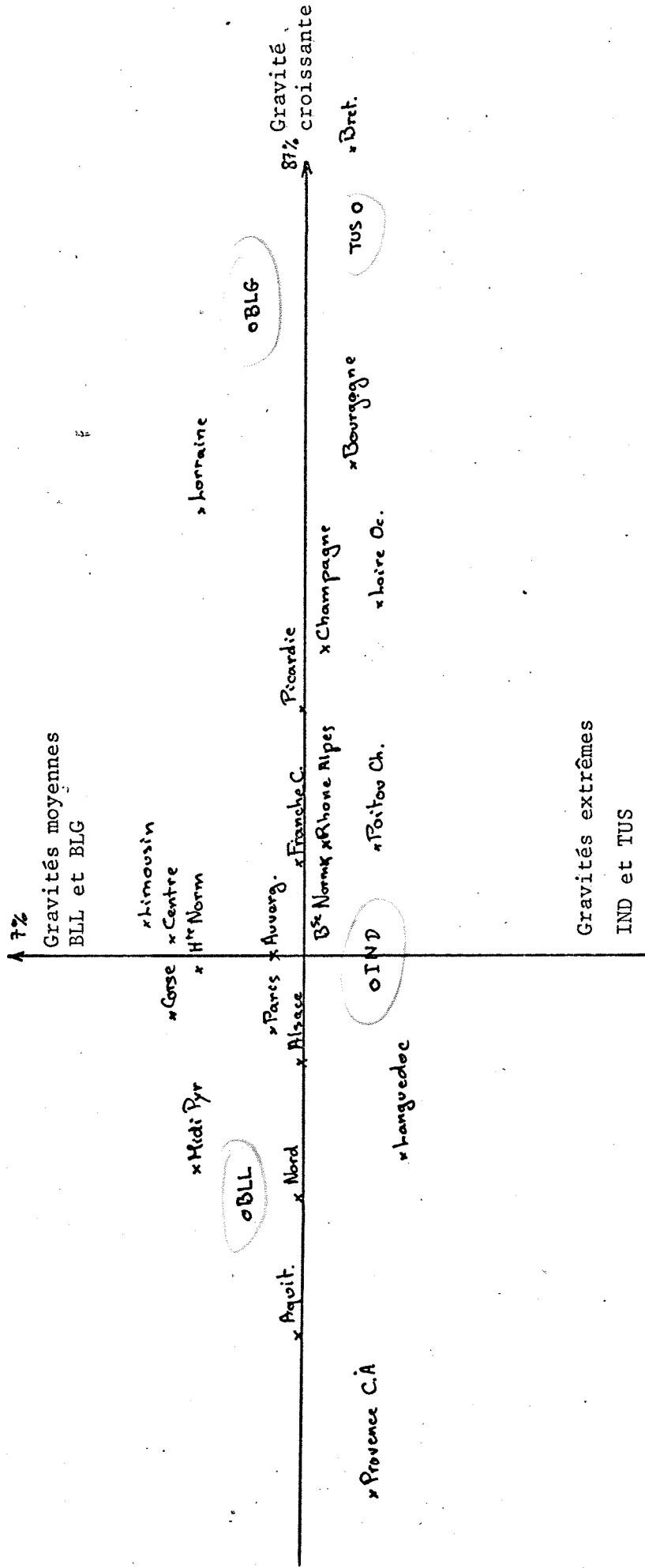
D. plus

- T tre
- G grave
- L Legu
- D J - de



3 P 139

Contingence Régions x Gravité



my

*graphique*

La troisième marge de dimension (4 x 4) confirme la forte liaison qui existe entre la gravité et le lieu, la variance vaut  $1,8 \cdot 10^{-2}$ .

G L

Les accidents en agglomération et en intersection sont les moins dangereux. On peut remarquer que les facteurs sur le lieu ou sur la gravité sont notablement différents des facteurs sur ces mêmes ensembles rencontrés dans les contingences précédentes. Cela laisse supposer que la contingence induite n'aura pas le pouvoir prévisionnel maximum qui correspondrait au produit des variances des tableaux (Région x lieu) et (lieu x gravité).

