

# REVUE DE STATISTIQUE APPLIQUÉE

V. CHOULAKIAN

J. DE TIBEIRO

## **Copules archimédiennes et tableaux de contingence à variables qualitatives ordinales**

*Revue de statistique appliquée*, tome 48, n° 4 (2000), p. 83-96

[http://www.numdam.org/item?id=RSA\\_2000\\_\\_48\\_4\\_83\\_0](http://www.numdam.org/item?id=RSA_2000__48_4_83_0)

© Société française de statistique, 2000, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

## COPULES ARCHIMÉDIENNES ET TABLEAUX DE CONTINGENCE À VARIABLES QUALITATIVES ORDINALES \*

V. Choulakian\*\*, J. de Tibeiro\*\*\*

\*\* *Département de Mathématiques et Statistique, Université de Moncton, Moncton, N.B.,  
Canada, E1A 3E9.*

\*\*\* *Secteur des Sciences, Université de Moncton, Campus de Shippagan, Canada, E8S 1P6.*

### RÉSUMÉ

Le modèle d'association U (uniforme) de Goodman tout comme le modèle de Plackett sont d'habitude utilisés pour analyser des tableaux de contingence dont les variables sont ordinales. Il n'en est pas moins vrai que dans certaines situations, ces deux modèles peuvent s'avérer infructueux pour ajuster convenablement de tels tableaux. Pour combler justement cette lacune, nous proposons une voie complémentaire à ces modèles, axée cette fois-ci sur l'utilisation de trois copules Archimédiennes bidimensionnelles à un paramètre. Le paramètre d'association de ces copules est estimé par la méthode du maximum de vraisemblance qui crée – rappelons-le – les conditions de choix d'un bon modèle. Nous proposons à cette fin deux exemples illustratifs.

**Mots-clés :** *Loi bidimensionnelle, loi multinomiale, dépendance quadrangulaire positive, mesure d'association, méthode du maximum de vraisemblance, modèle d'association U de Goodman, modèle de Plackett, copules Archimédiennes, AFC.*

### ABSTRACT

Usually, either Goodman's uniform association model or Plackett's model are applied for the analysis of association in contingency tables with ordered categories. In some cases, these two well known methods do not fit well the data set. We propose, in addition to these two methods, the use of three one-parameter bivariate Archimedean copulas. The parameter of association in these copulas is estimated by maximum likelihood, which at the same time helps us in the choice of a good model. Two examples are provided.

**Keywords :** *Bivariate distribution, multinomial distribution, positive quadrant dependence, measure of association, maximum likelihood, Goodman's U association model, Plackett's model, Archimedean copulas, correspondence analysis.*

---

\* Ce travail de recherche a été réalisé grâce à une subvention de Conseil de Recherches en Sciences Naturelles et en Génie du Canada (CRSNG) et à celle du fonds International de Coopération Universitaire (FICU).

## 1. Introduction

Dans ce travail, nous supposons au départ que les variables ordinales du tableau de contingence ne sont pas indépendantes, ce qui confère une certaine forme d'association entre ces variables. Au demeurant, notons que dans la plupart des cas, cette forme d'association est de nature positive. On pourrait donc s'attendre à ce qu'une variable ordinale devienne plus grande (ou plus petite) lorsque l'autre variable devient encore plus grande (ou encore plus petite).

Dans la littérature statistique classique, il existe plus d'une demi-douzaine de concepts d'association positive relatifs aux lois bidimensionnelles. Voir par exemple Hutchinson et Lai (1990, chapitre 3). Toutefois, l'un des plus grands concepts de base est celui de Dépendance Quadrangulaire Positive (DQP), introduite par Lehmann (1966). Lehmann a introduit aussi le concept de Dépendance Quadrangulaire Négative (DQN) qui est moins utilisé dans la pratique.

La loi de deux variables aléatoires  $X$  et  $Y$  est dite de DQP si  $P(X > x, Y > y) \geq P(X > x) P(Y > y)$  pour tout  $x, y$ . Une autre manière pour exprimer DQP est donnée par  $P(X > x | Y > y) \geq P(X > x)$  pour laquelle connaissant  $Y$  grand signifie que  $X$  l'est davantage pour autant qu'on ignore  $Y$ .

Un test statistique pour DQP basé sur la randomisation a été proposé par Nguyen et Sampson (1987); pour un tableau de contingence  $2 \times 2$ , ce dernier test est équivalent au test exact de Fisher basé sur la loi Hypergéométrique. A cause de l'aspect combinatoire du test de randomisation, le test statistique proposé par Nguyen et Sampson (1987) ne peut s'appliquer que pour de très petits ensembles de données.

Généralement, l'association entre des variables ordinales dans les tableaux de contingence est modélisée soit par la loi bidimensionnelle de Plackett – voir par exemple Plackett (1965) et Wahrendorf (1980) –, soit par le modèle d'association U (uniforme) de Goodman (1979, 1981). Voir aussi, Ogborn (1984) et Dale (1984). Il arrive que ces modèles ne conviennent pas pour ce type de données ordinales. Ainsi une bonne sélection de modèles pour analyser un tel ensemble de données requiert la considération de plusieurs modèles avant d'effectuer le choix d'un modèle approprié.

