RAIRO. RECHERCHE OPÉRATIONNELLE

I. C. LERMAN

Indices d'association partielle entre variables « qualitatives nominales »

RAIRO. Recherche opérationnelle, tome 17, nº 3 (1983), p. 213-259

<http://www.numdam.org/item?id=RO_1983__17_3_213_0>

R.A.I.R.O. Recherche opérationnelle/Operations Research
(vol. 17, n° 3, août 1983, p. 213 à 259)

INDICES D'ASSOCIATION PARTIELLE ENTRE VARIABLES « QUALITATIVES NOMINALES » (*)

par I. C. LERMAN [1]

Résumé. — *Pour construire des indices d'association partielle entre deux variables qualitatives* β *et* γ, *neutralisant l'influence d'une troisième variable qualitative* α, *les tentatives les plus connues ignorent la nature combinatoire du problème et, pour se baser sur des notions géométriques, plongent de façon plus ou moins arbitraire des échelles discrètes dans des échelles numériques.*

Au contraire, nous proposons une représentation mathématique fidèle ayant un caractère ensembliste des variables α, β *et* γ.

La démarche générale consiste alors à réduire un lien brut $s(\beta, \gamma)$ *par rapport à une situation aléatoire [hypothèse d'absence de lien (h. a. l.)] où à* β *(resp.* γ*), on associe une variable aléatoire* β' *(resp.* γ'*) définissant sur l'ensemble des individus le même type de structure que* β *(resp.* γ*) et dont la position par rapport à* α *est préservée.*

Trois formes fondamentales de l'h. a. l. sont analysées; chacune conduit à un calcul de la variance — de la variable aléatoire associée à $s(\beta, \gamma)$ *— et à une notion d'indépendance conditionnelle qui lui est spécifique.*

Mots clés : Corrélation généralisée, χ^2, Analyse non linéaire des données, Classification sous contrainte de variables.

Abstract. — *The most known approaches ignore the combinatorial nature of the problem concerning the construction of coefficients of partial association between two qualitative variables* β *and* γ, *neutralizing the effect of a third qualitative variable* α. *In fact they are based on geometrical considerations and plunge the discreet scale into a numerical one, in a more or less arbitrary way.*

Contrarily, we propose a faithful mathematical representation having set theoretic character.

Then, the general procedure consists in reducing a rough index of association $s(\beta, \gamma)$ *with respect to a random situation [hypothesis of non link (h. n. l)] where to the variable* β *(resp.* γ*), is associated a random variable* β' *(resp.* γ'*) defining the same type of structure on the set of individuals as* β *(resp.* γ*) and whose position relatively to* α *is preserved.*

Three fundamentals forms of the h. n. l. are analysed; each of which leads to a specific computation of the variance — of the random variable associated to $s(\beta, \gamma)$ *— and to a specific notion of conditional independance.*

Keywords: Generalized correlation, χ^2, Non linear analysis of data, Restricted classification of variables.

(*) Reçu mai 1982.
([1]) I.R.I.S.A., Laboratoire de Statistique, Université de Rennes-I, 35042 Rennes Cedex.

I. INTRODUCTION

Cet article est dans sa conception parallèle à celui qui vient de paraître dans les *Publications de l'Institut de Statistique des Universités de Paris*. Si ce dernier traite du cas où l'échelle descriptive et discrète des valeurs définie par une même variable est ordinale (cas « qualitatif ordinal » où la variable définit un préordre total sur l'ensemble des individus et cas « rang » où la variable définit un ordre total et strict sur l'ensemble des individus), l'article que nous proposons ici est concentré sur le cas « nominal » où la variable définit une partition sur l'ensemble des sujets.

Il est dans ces conditions naturel que ce texte emprunte certains points d'introduction et de conclusion dans l'article précité; d'autre part, l'approche et la démarche générale sont les mêmes et correspondent d'ailleurs aux idées qui ont prévalu à l'élaboration d'un indice de proximité ou d'association totale entre variables et qui sont d'ailleurs à la base de notre méthode de classification hiérarchique axée sur la vraisemblance des liens [Lerman (1973), (1976), (1981)]. Toutefois, on ne peut en aucune façon déduire l'analyse et les calculs du cas *disjoint* principalement étudié ici dont l'intérêt propre est entier.

E désignant l'ensemble des individus défini par l'échantillon étudié, nous distinguons en fait deux principaux types d'une variable de description; où pour le premier la variable peut être fidèlement représentée au niveau de E et où, pour le second type, la variable ne peut admettre une représentation mathématique fidèle qu'au niveau de $E \times E$. Dans la première classe définie par le premier grand type de variables, on a exactement l'« attribut descriptif » représentable par une partie de E et la variable quantitative représentable par une pondération sur E. D'autre part, toute variable définissant une relation binaire discrète ou pondérée sur E, rentre dans le cadre de la deuxième classe; ainsi en est-il de la variable qualitative nominale, de celle ordinale (totalement ou partiellement), de celle « rang » et enfin de celle définissant un graphe pondéré sur E.