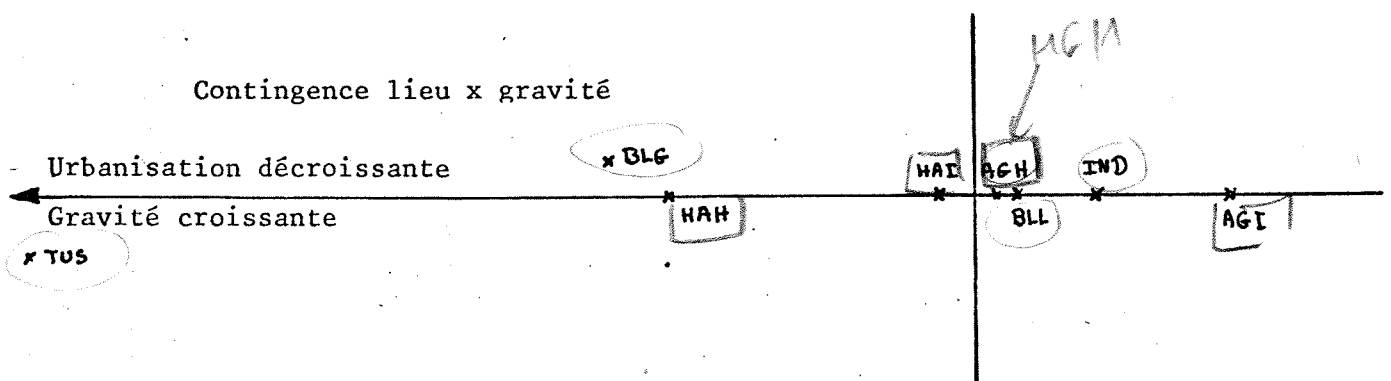
*Σ m. b. y. k. / n. k*

Cette contingence n'a qu'une variance égale à  $5,6 \cdot 10^{-4} \ll 4,6 \cdot 10^{-2} \times 1,8 \cdot 10^{-2} = 8,3 \cdot 10^{-4}$ .

R L L G

Le seul facteur significatif oppose les régions urbanisées qui sont maintenant les moins dangereuses, aux régions moins denses (Limousin, Bretagne, Corse) qui devraient si le lieu seul conditionnait la gravité des accidents, être les moins sûres.