Nous proposons dans cette étude un complément aux deux modèles évoqués ci-dessus : l'utilisation des copules Archimédiennes bidimensionnelles à un paramètre et ce à des fins d'analyse des tableaux de contingence dont les variables sont qualitatives ordinales. Le paramètre d'association dans ces copules Archimédiennes sera estimé par la méthode du maximum de vraisemblance qui concourt au choix d'un meilleur modèle.

Nous retenons à cet effet deux références sur des copules Archimédiennes : Genest et MacKay (1986) et Genest et Rivest (1993). Pour une bonne référence sur des lois bidimensionnelles continues, on pourra consulter Hutchinson et Lai (1990).

Ce travail s'articule comme suit. Dans le paragraphe 2, nous introduisons quelques notations usuelles d'usage. Nous présentons ensuite deux ensembles de données pour lesquels nous démontrerons par la suite qu'ils sont de Dépendance Quadrangulaire Positive (DQP). Dans le paragraphe 3, nous présentons cinq modèles comprenant les trois copules Archimédiennes bidimensionnelles lesquelles seront utilisées pour analyser les corpus de données retenus à cette fin. Dans le paragraphe 4,

nous procédons à l'analyse des deux ensembles de données constituant le corpus même des données.

Il y a dans ce droit fil d'autres approches proposées pour modéliser des tableaux de contingence à deux variables ordinales; voir à ce sujet, Lapp, Molenberghs et Lesaffre (1998), Agresti et Coull (1998), Rayner et Best (1996), Jefferson, May et Ravi (1989), et, Nishisato et Arri (1975).

## 2. Le Corpus de Données

Soit  $(X_\alpha, Y_\alpha)$  pour  $\alpha = 1, 2, \dots, n$  un échantillon aléatoire de taille  $n$  associé à une loi bidimensionnelle où  $X$  décrit l'ensemble des valeurs ordinales  $i$  pour  $i = 1, \dots, a$  et  $Y$  décrit l'ensemble des valeurs ordinales  $j$  pour  $j = 1, \dots, b$ .

Notons  $P = (p_{ij}) = m_{ij}/n$  la distribution conjointe de probabilité observée sous forme matricielle où  $m_{ij}$  est le nombre observé d'unités ( $X = i, Y = j$ ). Les lois de densité marginales pour  $X$  et  $Y$  sont respectivement indiquées sous forme vectorielle par  $P_r = (p_{i.})$  et  $P_c = (p_{.j})$ . Sous l'hypothèse de l'indépendance des variables  $X$  et  $Y$ , la fonction de densité conjointe est notée  $P_I = (p_{i.p.j})$ .

Pour un tableau à deux entrées  $T = (t_{ij})$ , nous définissons l'opération de sommation partielle cumulative sur la matrice  $T$ ,  $SPC(T) = (c_{ij})$ ; pour

$i = 1, \dots, a$ ,  $j = 1, \dots, b$ ; et  $c_{ij} = \left( \sum_{\alpha=1}^i \sum_{\beta=1}^j t_{\alpha\beta} \right)$ . Plus généralement, si  $T_1$  et  $T_2$

sont deux matrices  $a \times b$  de mêmes lignes et de mêmes colonnes, nous écrivons  $T_1 \geq T_2$  si  $SPC(T_1 - T_2) \geq 0$ . La fonction de densité conjointe de  $P$  est alors DQP si  $P \geq P_I$ , et  $P$  est strictement de DQP si  $P > P_I$  et  $P \neq P_I$ .

Nous avons retenu dans ce travail, deux tableaux de données qui seront utilisés ultérieurement à des fins d'application. Le tableau 1 dû à Srole *et al.* (1974) ventile des individus à travers le Centre-Ville de Manhattan classés d'après le statut sanitaire mental ( $X$ ) et le statut socio-professionnel parental ( $Y$ ). Dans ce tableau de contingence, cohabitent des données mises entre parenthèses qui correspondent aux valeurs du tableau  $nSPC(P - P_I)$ . Visuellement, nous pouvons voir que  $P > P_I$ , raison pour laquelle les deux variables étudiées, à savoir le statut sanitaire mental et le statut socio-économique parental sont strictement de DQP. Hutchinson et Lai (1990) ont cité 11 références relatives à l'analyse statistique du tableau 1. Le tableau 2, dû à Glass et Hall (1954), présente des sujets britanniques de sexe masculin classés d'après leur statut d'emploi ( $Y$ ) et celui de leurs pères ( $X$ ). De manière analogue, on voit que les deux variables sont strictement de DQP. Ce tableau de données a été successivement analysé par Wahrendorf (1980) et Goodman (1981).