Il y a lieu de commencer par rappeler notre démarche générale dans l'élaboration d'un coefficient d'association totale entre variables de description statistique. Cette dernière peut être schématisée par le diagramme suivant (cf. *fig.* 1) que nous allons commenter dans deux cas différents : le premier (resp. le second) qui s'inscrit dans le premier (resp. second) type, est celui où les variables sont des attributs (resp. des variables qualitatives nominales ou « partition »).

Dans le premier cas d'illustration α et β sont deux attributs descriptifs dont on désignera par $n(\alpha)$ et $n(\beta)$ les fréquences absolues. A (resp. B) est l'ensemble

$$(\alpha,\beta) \in A \times B \longrightarrow (R(\alpha), R(\beta)) \in \Omega \times \Omega \qquad (1)$$

$$\downarrow$$

$$s = \mathrm{card}(R(\alpha) \cap R(\beta)) \qquad (2)$$

$$\downarrow$$

$$\boxed{\text{h.a.l. "respectant les caractéristiques de cardinalité" de } \alpha \text{ et } \beta}$$

$$\downarrow$$

$$S = \mathrm{card}(R(\alpha') \cap R(\beta')) \qquad (3)$$

$$\downarrow$$

$$Q(\alpha,\beta) = (s - \mathcal{E}(S))/\sigma(S) \qquad (4)$$

$$\downarrow$$

$$P(\alpha,\beta) = Pr^{N}\{S < s\} \cong \Phi(Q(\alpha,\beta)). \qquad (5)$$

Figure 1

des parties de E de même cardinal $n(\alpha)$ [resp. $n(\beta)$]. $R(\alpha)$ [resp. $R(\beta)$] est confondu avec l'ensemble $E(\alpha)$ [resp. $E(\beta)$] des individus possèdant l'attribut α (resp. β). Ω est l'ensemble des parties de E.

L'hypothèse d'absence de lien (h. a. l.), passage du niveau (2) au niveau (3), associe au couple (α, β) un couple (α', β') d'éléments aléatoires indépendants où α' (resp. β') est un objet aléatoire de même nature que α (resp. β) qui respecte d'une « certaine façon » les caractéristiques cardinales de α (resp. β). Dans le cas de la comparaison de deux parties $E(\alpha)$ et $E(\beta)$ d'un même ensemble E (représentant en l'occurrence deux attributs descriptifs) nous avons pu mettre en évidence trois formes fondamentales de l' h. a. l. respectant chacune d'une certaine façon les caractéristiques cardinales de $E(\alpha)$ et de $E(\beta)$ [Lerman (1981)].

La première forme N_1 a un caractère combinatoire; elle consiste à associer à une même partie D de cardinal d, une partie aléatoire D' qui représente un élément aléatoire dans l'ensemble, muni d'une probabilité uniformément répartie des parties de E de même cardinal d. Cette forme de l' h. a. l. conduit à une distribution hypergéométrique de la v. a. S [*cf.* niveau (3) du schéma].

Alors que dans le cas que nous venons de mentionner toute la mesure de probabilité est concentrée sur un même niveau de l'ensemble des parties de E, la deuxième forme N_2 de l' h. a. l. consiste à munir cet ensemble de parties d'une mesure de probabilité plus diffuse où le k-ième niveau, formé des parties à k éléments, est affecté de la probabilité $\binom{n}{k}\delta^k(1-\delta)^k$ où $\delta=d/n$. D'autre part, cette dernière probabilité est également répartie sur ce niveau. Cette deuxième forme de l' h. a. l. conduit à une distribution binomiale de la v. a. S.

La troisième forme N_3 de l' h. a. l. commence par associer à E, l'ensemble aléatoire $\mathscr{E}$ où l'aléa ne concerne que la v. a. entière $\mathscr{N}=\text{card}(\mathscr{E})$ qui est considérée comme une v. a. de Poisson de paramètre $n=\text{card}(E)$. Relativement à une réalisation $\mathscr{E}_0$ de cardinal m, les deux autres pas du modèle de choix correspondent à N_2.

Dans le deuxième exemple choisi α et β sont deux partitions où nous désignons par $t(\alpha)$ [resp. $t(\beta)$] le type de la partition α(resp. β); c'est-à-dire la suite ordonnée des cardinaux de ses classes. Dans ces conditions A (resp. B) est l'ensemble des partitions sur E de type $t(\alpha)$ [resp. $t(\beta)$]. $R(\alpha)$ [resp. $R(\beta)$] est l'ensemble des paires [sous-ensemble de l'ensemble $P_2(E)$ des parties à deux éléments de E] dont les deux composantes sont réunies dans une même classe de la partition α(resp. β) Ω peut être défini comme l'ensemble des parties de $P_2(E)$ dont chacune correspond à la représentation d'une relation d'équivalence sur E.

Dans cet exemple une des formes de l' h. a. l. consiste à associer à α(resp. β) un élément aléatoire dans l'ensemble A (resp. B) des partitions sur E, muni d'une probabilité uniformément répartie.