*graphique*  
*de la gravité*

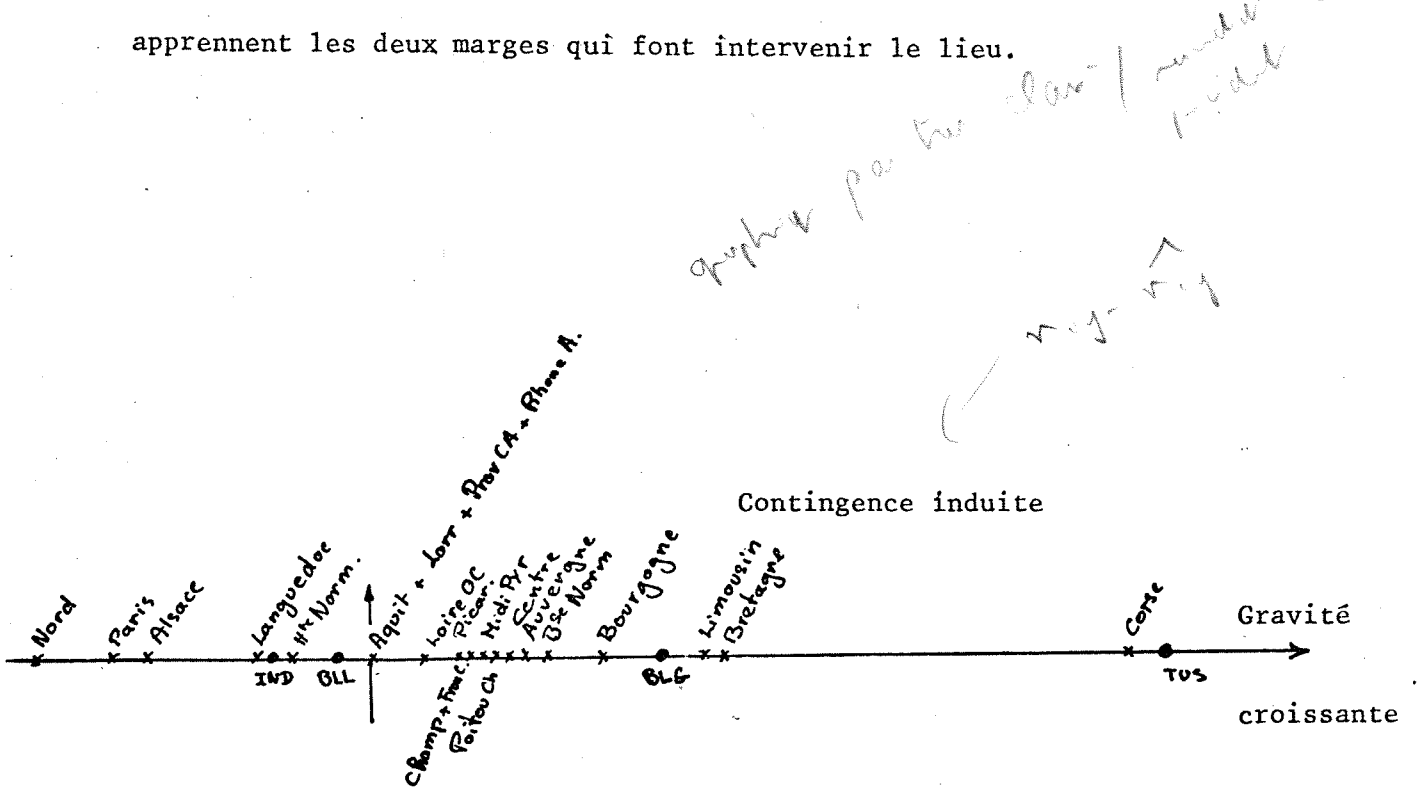


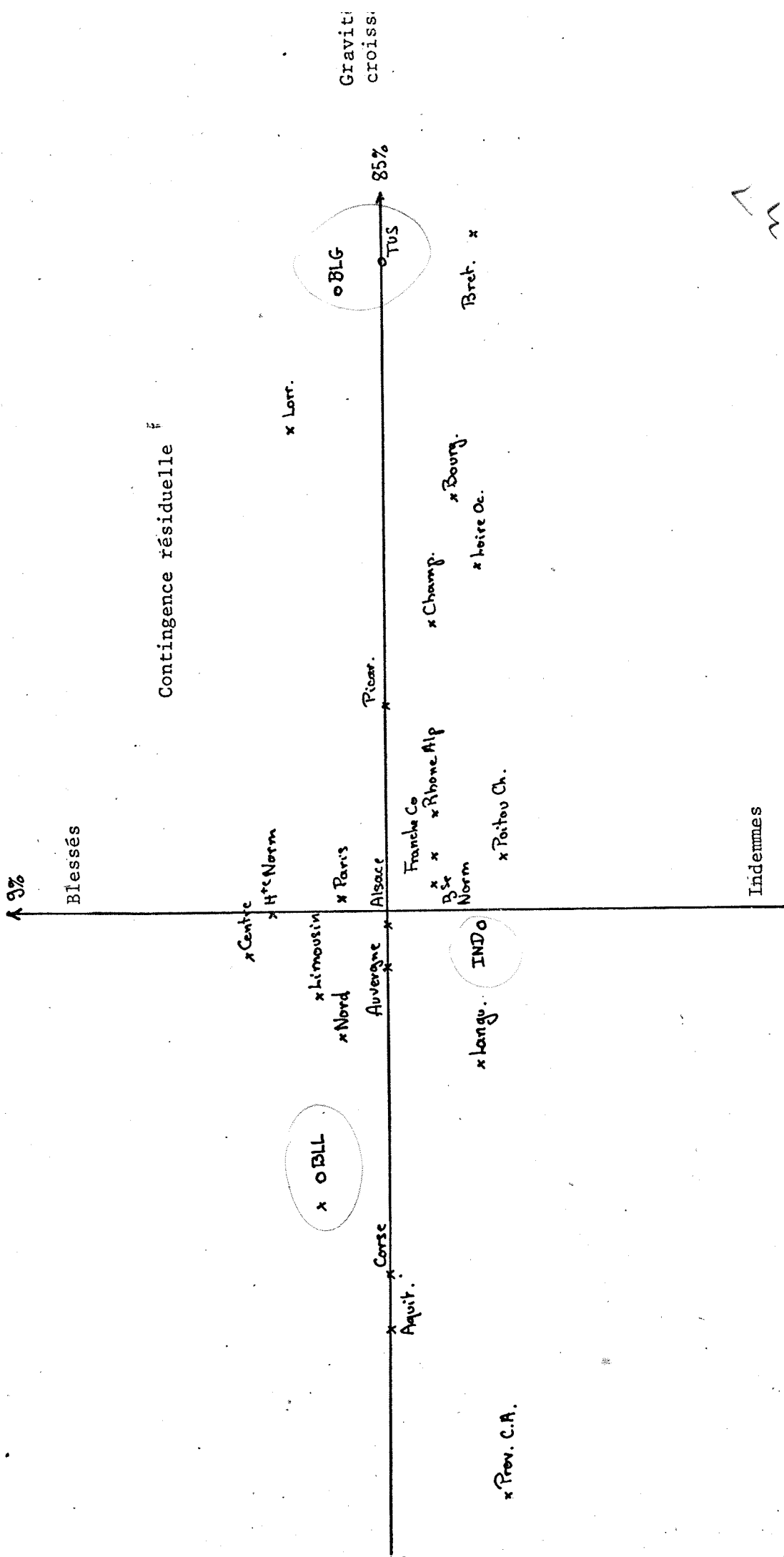
Dans cet exemple il n'y a pas moyen de comparer directement les deux contingences (gravité-région) puisque les facteurs sur la gravité sont différents. Il faut donc procéder à une décomposition de la contingence résiduelle  $(n_{IJ} - \hat{n}_{IJ})$ . La table de fréquence donne une variance de  $7,4 \cdot 10^{-3}$ , 10 % plus faible que celle de la table observée ( $f_{IJ}$ ).

*(R.G. = 18,410%)  
p. 145*

La variable lieu semble pouvoir expliquer une partie des fluctuations de la gravité entre les régions. Il reste 90 % du phénomène à expliquer par d'autres facteurs non introduits tels que le climat, l'état des routes. Ces facteurs devront être de préférence indépendants du premier caractère explicatif le lieu de l'accident.

Le dernier graphique compare les deux contingences ; les déplacements des points "Régions" sur l'échelle de gravité correspondent à ce que nous apprennent les deux marges qui font intervenir le lieu.