TABLEAU 1

**Individus du centre-ville de Manhattan classés d'après le statut sanitaire mental et le statut socio-économique parental. Les couches (sociales) A à F sont ordonnées de haut en bas. Entre parenthèses sont données les entrées de la matrice de somme partielle cumulative  $n$  SPC( $P - P_1$ ), qui confirment la nature DQP des variables.**

Statut sanitaire mental	Couches socio-économiques parentales					
	A	B	C	D	E	F
Bon état de santé (Bon)	64 (15)	57 (27)	57 (31)	72 (32)	36 (19)	21 (00)
Formation légère de symptômes (leg)	94 (14)	94 (31)	105 (36)	141 (39)	97 (26)	71 (00)
Formation modérée de symptômes (mod)	58 (15)	54 (32)	65 (40)	77 (36)	54 (20)	54 (00)
Formation marquée de symptômes (mrq)	31 (11)	21 (17)	34 (21)	51 (17)	43 (09)	39 (00)
Formation sévère de symptômes (sev)	10 (02)	11 (01)	23 (05)	32 (05)	27 (04)	22 (00)
Invalidité (inv)	5 (00)	8 (00)	3 (00)	11 (00)	8 (00)	10 (00)

TABLEAU 2

**Sujets britanniques de sexe masculin classés par leur propre statut et par celui de leurs parents. Les colonnes représentent des catégories des statuts de leurs parents. Entre parenthèses sont données les entrées de la matrice de somme partielle cumulative  $n$  SPC( $P - P_1$ ) qui confirment la nature DQP des 2 variables.**

	$P_1$	$P_2$	$P_3$	$P_4$	$P_5$
$F_1$	50(46)	45(73)	8(64)	18(29)	8(0)
$F_2$	28(59)	174(191)	84(201)	154(118)	55(0)
$F_3$	11(54)	78(192)	10(244)	223(173)	96(0)
$F_4$	14(23)	150(099)	185(130)	714(165)	447(0)
$F_5$	3(0)	42(0)	72(0)	320(0)	411(0)

### 3. Modèles Bidimensionnels

Nous allons d'abord introduire quelques notations. Sous l'hypothèse de l'existence d'une certaine association entre  $X$  et  $Y$ , nous désignons la fonction de densité bidimensionnelle par  $h_{ij}$ , et la fonction de répartition cumulative par  $H_{ij}$ . Cinq

modèles bidimensionnels sont représentés dans cette section, les deux premiers étant bien connus dans la littérature statistique.

### 3.1. Modèle d'Association Uniforme de Goodman

Considérons d'abord le modèle log-linéaire saturé

$$h_{ij} = \alpha_i \beta_j \eta_{ij}$$

où  $\alpha_i$  représente l'effet marginal de la ligne  $i$ ,  $\beta_j$  l'effet de la colonne  $j$  et  $\eta_{ij}$  l'interaction entre la ligne  $i$  et la colonne  $j$ . Ces paramètres vérifient les conditions suivantes :  $\sum_i \ln \alpha_i = \sum_j \ln \beta_j = \sum_i \ln \eta_{ij} = \sum_j \ln \eta_{ij} = 0$ . Le modèle log-linéaire sous l'hypothèse de l'indépendance s'obtient en fixant  $\eta_{ij} = 1$ . De plus, on sait que ces deux modèles log-linéaires ne tiennent pas compte du caractère ordinal des variables. Cette lacune a déjà été contournée par le modèle d'association uniforme de Goodman (1979) relatif à la loi de densité de probabilité conjointe entre  $X$  et  $Y$

$$h_{ij} = \alpha_i \beta_j \exp(\varphi \mu_i \nu_j) \tag{1}$$

où  $\mu_i = i - \sum_{i=1}^a i/a$  pour  $i = 1, \dots, a$  et  $\nu_j = j - \sum_{j=1}^b j/b$  pour  $j = 1, \dots, b$ , ce qui implique  $\mu_{i+1} - \mu_i = \nu_{j+1} - \nu_j = 1$ . Les  $\alpha_i$ ,  $\beta_j$  et  $\varphi$  sont des paramètres à estimer pour  $0 \leq \varphi < \infty$ .

L'interprétation de  $\varphi$  est très simple. Elle représente les logarithmes des quotients impairs locaux constants basés sur les probabilités

$$\log [(h_{ij} h_{i+1,j+1}) / (h_{i,j+1} h_{i+1,j})] = \varphi \tag{2}$$

Si  $\varphi > 0$ , alors  $X$  et  $Y$  sont de DQP; si  $\varphi < 0$ , alors  $X$  et  $Y$  sont de DQN et si  $\varphi = 0$ , alors  $X$  et  $Y$  sont indépendantes, voir Goodman (1979). Le modèle d'association uniforme de Goodman est aussi un modèle log-linéaire ayant un paramètre de plus que le modèle log-linéaire sous l'hypothèse de l'indépendance.

### 3.2. Copules

Les quatre autres modèles bidimensionnels sont des copules, ce qui fait de la fonction de répartition bidimensionnelle  $H_{ij} = \sum_{\alpha=1}^i \sum_{\beta=1}^j h_{\alpha\beta}$  une simple fonction de

la classe des fonctions de répartition marginale des données observées  $F_i = \sum_{\alpha=1}^i p_{\alpha}$ .

et  $G_j = \sum_{\beta=1}^j p_{\beta}$  et du paramètre d'association  $\varphi$ .