$\mathscr{E}(S)$ et $\sigma^2(S)$ sont la moyenne et la variance de la variable aléatoire (v. a) S et $Q(\alpha, \beta)$ est ce que nous appelons l'indice de proximité « centré réduit ». Le passage de (3) à (4) est décisif puisque cet indice est celui d'association de K. Pearson s'il s'agit de comparer (dans le cadre de l' h. a. l. N_1) deux attributs de description; à un coefficient multiplicatif près, celui de corrélation dans le cas de comparaison de deux variables numériques. D'autre part, dans le cas de la comparaison de deux variables « rang », le numérateur de Q est exactement le numérateur de l'indice τ de M. G. Kendall; mais dans notre indice nous remplaçons le dénominateur de τ qui est le maximum de la valeur absolue du numérateur par l'écart-type de la v. a. associée au numérateur. Toutefois, nous découvrons des indices nouveaux dans le cas de la comparaison de variables qualitatives nominales (variables « partition ») et ordinales (variables « préordre total ») où nous mettons en évidence un biais dans l'indice proposé pour cette situation par M. G. Kendall [Lerman (1973) repris dans (1981)]. D'autre part, comme dans le cas de la comparaison des attributs, il y a lieu de découvrir de nouveaux indices en considérant, pour les variables qualitatives, la forme N_3 de l' h. a. l..

Ce qui va changer dans la comparaison entre deux variables α et β, partielle relativement à la variable γ, se situe exactement entre les étapes (2) et (3) du schéma de la figure 1, au niveau de la définition de l' h. a. l.. La variable aléatoire α' (resp. β') évoluera dans un espace tel que sa position relative par rapport à γ restera invariante.

Pour terminer cette introduction signalons qu'une des motivations source de ce travail est que les tentatives connues ignorent complètement la nature combinatoire du problème. Ainsi, considérons le cas où α, β et γ sont des variables qualitatives nominales et sans craindre aucune ambiguïté, désignons par α, β et γ les trois partitions qu'elles définissent respectivement sur l'ensemble E :

$$\alpha=\{E_i/1\leqq i\leqq I\},$$
$$\beta=\{F_j/1\leqq j\leqq J\} \qquad \text{et} \qquad \gamma=\{G_k \mid 1\leqq k\leqq K\}.$$

Relativement à un couple de telles variables, (α, β) par exemple, introduisons les proportions :

$$p(i\wedge j)=\text{card}(E_i\cap F_j)/\text{card}(E)$$
$$p(i)=\text{card}(E_i)/\text{card}(E)$$

et :

$$p(j)=\text{card}(F_j)/\text{card}(E), \qquad 1\leqq i\leqq I, \quad 1\leqq j\leqq J.$$

En posant :

$$\varphi^2_{\alpha\beta}=\sum\{[p^2(i\wedge j)/p(i)p(j)]/1\leqq i\leqq I,\ 1\leqq j\leqq J\}-1,$$

le carré du coefficient de Tschuprow entre les variables α et β, se met sous la forme :

$$T_{\alpha\beta}=\varphi^2_{\alpha\beta}/\sqrt{(I-1)(J-1)}.$$

G. Saporta [Saporta (1976)] montre qu'on peut définir $\varphi^2_{\alpha\beta}$ comme le produit de deux opérateurs de normes respectives $\sqrt{I-1}$ et $\sqrt{J-1}$. $T_{\alpha\beta}$ est de la sorte un indice de liaison entre variables qualitatives nominales pouvant correspondre au coefficient de corrélation usuel entre variables numériques. Par analogie, G. Saporta propose alors le coefficient $T_{\alpha\beta.\gamma}$ suivant comme indice de liaison partielle entre α et β relativement à γ

$$T_{\alpha\beta.\gamma}=\frac{T_{\alpha\beta}-T_{\alpha\gamma}T_{\beta\gamma}}{\sqrt{1-T^2_{\alpha\gamma}}\sqrt{1-T^2_{\beta\gamma}}}.$$

J. Daudin [Daudin (1979)] met en évidence le grave défaut de ce dernier indice de ne pas être nul en cas d'indépendance conditionnelle où pour toute valeur de γ, α et β sont des variables indépendantes.

Signalons tout de suite que les indices que nous proposerons n'auront pas cet inconvénient majeur.

Ce que propose M. G. Kendall [Kendall (1942), *cf.* dans (1970)] pour la comparaison partielle de variables « rang »; entre ω et ϖ, relativement à θ, est dans son principe essentiellement différent et correspond formellement à commencer par représenter chacun des ordres totaux ω, ϖ et σ par son graphe $R(\omega)$, $R(\varpi)$ et $R(\sigma)$ dans $E \times E$; où par exemple :

$$R(\sigma) = \{ (x, y)/(x, y) \in E \times E, x < y(\sigma) \},$$

avec des expressions analogues pour $R(\omega)$ et $R(\varpi)$. Il se situe ensuite dans $R(\sigma)$ et définit dans ce cadre le coefficient d'association de K. Pearson entre les deux attributs représentés par les deux parties $R(\sigma) \cap R(\omega)$ et $R(\sigma) \cap R(\varpi)$. Notre analyse dans [Lerman (1982)] nous montre que de la sorte, on ne tient pas assez étroitement compte de la nature des structures des variables à comparer.