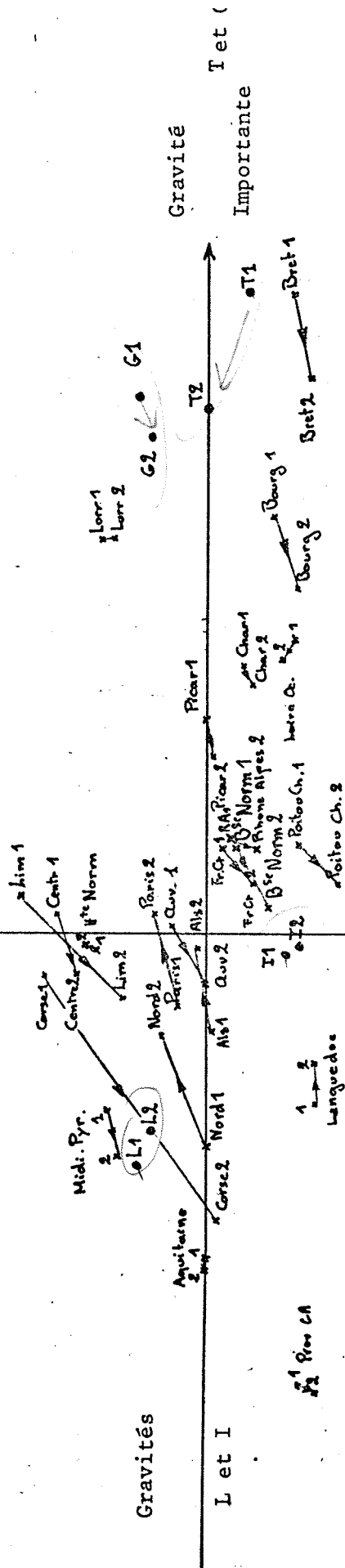




ch p 143

Gravités moyennes G et L

Comparaison des deux contingences  
Régions x gravités



Le troisième exemple est tiré d'une enquête de l'Institut de Recherche des Transport sur les déplacements touristiques dans la région Lyonnaise.

Il montre que la catégorie socio-professionnelle ne permet pas de définir des couches homogènes dans la population enquêtée.

Le tableau étudié était le suivant :

- CSP - 9 modalités (agriculteurs, patrons, professions libérales, cadres moyens, employés, ouvriers, services, autres actifs, inactifs).
- Mode de transport - 5 modalités (voiture, train, car, avion, autre)
- Mode d'hébergement - 7 modalités (hôtel, location, résidence secondaire, chez de parents, chez des amis, camping, autre).

L'enquête portait sur 1300 personnes. Les trois calculs de  $\chi^2$  s les marges ont donné les résultats ci-dessous :

$\chi^2$  (Transport-hébergement) = 260 pour 6 x 4 = 24 d.d.1 cette valeur est due pour plus de 50 % à l'association (car, autre hébergement). En supprimant ces deux modalités particulières, le  $\chi^2$  reste très significatif (120 pour 15 dd1).

La table de fréquence a une variance de :  $\frac{\chi^2}{N} = 0,2$

$\chi^2$  (CSP x hébergement) = 260 pour 8 x 6 = 48 d.d.1.

Il est très significatif, et la contribution des modalités est assez équilibrée (  $\frac{\chi^2}{N} = 0,2$  pour les fréquences).

$\chi^2$  (CSP x Transport) 160 pour 8 x 4 = 32 d.d.1



Il y a une dépendance significative. La contribution des inactifs est très importante, ce sont des touristes particuliers vis-à-vis des moyens de transports (50 % d'entre eux utilisent l'automobile, ce qui est très peu). Pour les fréquences  $\frac{\chi^2}{N} \approx 0,12$ .

La prévision qu'il est possible d'effectuer à l'aide de la variable CSP donnera un tableau de fréquence dont la variance ne pourra pas dépasser la valeur  $\frac{\chi^2_{CSP.T}}{N} \times \frac{\chi^2_{CSP.H}}{N} = 0,12 \times 0,2 = 0,024$ . Cette valeur est très faible vis-à-vis de la variance de la contingence Transport-hébergement. La prévision donne exactement une variance de 0,005 bien plus faible que le maximum possible nous avons  $\frac{\hat{f}_{11}}{f_1 f_2} \neq 1$  et la contingence résiduelle est très proche de la contingence observée ( $\frac{\chi^2_R}{N} = 0,204$  contre 0,207). A peine 2 % de la contingence transport hébergement est expliquée par la CSP.

Cette conclusion négative conduit à penser que les vacanciers constituent une population particulière au niveau de vie assez constant qui se distingue plus par la possession de plusieurs véhicules ou d'une résidence secondaire que par les catégories socio-professionnelles. On peut aussi penser que les 9 CSP définies par l'INSEE ne constituent pas des sous-ensembles suffisamment homogènes, les patrons regroupent par exemple les industriels et les petits commerçants.