Rappelons qu'une copule est dite Archimédienne s'il existe une fonction convexe décroissante  $s$  définie sur l'intervalle  $[0, 1]$  telle que  $s(1) = 0$  et la relation  $s(H_{ij}) = s(F_i) + s(G_j)$  est satisfaite. La fonction  $s$  est appelée fonction de dépendance. Nous allons utiliser dans cette étude trois copules Archimédiennes : les copules de Gumbel, Frank et Clayton. La modélisation des tableaux de contingence dont les variables sont qualitatives ordinales par des copules associées à la propriété de DQP, peut justement être motivée par l'argument suivant émis par Schriever (1986). Supposons que  $(X, Y)$  soit un couple de variables aléatoires bidimensionnelles de fonction de répartition conjointe continue. Supposons que l'on obtienne la matrice de probabilité  $P$  en discrétisant  $(X, Y)$  de telle manière qu'il existe deux suites croissantes de nombres  $A_0 < A_1 < \dots < A_a$  et  $B_0 < B_1 < \dots < B_b$  telle que  $p_{ij} = \Pr(X \in ]A_{i-1}, A_i[ \text{ et } Y \in ]B_{j-1}, B_j[)$ ,  $p_{i.} > 0$  et  $p_{.j} > 0$  pour  $i = 1, \dots, a$  et  $j = 1, \dots, b$ . Si  $(X, Y)$  est de DQP, alors  $P$  l'est aussi. Ainsi, par exemple si le coefficient de corrélation  $\rho$  est positif, une discrétisation de la loi normale bidimensionnelle devrait toujours produire une autre distribution  $P$  de DQP.

### 3.2.1. Copule de Plackett

La loi bidimensionnelle de Plackett est définie comme suit :

$$H_{ij} = \left[ \frac{[1 + (\varphi - 1)(F_i + G_j)] - [\{1 + (\varphi - 1)(F_i + G_j)\}^2 - 4\varphi(\varphi - 1)F_iG_j]^{1/2}}{[2(\varphi - 1)]} \right] \quad \text{pour } \varphi \neq 1 \quad (3)$$

et en passant à la limite

$$H_{ij} = F_i G_j \quad \text{si } \varphi = 1$$

Le paramètre d'association  $\varphi$ , pour  $0 \leq \varphi$  admet une interprétation analogue à (2). On obtient un tableau de probabilité  $2 \times 2$  en dichotomisant le tableau de contingence initial de dimension  $a \times b$  à la ligne  $i$  et à la colonne  $j$ , et en additionnant les probabilités à l'intérieur de chaque quadrant. Le quotient impair global au couple  $(i, j)$  pour la loi (3) est donné par la relation suivante :

$$\frac{\left( \sum_{\alpha=1}^i \sum_{\beta=1}^j h_{\alpha\beta} \right) \left( \sum_{\alpha=i+1}^a \sum_{\beta=j+1}^b h_{\alpha\beta} \right)}{\left( \sum_{\alpha=1}^i \sum_{\beta=j+1}^b h_{\alpha\beta} \right) \left( \sum_{\alpha=i+1}^a \sum_{\beta=1}^j h_{\alpha\beta} \right)} = \frac{[H_{ij}(1 - F_i - G_j + H_{ij})]}{[(F_i - H_{ij})(G_j - H_{ij})]} \quad (4)$$

$$= \varphi$$

Si  $\varphi > 1$  dans (4), alors  $X$  et  $Y$  sont de DQP; si  $\varphi < 1$ , alors  $X$  et  $Y$  sont de DQN, et si  $\varphi = 1$ , alors  $X$  et  $Y$  sont indépendantes, voir Agresti (1984, page 29). L'équation (3) est connue comme la copule de Plackett qui n'est pas Archimédienne.

#### 4. Copules Archimédiennes

Voici les trois copules Archimédiennes munies de leurs fonctions de dépendance  $s$ .

##### 4.1. Copule de Gumbel

$$H_{ij} = \exp \left[ - \{ (-\log F_i)^\varphi + (-\log G_j)^\varphi \}^{1/\varphi} \right] \text{ pour } \varphi \geq 1 \quad (5)$$

$$s(v) = (-\log v)^\varphi$$

Voir par exemple, Gumbel (1960) et Hougaard (1986). Cette copule est aussi appelée copule de Gumbel-Hougaard par Hutchinson et Lai (1990). Pour  $\varphi > 1$  dans (5),  $X$  et  $Y$  sont de DQP et si  $\varphi = 1$ ,  $X$  et  $Y$  sont indépendantes. Notons que la copule de Gumbel ne peut pas être de DQN.