Dans ces conditions et pour des raisons d'ailleurs de meilleure compréhension générale, nous commencerons par l'étude des deux situations suivantes qui ont un intérêt propre :

— Association entre deux attributs, partielle relativement à un troisième attribut.

— Association entre deux variables numériques partielle relativement à un attribut ou à une variable qualitative nominale, puis nous analyserons à travers trois formes de l' h.a.l. la comparaison partielle de variables « partition ».

II. ASSOCIATION ENTRE DEUX ATTRIBUTS, PARTIELLE RELATIVEMENT A UN TROISIÈME ATTRIBUT

On désigne par a, b et c trois attributs descriptifs et par E l'ensemble des individus défini par l'échantillon étudié. Soient $E(a)$, $E(b)$ et $E(c)$ les parties de E représentant respectivement a, b et c; en d'autres termes $E(a)$ est le sous-ensemble de E formé des sujets qui possèdent l'attribut a, $E(b)$ celui b et $E(c)$ celui c. Nous schématisons la situation par le diagramme naïf suivant :

Il s'agit de définir un indice d'association partielle entre b et c, neutralisant l'influence de a. Nous nous proposons de le faire à partir d'une construction adéquate de l' h.a.l. de même nature que N_1 (*cf.* § I ci-dessus).

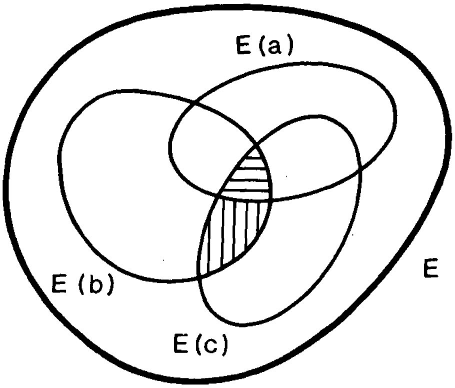

Figure 2

L'indice brut d'association entre b et c a toujours la même expression :

$$s(b, c) = n(b \wedge c) = \text{card}(E(b) \cap E(c)), \qquad (1)$$

Comme dans le cas total l' h. a. l. peut avoir une forme unilatérale, en fixant par exemple $E(b)$ et en associant à $E(c)$ une partie aléatoire Z. Pour neutraliser l'influence de $E(a)$, cette h. a. l. doit préserver les positions relatives observées entre $E(a)$ et $E(b)$ d'une part, $E(a)$ et $E(c)$ d'autre part. Comme $E(a)$ et $E(b)$ sont fixées, la partie aléatoire Z doit avoir par rapport à $E(a)$, la même situation relative que celle qu'avait $E(c)$.

Dans ces conditions, Z sera choisie comme formée de la réunion de deux parties aléatoires indépendantes $Z(a)$ et $Z(\bar{a})$ où $Z(a)$ [resp. $Z(\bar{a})$] est un élément aléatoire dans l'ensemble, muni d'une probabilité uniformément répartie, des parties de $E(a)$ [resp. $E(\bar{a})$] de même cardinal $n(a \wedge c)$ [resp. $n(\bar{a} \wedge c)$], où nous avons noté $\bar{a}$ l'attribut opposé de a; de sorte que $E(\bar{a}) = E - E(a)$ (différence ensembliste).

La v. a. S se met alors sous la forme suivante d'une somme de deux v. a. indépendantes :

$$S(b, c'; a) = \text{card}(E(b) \cap Z(a)) + \text{card}(E(b) \cap Z(\bar{a})), \qquad (2)$$

Il est facile de se rendre compte que la première (resp. seconde) v. a. est hypergéométrique de paramètres $[n(a), n(a \wedge b), n(a \wedge c)]$ [resp. $(n(\bar{a}), n(\bar{a} \wedge b), n(\bar{a} \wedge c))$]. On détermine dans ces conditions la moyenne et la variance de $S(b, c'; a)$. En confondant $[n(a) - 1]$ avec $n(a)$ [resp. $(n(\bar{a}) - 1)$ avec

$n(\bar{a})]$, on obtient par centrage et réduction le coefficient d'association partielle $\rho(b, c; a)$ qui, au facteur $\sqrt{n}$ près prend la forme suivante :

$$\rho(b, c; a) = \frac{[p(b \wedge c) - ((p(a \wedge b)\,p(a \wedge c)/p(a)) + (p(\bar{a} \wedge b)\,p(\bar{a} \wedge c)/p(\bar{a})))]}{\sqrt{(p(a \wedge b)\,p(a \wedge \bar{b})\,p(a \wedge c)\,p(a \wedge \bar{c})/p(a)^3) + (p(\bar{a} \wedge b)\,p(\bar{a} \wedge \bar{b})\,p(\bar{a} \wedge c)\,p(\bar{a} \wedge \bar{c})/p(\bar{a})^3)}} \tag{3}$$

où une expression de la forme $p(a)$ [resp. $p(a \wedge b)$] représente la proportion $n(a)/n$ [resp. $n(a \wedge b)/n$] et où $\bar{a}$ désigne l'attribut opposé à a.