BIBLIOGRAPHIE

- Bartlett M.S - - Contingency table interactions J.R. Stat. Soc. 1935
- Benzecri J.P. - Théorie de l'information et classification d'après un  
tableau de contingence
- Lois de probabilité sur un ensemble produit Les diverses  
notions d'indépendance et le critère d'entropie maximale
- Sur l'analyse des tableaux binaires associés à une  
correspondance multiple
- Décomposition de fonctions et de mesures sur un espace  
produit
- Laboratoire de Statistiques mathématiques. Université  
Paris VI
- Carroll JD et Chang JJ - Analysis of individual differences in multidimensional  
Scaling via an N-way généralization of "Eckart-Young"  
décomposition
- Psychométrica Sept. 70.
- Caussinus H. - Recherche sur divers problèmes relatifs aux tables de  
contingence
- Thèse 3ème Cycle 1962 Toulouse
- Cramer H - Mathematical Methods of Statistics Chap. 30  
Prince ton Univ. Press
- Darroch JN - Interactions in multifacteurs contingency tables  
JR Stat. Soc. 1962
- Dugué D et Girault M - Analyse de variance et plans d'expériences  
Dunod 1969
- Escoffier-Cordier B - L'analyse factorielle des correspondances  
Thèse 3ème cycle 1965 Rennes

- Kastenbaum MA et Lamphicar DE - Calculation of Chi-Square to test the no-three factors interaction hypothesis  
Biometrics 1959
- Kendall MG - The advanced theory of statistics  
Chap. 33 "Categorized Data"  
Charles Griffin Londres
- Kullback S - Information Theory and Statistics  
Chap 8' Contingency tables' J. Wiley NY
- Lancaster HO - The chi-Square distribution  
Interscience 1959
- Complex contingency tables treated by the partition of  $\chi^2$   
JR Stat. Soc. 1951
- Lewis BN - On analysis of interaction in multidimensional contingency tables  
JR Stat Soc 1962
- Mitra SK - Contributions to the statistical analysis of categorical data  
North Carolina Institute of statistics 1955
- Plackett RL - A note on interactions in contingency tables  
JR Stat. Soc 1962
- Roy SN et Kastenbaum MA - On hypothesis of no interaction in a multi-way contingency table  
Ann. Math. Stat. 1956
- Roy SN - Some aspects of multivariate analysis  
Chap 15 "Some non parametric generalisation of analysis of variance and multivariate analysis"  
J. Wiley NY

- Kastenbaum MA et Lamphicar DE - Calculation of Chi-Square to test the no-three factors interaction hypothesis  
Biometrics 1959
- Kendall MG - The advanced theory of statistics  
Chap. 33 "Categorized Data"  
Charles Griffin Londres
- Kullback S - Information Theory and Statistics  
Chap 8 "Contingency tables" J. Wiley NY
- Lancaster HO - The chi-Square distribution  
Interscience 1959
- Complex contingency tables treated by the partition of  $\chi^2$   
JR Stat. Soc. 1951
- Lewis BN - On analysis of interaction in multidimensional contingency tables  
JR Stat Soc 1962
- Mitra SK - Contributions to the statistical analysis of categorical data  
North Carolina Institute of statistics 1955
- Plackett RL - A note on interactions in contingency tables  
JR Stat. Soc 1962
- Roy SN et Kastenbaum MA - On hypothesis of no interaction in a multi-way contingency table  
Ann. Math. Stat. 1956
- Roy SN - Some aspects of multivariate analysis  
Chap 15 "Some non parametric generalisation of analysis of variance and multivariate analysis"  
J. Wiley NY

Snedecor GW

- "Chi-Squares of Bartlett, Mood and Lancaster  
in a 2<sup>3</sup> contingency table"

Biometrics 1958

Wilks SS

- Mathematical Statistics Chap. 13 "Testing  
Parametric statistical hypothesis"

J. Wiley NY