##### 4.2. Copule de Frank

$$H_{ij} = [\log \{ 1 + (\varphi^{F_i} - 1)(\varphi^{G_j} - 1)/(\varphi - 1) \}] / \log \varphi \text{ pour } \varphi > 0 \text{ et } \varphi \neq 1$$

$$s(v) = \log [(1 - \varphi^v)/(1 - \varphi)] \quad (6)$$

et par passage à la limite

$$H_{ij} = F_i G_j \text{ et } s(v) = \log v \text{ si } \varphi = 1$$

Voir par exemple Frank (1979) et Genest (1987). Si  $0 < \varphi < 1$  dans (6), alors  $X$  et  $Y$  sont de DQP; si  $\varphi = 1$ ,  $X$  et  $Y$  sont indépendantes, et si  $\varphi > 1$ ,  $X$  et  $Y$  sont de DQN. C'est la seule copule Archimédienne qui est telle que la fonction de survie bidimensionnelle  $\bar{H}_{ij} = \Pr(X > i, Y > j)$  s'obtient à partir de (6) en remplaçant  $F_i$  et  $G_j$  par  $1 - F_i$  et  $1 - G_j$  respectivement; ce qui revient à dire que (6) est encore valable si on inverse l'ordre des modalités de  $X$  et de  $Y$ .

##### 4.3. Copule de Clayton

$$H_{ij} = (F_i^{1-\varphi} + G_j^{1-\varphi} - 1)^{(1-\varphi)^{-1}} \text{ pour } \varphi > 0 \text{ et } \varphi \neq 1 \quad (7)$$

$$s(v) = (v^{1-\varphi} - 1)/(\varphi - 1)$$

et par passage à la limite

$$H_{ij} = F_i G_j \text{ et } s(v) = \log v \text{ si } \varphi = 1$$



Voir Clayton (1978) et Cook et Johnson (1981). Cette copule est aussi appelée copule de Pareto par Hutchinson et Lai (1990), copule de Johnson et Cook par Genest et MacKay (1986). Si  $1 < \varphi$  dans (7), alors  $X$  et  $Y$  sont de DQP; si  $\varphi = 1$ ,  $X$  et  $Y$  sont indépendantes, et si  $0 < \varphi < 1$ ,  $X$  et  $Y$  sont de DQN. Hutchinson (1979) a utilisé cette copule pour analyser des tableaux de contingence qui renferment des variables ordinales.

## 5. Remarques

Le paramètre d'association  $\varphi$  dans les trois copules Archimédiennes n'a pas une simple interprétation comme dans les relations (2) et (4). Ces trois modèles peuvent toutefois être interprétés en utilisant les fonctions de dépendance  $s$  : les différences tétrachoriques adjacentes de la fonction de répartition sur l'échelle  $s$  sont nulles, ce qui entraîne la relation suivante :

$$s(H_{ij}) + s(H_{i+1,j+1}) - s(H_{i+1,j}) - s(H_{i,j+1}) = 0$$

Notons que pour les copules de Gumbel et Clayton, on peut modéliser soit la fonction de répartition bidimensionnelle  $H$  ou la fonction de survie bidimensionnelle  $\bar{H}$ . Les trois autres modèles produisent des résultats identiques si l'on renverse l'ordre des variables ordinales  $X$  et  $Y$ .

## 6. Estimation des Paramètres

Chacune des quatre copules présentées ci-haut est caractérisée seulement par un paramètre d'association  $\varphi$ . Nous estimons  $\varphi$  par la méthode du maximum de vraisemblance, en utilisant un algorithme de recherche directe présenté en annexe. En utilisant le théorème de Birch, voir par exemple Birch (1964), Cox (1984) ou Bishop, Fienberg et Holland (1975), il vient asymptotiquement

$$\sqrt{n}(\hat{\varphi} - \varphi) \sim N(0, \hat{J}^{-1}) \quad (8)$$

où  $\hat{J}$  est la mesure d'information estimée de Fisher, et qui vaut

$$\hat{J} = \left[ \sum_i \sum_j (dh_{ij}/d\varphi)^2 / h_{ij} \right]_{\varphi=\hat{\varphi}} \quad (9)$$

où  $dh_{ij}/d\varphi$  est la dérivée de  $h_{ij}$  en fonction de  $\varphi$ . En utilisant (8), on peut facilement construire un intervalle de confiance pour  $\varphi$ .

Dans le cas du modèle uniforme de Goodman, les paramètres  $\alpha_i$ ,  $\beta_j$  et  $\varphi$  sont estimés en maximisant la vraisemblance de la loi multinomiale; pour les détails du calcul voir Goodman (1979).

## 7. Analyse des Données

Avant de modéliser ces deux tableaux de contingence évoqués ci-haut à la section 2, nous avons effectué d'abord une Analyse Factorielle des Correspondances afin de visualiser et de décrire les éventuelles affinités et dichotomies entre les deux variables.