Il est naturel de chercher à rapprocher l'expression qu'on vient de trouver pour l'indice $\rho(b, c; a)$ avec celle :

$$\frac{\rho_{bc} - \rho_{ab}\,\rho_{ac}}{\sqrt{(1-\rho_{ab}^2)(1-\rho_{ac}^2)}} \tag{4}$$

qu'on aurait obtenue en procédant par analogie directe entre le cas où les variables sont numériques et celui qui nous occupe ici. On peut en effet se rendre compte que le numérateur de l'indice (3) trouvé se met, dans le cadre de l' h. a. l. considérée, à un coefficient multiplicatif près, sous la forme $(\rho_{bc} - \rho_{ab}\,\rho_{ac})$. Cependant, on n'a pas une propriété analogue de rapprochement en ce qui concerne le dénominateur.

En effet, en rappelant qu'une expression du type ρ_{bc} s'écrit :

$$\rho_{bc} = \frac{(p(b \wedge c) - p(b)\,p(c))}{\sqrt{p(b)\,p(c)\,p(\bar{b})\,p(\bar{c})}}, \tag{5}$$

on obtient en développant le numérateur de l'expression (3) :

$$\text{num}.\,(\rho(b, c; a)) = p(b)\,p(\bar{b})\,p(c)\,p(\bar{c})\,(\rho_{bc} - \rho_{ab}\,\rho_{ac}). \tag{6}$$

Mais, il est inutile de poursuivre le rapprochement de :

$$\frac{\rho(b, c; a)}{\rho_{bc} - \rho_{ab}\,\rho_{ac}} = \left\{ \frac{p(b)\,p(\bar{b})\,p(c)\,p(\bar{c})}{p(a \wedge b)\,p(a \wedge \bar{b})\,p(a \wedge c)\,p(a \wedge \bar{c})/p(a)^3) + p(\bar{a} \wedge b)\,p(\bar{a} \wedge \bar{b})\,p(\bar{a} \wedge c)\,p(\bar{a} \wedge \bar{c})/p(\bar{a})^3} \right\}^{1/2} \tag{7}$$

avec :

$$1/\{(1-\rho_{ab}^2)(1-\rho_{ac}^2)\}^{1/2}; \tag{8}$$

en effet, l'exemple numérique suivant :

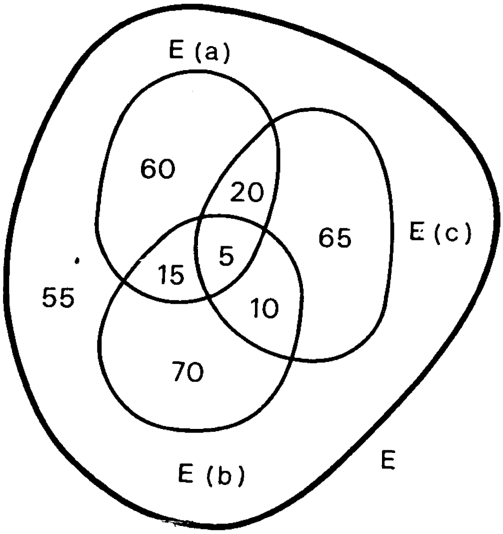

Figure 3

donne, respectivement, les valeurs 0,047 5 et 0,0466 6... pour les expressions (7') et (8') ci-dessous dont l'identification est équivalente à celle de (7) et (8) ci-dessus :

$$\frac{p(a \wedge b)p(a \wedge \bar{b})p(a \wedge c)p(a \wedge \bar{c})}{p(a)^3} + \frac{p(\bar{a} \wedge b)p(\bar{a} \wedge \bar{b})p(\bar{a} \wedge c)p(\bar{a} \wedge \bar{c})}{p(\bar{a})^3} \tag{7'}$$

et :

$$\left\{p(b)p(\bar{b}) - \frac{[p(a \wedge b)-p(a)p(b)]^2}{p(a)p(\bar{a})}\right\} \times \left\{p(c)p(\bar{c}) - \frac{[p(a \wedge c)-p(a)p(c)]^2}{p(a)p(\bar{a})}\right\} \tag{8'}$$

Cet aspect sera également considéré au paragraphe suivant.

PROPOSITION : *Le coefficient d'association partielle* $\rho(b, c; a)$ *peut se mettre à un facteur multiplicatif près qui ne dépend que des liens entre a et b d'une part, a et c d'autre part, sous la forme* $(\rho_{bc}-\rho_{ab}\,\rho_{ac})$. *Le facteur* (7) *ci-dessus ne se réduit pas comme dans le cas linéaire et numérique au coefficient* (8) *ci-dessus.*

Enfin, compte tenu de la manière dont il a été construit, le comportement asymptotique de $\sqrt{n}\,\rho(b, c; a)$ *dans l'h. a. l. est celui d'une v. a. normale centrée réduite.*

III. COEFFICIENT D'ASSOCIATION ENTRE DEUX VARIABLES NUMÉRIQUES PARTIELLE, RELATIVEMENT A UN ATTRIBUT OU A UNE VARIABLE QUALITATIVE NOMINALE

III. A. Cas où la liaison est relative à un attribut

1. INDICE BRUT ET H. A. L.

$I=\{1, 2, \ldots, i, \ldots, n\}$ désignera l'ensemble d'indexation de l'ensemble E des individus.