### 7.1. Analyse du Tableau 1

Le tableau 1 a été soumis à l'Analyse Factorielle des Correspondances. Les valeurs propres et les taux d'inertie (entre parenthèses) des trois premiers axes obtenus sont respectivement : 0,0264 (81,61 %), 0,0034 (10,57 %), 0,0016 (4,87 %). La figure 1 issue de cette analyse révèle un premier axe interprétable et prédominant associé toutefois à une très faible valeur propre. On constate que les chapelets des modalités des deux variables ordinales s'échelonnent régulièrement de la droite vers la gauche dans l'ordre du statut sanitaire mental et du statut socio-économique parental.

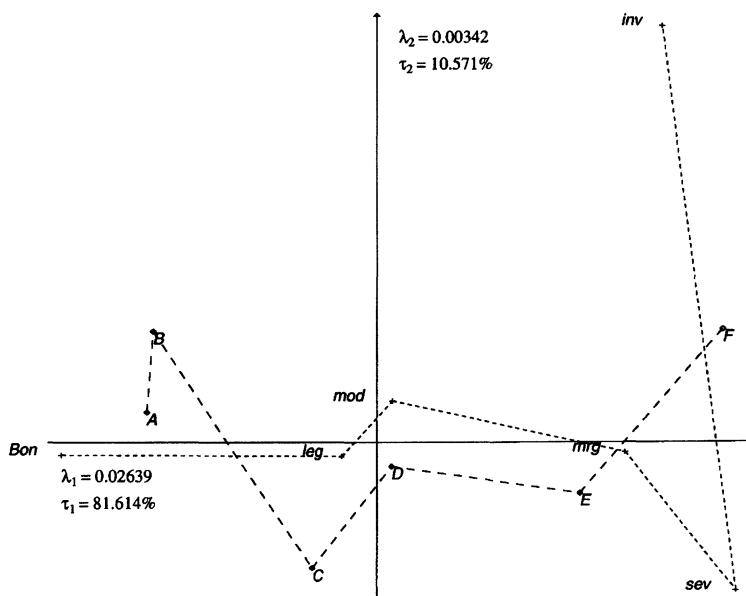


FIGURE 1

*Plan (1,2) : Individus du centre-ville de Manhattan classés d'après le statut sanitaire mental et le statut socio-économique parental*

Le tableau 3 donne des statistiques sommaires pour les modèles considérés qui s'ajustent aux données du tableau 1. RV désigne la statistique du chi-deux du rapport de vraisemblance obtenue en comparant le modèle proposé au modèle saturé,  $X^2$

est la statistique d'ajustement du chi-deux de Pearson et ces deux statistiques suivent asymptotiquement sous l'hypothèse de l'indépendance la loi du  $\chi^2$  à  $(a-1)(b-1)-1$  degrés de liberté et ce pour les cinq modèles. Les lignes Gumbel 1 et Clayton 1 font observer les résultats lorsqu'ils s'ajustent à la loi de survie, ce qui est le cas en inversant l'ordre des lignes et des colonnes. En utilisant la loi du chi-deux, nous réalisons que les 5 modèles s'ajustent assez bien aux données. Toutefois, en comparant RV et  $X^2$ , nous voyons que la copule de Frank et le modèle U de Goodman s'ajustent convenablement aux données.

TABLEAU 3

**Résumé exhaustif des statistiques pour les différents modèles**  
*qui se sont ajustés aux données du tableau 1. Gumbel 2 et Clayton 2 correspondent aux modèles (5) et (7) respectivement, tandis que Gumbel 1 et Clayton 1 correspondent aux mêmes modèles ajustés sur la fonction de survie (i.e. en inversant l'ordre de lignes et des colonnes).*

Modèles	RV	$X^2$	Degrés de liberté	$\varphi$ (écart-type de $\varphi$ )
<b>Gumbel 1</b>	25.19	24.76	24	1.096(0.019)
<b>Gumbel 2</b>	26.75	25.60	24	1.101(0.020)
<b>Frank</b>	19.59	19.02	24	0.387(0.061)
<b>Clayton 1</b>	30.95	29.51	24	6.66(1.195)
<b>Clayton 2</b>	25.55	25.52	24	6.66(1.097)
<b>Plackett</b>	20.10	19.47	24	1.59(0.122)
<b>Goodman</b>	19.15	19.18	24	1.30

L'estimateur par la méthode du maximum de vraisemblance du paramètre d'association  $\varphi$  dans la copule de Frank vaut 0.387 avec un écart-type de 0.061 ; d'où (0.265, 0.509) qui forme un intervalle de confiance pour  $\varphi$  au niveau de confiance de 95%.

### 7.2. Analyse du Tableau 2

Le tableau 2 a été également traité par l'Analyse Factorielle des Correspondances. Les valeurs propres et les taux d'inertie (entre parenthèses) des trois premiers axes obtenus sont respectivement : 0,246 (73,29 %), 0,071 (21,18 %), 0,015 (4,49 %). Avec près de 95 % de l'inertie, le plan (1,2) offre une image fidèle de l'ensemble des résultats de l'analyse. Sur le plan (1,2) des deux premiers facteurs de la figure 2, on remarque que le nuage des libellés des modalités des deux variables s'inscrit dans le croissant caractéristique de l'effet Guttman.