$\{v(i)/i\in I\}$ [resp. $\{w(i)/i\in I\}$] désignera la suite des valeurs de la première (resp. seconde) variable numérique. On note a l'attribut de description dont il s'agit de neutraliser l'influence dans la comparaison entre v et w; $a(i)=1$ (resp. 0) si l'attribut est présent (resp. absent) chez l'individu codé i. On indiquera par $I(a)$ [resp. $I(\bar{a})$] l'ensemble des indices des sujets possédant (resp. ne possédant pas) l'attribut a; on note ainsi $\bar{a}$ l'attribut opposé à a.

Comme dans le cas total l'indice brut de proximité se met sous la forme :

$$\sum\{v(i)\,w(i)/i\in I\} \tag{1}$$

qui ici, peut se décomposer comme suit :

$$\sum\{v(i)\,w(i)/i\in I(a)\}+\sum\{v(i)\,w(i)/i\in I(\bar{a})\}, \tag{2}$$

$$=\sum_{i\in I} v(i)\,w(i)\,a(i)+\sum_{i\in I} v(i)\,w(i)\,\bar{a}(i). \tag{3}$$

Nous avons pu mettre en évidence que pour passer du cas discret au cas quantitatif dans l'élaboration de l'hypothèse d'absence de lien, il y avait lieu de substituer à l'expression ensembliste, une expression permutationnelle [Lerman (1976)]; alors que dans le cas discret l'élément aléatoire est un ensemble, il s'agira dans le cas numérique d'une permutation.

La permutation aléatoire qu'il s'agit de construire ici doit être telle que les positions relatives entre v et a d'une part, w et a d'autre part, soient préservées.

En d'autres termes, la suite des quatre quantités suivantes doit être invariante après avoir opéré sur I une telle permutation.

$$\{\sum a(i)\,v(i), \sum \bar{a}(i)\,v(i), \sum a(i)\,w(i), \sum \bar{a}(i)\,w(i)/i \in I\}. \tag{4}$$

La permutation σ cherchée sera par conséquent formée d'un couple $(\sigma_a, \sigma_{\bar{a}})$ de deux permutations aléatoires indépendantes, où σ_a (resp. $\sigma_{\bar{a}}$) est un élément aléatoire dans l'ensemble, muni d'une probabilité uniformément répartie, des permutations sur $I(a)$ [resp. $I(\bar{a})$].

On pourra facilement montrer que, comme dans le cas total, les deux v. a. duales suivantes associées à l'indice brut (1), sont de même loi asymptotiquement normale :

$$\left.\begin{aligned} &X(v, w') = \sum \{v(i)\,w(\sigma(i))/i \in I\} \\ \text{et :}\quad & \\ &X(v', w) = \sum \{v(\sigma(i))\,w(i)/i \in I\}. \end{aligned}\right\} \tag{5}$$

Reprenons la première des deux v. a., elle peut se mettre sous la forme :

$$X(v, w') = \sum \{v(i)\,w(\sigma_a(i))/i \in I(a)\} + \sum \{v(i)\,w(\sigma_{\bar{a}}(i))/i \in I(\bar{a})\}. \tag{6}$$

Le calcul de la moyenne et de la variance de cette dernière v. a. permet d'obtenir, en centrant et en réduisant l'indice brut, le coefficient d'association partielle $\sigma(v, w; a)$. Comme c'est le cas pour la comparaison des attributs de description, le numérateur de cet indice peut se réduire à la forme $(\rho(v, w) - \rho(a, v)\,\rho(a, w)]$ où ρ est le coefficient de corrélation; mais le dénominateur de cet indice a une forme essentiellement différente de $\sqrt{(1-\rho^2(a, v))(1-\rho^2(a, w))}$.

2. Moyenne et variance de la v.a. $X(v, w')$

Les deux permutations aléatoires σ_a et $\sigma_{\bar{a}}$ étant indépendantes, $X(v, w')$ se présente sous la forme d'une somme de deux v. a. indépendantes qu'on peut d'ailleurs noter $X(v, w'_a)$ et $X(v, w'_{\bar{a}})$. On a :

$$\mathscr{E}[X(v, w'_a)) = \sum_{i \in I(a)} v(i)\,\mathscr{E}\{w[\sigma_a(i)]\} \tag{7}$$

mais, pour tout i :

$$\mathscr{E}\{w[\sigma_a(i)]\} = \frac{1}{n(a)} \sum_{i' \in I(a)} w(i'). \tag{8}$$

$$\mathscr{E}[X(v, w'_a)] = n(a) \times \left[\frac{1}{n(a)} \sum_{i \in I(a)} v(i)\right] \times \left[-\frac{1}{n(a)} \sum_{i \in I(a)} w(i)\right]$$

$$= n(a)\,\mu_a(v)\,\mu_a(w); \tag{9}$$

où, on le comprend, $\mu_a(v)$ [resp. $\mu_a(w)$] représente la moyenne sur $I(a)$ de la variable numérique v (resp. w).