Le tableau 4 ventile des statistiques sommaires pour les modèles qui s'ajustent aux ensembles des données du tableau 2. Parmi tous les modèles retenus pour cette

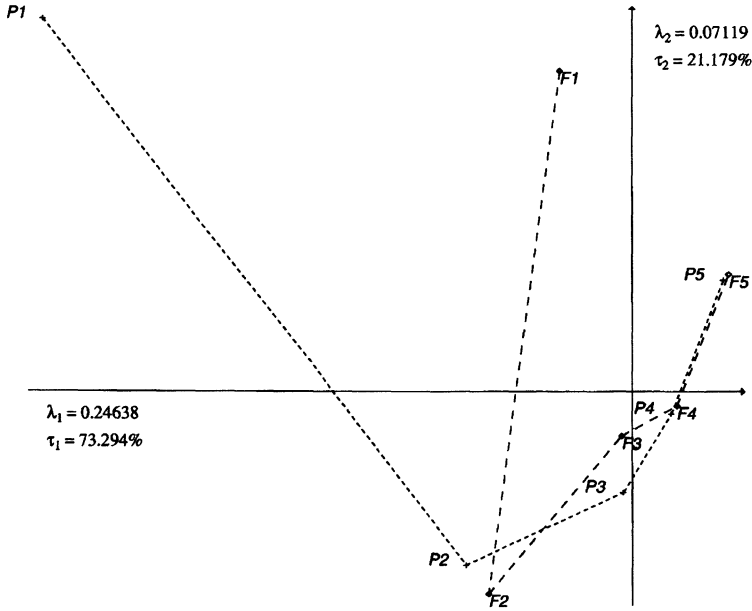


FIGURE 2  
 Plan (1,2) : Sujets britanniques de sexe masculin classés par leur propre statut et par celui de leurs parents

TABLEAU 4  
 Résumé exhaustif des statistiques pour les différents modèles qui se sont ajustés aux données du tableau 2. (Même légende que pour le tableau 3).

Modèles	RV	X <sup>2</sup>	Degrés de liberté	φ (écart-type de φ)
Gumbel 1	192.44	215.62	15	1.47(0.025)
Gumbel 2	15.62	15.17	15	1.43(0.022)
Frank	177.12	227.10	15	0.05(0.006)
Clayton 1	41.09	41.95	15	2.31(0.060)
Clayton 2	354.40	504.45	15	2.33(0.077)
Plackett	142.36	170.66	15	4.42(0.242)
Goodman	78.97	82.86	15	4.12

étude, seul celui de Gumbel2 s’ajuste convenablement aux données avec une p-valeur calculée approximativement à 50 %. La valeur estimée par la méthode du maximum de vraisemblance du paramètre d’association φ dans la copule de Gumbel est 1.43 avec un écart-type de 0.022; d’où un intervalle de confiance (1.39; 1.47) pour φ avec

un niveau de confiance de 95%. Faisons remarquer que l'hétérogénéité constatée dans le tableau 4 entre les valeurs RV et  $X^2$  pour les modèles de Gumbel1, Frank, Clayton2, Plackett et Goodman s'explique par le fait que ces modèles n'ajustent pas bien les données.

## 8. Conclusion

Nous avons présenté dans ce travail une collection de cinq modèles pouvant être utilisés dans l'analyse des tableaux de contingence dont les variables sont ordinales. D'autres modèles peuvent s'y ajouter, par exemple, modéliser le paramètre d'association  $\varphi$  par une fonction bilinéaire,  $\varphi_{ij} = \sum_{k=1}^m \mu_i \nu_j$ .

## Annexe

Nous supposons que les fréquences observées  $m_{ij}$  suivent une loi multinomiale avec des probabilités correspondantes  $h_{ij}$ . La fonction du logarithme de vraisemblance sera équivalente à

$$l(\varphi) = \sum_{i=1}^a \sum_{j=1}^b m_{ij} \log h_{ij}$$

où  $h_{ij}$  est une fonction de la fonction de répartition  $H_{ij}$ , c'est-à-dire,

$$\begin{aligned} h_{11} &= H_{11} \\ h_{1j} &= H_{1j} - H_{1,j-1} && \text{pour } j = 2, \dots, b \\ h_{i1} &= H_{i1} - H_{i-1,j} && \text{pour } i = 2, \dots, a \\ h_{ij} &= H_{ij} + H_{i-1,j-1} - H_{i,j-1} - H_{i-1,j} && \text{pour } i = 2, \dots, a \text{ et } j = 2, \dots, b \end{aligned}$$

Le but principal est de maximiser  $l(\varphi)$  en fonction de  $\varphi$ , où  $0 < \varphi < 1$ . Nous avons appliqué une méthode de recherche directe en sous-divisant l'intervalle  $(0, 1)$  en plus petits sous-intervalles, disons  $L$  et en calculant la valeur de  $l(\varphi)$  pour différentes valeurs de  $\varphi$  représentant le centre de ces sous-intervalles  $L$ . Nous notons enfin que dans le cas où  $\varphi > 1$ , ce qui est vrai par exemple pour le copule de Gumbel, nous reparamétrisons  $\varphi$  par  $1/\varphi$ .