De même :

$$\mathscr{E}[X(v, w'_{\bar{a}})] = n(\bar{a})\,\mu_{\bar{a}}(v)\,\mu_{\bar{a}}(w), \tag{10}$$

où $\mu_{\bar{a}}(v)$ [resp. $\mu_{\bar{a}}(w)$] représente la moyenne sur $I(\bar{a})$ de la variable numérique v (resp. w).

Finalement :

$$\mathscr{E}[X(v, w')] = n(a)\,\mu_a(v)\,\mu_a(w) + n(\bar{a})\,\mu_{\bar{a}}(v)\,\mu_{\bar{a}}(w). \tag{11}$$

Compte tenu de relations de la forme :

$$\sum_{i \in I} v(i)\,w(i) = n\,\mathrm{cov}(v, w) + n\,\mu(v)\,\mu(w)$$

et :

$$\sum_{i \in I(a)} v(i) = \sum_{i \in I} a(i)\,v(i),$$

l'indice centré peut s'écrire comme suit :

$$n\,\mathrm{cov}(v, w) + n\,\mu(v)\,\mu(w) - \frac{1}{n(a)}[n\,\mathrm{cov}(a, v) + n(a)\,\mu(v)]$$

$$\times [n\,\mathrm{cov}(a, w) + n(a)\,\mu(w)]$$

$$-\frac{1}{n(\bar{a})}[n\,\mathrm{cov}(\bar{a}, v) + n(\bar{a})\,\mu(v)]\,[n\,\mathrm{cov}(\bar{a}, w) + n(\bar{a})\,\mu(w)].$$

En développant cette expression et en tenant compte des relations de type $\operatorname{cov}(\bar{a}, v) = -\operatorname{cov}(a, v)$ et $n(a) + n(\bar{a}) = n$, on obtient le résultat suivant :

$$= n\left[\operatorname{cov}(v, w) - \frac{1}{p(a)\, p(\bar{a})} \operatorname{cov}(a, v) \operatorname{cov}(a, w)\right],$$

où :
$p(a) = n(a)/n$ [resp. $p(\bar{a})] = n(\bar{a})/n$),

$$= ns(v)\, s(w) [\rho(v, w) - \rho(a, v)\, \rho(a, w)], \quad (12)$$

où $s^2(v)$ [resp. $s^2(w)$] désigne la variance sur I de la variable numérique v (resp. w) et où $\rho(a, v)$ [resp. $\rho(a, w)$ et $\rho(v, w)$] désigne le coefficient de corrélation entre a et v (resp. entre a et w et entre v et w). Ainsi :

$$\rho(a, v) = \frac{\operatorname{cov}(a, v)}{\sqrt{\operatorname{var}(a) \operatorname{var}(v)}} = \frac{(1/n) \sum_{1 \leq i \leq n} a(i)\, v(i) - p(a)\, \mu(v)}{\sqrt{p(a)\, p(\bar{a})\, s^2(v)}}. \quad (13)$$

Pour le calcul de la variance, commençons par rappeler que si τ est un élément aléatoire dans l'ensemble, muni d'une probabilité uniforme, de toutes les permutations sur I, la variance d'une v. a. de la forme $\sum \{ \xi_i \eta_{\sigma(i)} \mid i \in I \}$ s'écrit $(n^2/(n-1)) \operatorname{var}(\xi) \operatorname{var}(\eta)$. Dans ces conditions et en tenant compte de la nature de la variable a, on peut écrire :

$$\begin{aligned} \operatorname{var}[X(v, w'_a)] = \frac{n(a)^2}{[n(a)-1]} & \left\{ \frac{1}{n(a)} \sum_{i \in I} v^2(i)\, a^2(i) - \left[\frac{1}{n(a)} \sum_{i \in I} v(i)\, a(i) \right]^2 \right\} \left\{ \frac{1}{n(a)} \sum_{i \in I} w^2(i)\, a^2(i) \right. \\ & \left. - \left[\frac{1}{n(a)} \sum_{i \in I} w(i)\, a(i) \right]^2 \right\} \quad (14) \end{aligned}$$

$$\begin{aligned} \operatorname{var}[X(v, w'_{\bar{a}})] = \frac{n(\bar{a})^2}{[n(\bar{a})-1]} & \left\{ \frac{1}{n(\bar{a})} \sum_{i \in I} v^2(i) [1 - a^2(i)] - \left(\frac{1}{n(\bar{a})} \sum_{i \in I} v(i) [1 - a(i)] \right)^2 \right\} \left\{ \frac{1}{n(\bar{a})} \sum_{i \in I} w^2(i) [1 - a^2(i)] \right. \\ & \left. - \left(\frac{1}{n(\bar{a})} \sum_{i \in I} w(i) [1 - a(i)] \right)^2 \right\}. \quad (15) \end{aligned}$$

Ici encore, il est tentant de chercher à rapprocher l'expression de notre indice d'association partielle avec celui qu'on aurait obtenu dans le cas linéaire et numérique [*cf.* formule (4), § II ci-dessus].