## Références

- AGRESTI A. (1984), *Analysis of Ordinal Data*, Wiley, N.Y.  
 AGRESTI A., COULL B.A. (1998), Order-restricted inference for monotone trend alternatives in contingency tables, *Computational Statistics and Data Analysis*, 28, 2, 139-156.

- BIRCH M.W. (1964), A new proof of the Pearson-Fisher theorem, *Annals of Mathematical Statistics*, 35, 818-824.
- BISHOP Y.M.M., FIENBERG S.E., HOLLAND P.W. (1975), Discrete Multivariate Analysis, *Theory and Practice*, Cambridge, Mass, MIT Press.
- CLAYTON D.G. (1978), A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence, *Biometrika*, 65, 141-151.
- COOK R.D., JOHNSON M.E. (1981), A family of distributions for modelling non-elliptically symmetric multivariate data, *Journal of the Royal Statistical Society*, ser. B., 43, 210-218.
- COX C. (1984), An elementary introduction to maximum likelihood estimation for multinomial models : Birch's theorem and the delta method, *The American Statistician*, 38, 283-287.
- DALE J.R. (1984), Local versus global association for bivariate ordered responses, *Biometrika*, 71, 507-514.
- FRANK M.J. (1979), On the simultaneous associativity of  $F(x, y)$  and  $x+y-F(x, y)$ , *Aequationes Mathematicae*, 19, 194-226.
- GENEST C. (1987), Frank's family of bivariate distributions, *Biometrika*, 74, 549-555.
- GENEST C., MacKay R.J. (1986), Copules Archimédiennes et familles de lois bidimensionnelles dont les marges sont données, *The Canadian Journal of Statistics*, 14, 145-159.
- GENEST C., RIVEST L.P. (1993), Statistical inference procedures for bivariate Archimedean copulas, *Journal of the American Statistical Association*, 88, 1034-1043.
- GLASS D.V., HALL J.R. (1954), Social mobility in Great Britain : A study of inter-generation changes in status, In D.V. Glass (Editor), *Social Mobility in Britain*, 177-217, London : Routledge and Kegan Press.
- GOODMAN L. A. (1979), Simple models for the analysis of cross-classifications having ordered categories, *Journal of the American Statistical Association*, 74, 537-552.
- GOODMAN L. A. (1981), Association models and the bivariate normal for contingency tables with ordered categories, *Biometrika*, 68, 347-355.
- GUMBEL E.J. (1960), Bivariate exponential distributions, *Journal of the American Statistical Association*, 55, 698-707.
- JEFFERSON T.R., MAY J.H., RAVI N. (1989), An entropy approach to the scaling of ordinal categorical data, *Psychometrika*, 54, 2, 203-215.
- HOUGAARD P. (1986), A class of multivariate failure time distribution, *Biometrika*, 73, 671-678, (Correction, 75, 395).
- HUTCHINSON T.P. (1979), Four applications of a bivariate Pareto distribution, *Biometrical Journal*, 21, 553-563.

- HUTCHINSON T.P., LAI C.D. (1990), Continuous Bivariate Distributions, Emphasising Applications, Rumsby Scientific Publishing : Australia.
- LAPP K., MOLENBERGHS G., LESAFFRE E. (1998), Models for the association between ordinal variables, *Computational Statistics & Data Analysis*, 28, 4, 387-412.
- LEHMANN E.L. (1966), Some concepts of dependence, *Annals of Mathematical Statistics*, 37, 1137-1153.
- NGUYEN T.T., SAMPSON A.R. (1987), Testing for positive quadrant dependence in ordinal contingency tables, *Naval Research Logistics*, 34, 859-877.
- NISHISATO S., Arri P.S. (1975), Nonlinear programming approach to optimal scaling of partially ordered categories, *Psychometrika*, 40, 525-548.
- OGBORN J. (1984), Cumulative and local odds-ratio measures of uniformity of association in two-way ordered contingency tables, *British Journal of Mathematical and Statistical Psychology*, 37, 89-99.
- PLACKETT R.L. (1965), A class of bivariate distributions, *Journal of the American Statistical Society*, 60, 516-522.
- RAYNER J.C.W., BEST D.J. (1996), Smooth extensions of Pearson's product moment correlation and Spearman's rho, *Statistics & Probability Letters*, 30, 171-177.
- SHRIEVER B.F. (1986), Order Dependence, CWI-Tract 20, *Centre for Mathematics and Computer Science*, Amsterdam.
- SROLE L., LANGNER T.S., MICHAEL S.T., OPLER M.K., RENNIE T.A.C. (1974), Mental Health in the Metropolis, The Midtown Manhattan Study, Revised edition, N.Y., McGraw-Hill.
- WAHRENDORF J. (1980), Inference in contingency tables with ordered categories using Plackett's coefficient of association for bivariate distributions, *Biometrika*, 67, 15-21.