Compte tenu de la formule (12) ci-dessus, il y a lieu de comparer :

$$\{\operatorname{var}[X(v, w'_a)] + \operatorname{var}[X(v, w'_{\bar{a}})]\}/ns^2(v)\,s^2(w) \tag{16}$$

avec :

$$[1-\rho^2(a, v)][1-\rho^2(a, w)] \tag{17}$$

où, rappelons-le ici, une expression de type $\rho(a, v)$ peut prendre la forme $[\mu(a \wedge v) - p(a)\,\mu(v)]/\sqrt{p(a)\,p(\bar{a})\,\mu(v)\,\mu(\bar{v})}$.

L'exemple considéré dans le cas où v et w sont des attributs (*cf.* § précédent) montre déjà que le rapprochement formel entre les expressions (16) et (17) ne peut aller jusqu'à l'identification. En effet, la situation considérée ici peut apparaître comme une généralisation du cas précédent qu'on retrouverait ici en supposant que les deux variables v et w ne sont susceptibles de prendre que les deux valeurs 0 et 1.

Nous allons néanmoins chercher à rapprocher au mieux les expressions de (16) et de (17), ce qui nous permettra de nous rendre compte de leur différence formelle. Pour ce calcul, on supposera sans restreindre la généralité que les variables v et w sont centrées réduites.

Dans ces conditions (17) peut se mettre sous la forme

$$\left[p(a)\,p(\bar{a}) - \left(\frac{1}{n}\langle a, v\rangle\right)^2\right]\left[p(a)\,p(\bar{a}) - \left(\frac{1}{n}\langle a, w\rangle\right)^2\right]\Big/[p(a)\,p(\bar{a})]^2, \tag{18}$$

où on a noté $\langle a, v\rangle$ une expression du type $\sum\{a_i v_i / i \in I\}$. Alors que l'expression (16) ne peut se réduire qu'à la forme suivante :

$$\begin{aligned}\frac{1}{[p(a)\,p(\bar{a})]^2}\Bigg\{&\frac{(p(\bar{a}))^2}{p(a)}\left[p(a)\frac{1}{n}\langle a^2, v^2\rangle - \left(\frac{1}{n}\langle a, v\rangle\right)^2\right]\\ &\times\left[p(a)\times\frac{1}{n}\langle a^2, w^2\rangle - \left(\frac{1}{n}\langle a, w\rangle\right)^2\right]\\ &+\frac{(p(a))^2}{p(\bar{a})}\left[p(\bar{a})\left(1-\frac{1}{n}\langle a^2, v^2\rangle\right) - \left(\frac{1}{n}\langle a, v\rangle\right)^2\right]\\ &\times\left[p(\bar{a})\left(1-\frac{1}{n}\langle a^2, w^2\rangle\right) - \left(\frac{1}{n}\langle a, w)\right)^2\right]\Bigg\};\end{aligned} \tag{19}$$

où on a noté $\langle a^2, v^2 \rangle$ une expression de type $\sum \{ a_i^2 v_i^2 / i \in I \}$ qui est nécessairement inférieure à n.

PROPOSITION : *Le coefficient d'association partielle entre les deux variables numériques v et w, relativement à l'attribut a, peut se mettre à un coefficient multiplicatif près qui ne dépend que des liens entre a et v d'une part, a et w d'autre part, sous la forme* $(\rho_{vw} - \rho_{av} \rho_{aw})$. *Ce coefficient dont le carré est défini par l'expression* (16), *ne peut se réduire à* $\sqrt{(1-\rho_{av}^2)(1-\rho_{aw}^2)}$ [cf. *formules* (18) *et* (19) *ci-dessus*]. *D'autre part, dans l'h. a. l. à caractère permutationnel, le comportement asymptotique de l'indice est celui d'une v. a. centrée réduite.*

La dernière assertion de cette proposition résulte de l'application du théorème de Wald, Wolfowitz et Nœther [Wald et Wolfowitz (1944)], [Nœther (1949)].

III. B. Cas où la liaison est conditionnée par une variable qualitative nominale

Le cas où la liaison est conditionnée par une variable qualitative nominale est une généralisation naturelle du cas précédent. Désignons par $a^1, a^2, \ldots, a^j, \ldots, a^J$, les attributs-modalités de la variable qualitative a dont il s'agit de neutraliser l'influence dans la comparaison entre deux variables quantitatives numériques v et w. On a, pour tout individu x, $a^j(x) = 1$ (resp. 0) selon que x possède (resp. ne possède pas) la modalité a^j du caractère; on notera I_j l'ensemble des indices des sujets possédant la modalité a^j, $1 \leqq j \leqq J$.

Comme dans le cas total, l'indice brut de proximité se met sous la forme :

$$\sum \{ v(i)\, w(i) / i \in I \} \tag{1'}$$

et les deux v. a. duales, respectivement, sous la forme :

$$\sum \{ v(i)\, w(\sigma(i)) / i \in I \}, \sum \{ v(\tau(i))\, w(i) / i \in I \} \tag{2'}$$

où σ (resp. τ) est un élément aléatoire d'un espace probabilisé de permutations, judicieusement choisi.

En effet, l'h. a. l. à caractère permutationnel doit préserver les positions relatives des deux variables sur chaque I_j, $1 \leqq j \leqq J$. Dans ces conditions, décomposons l'expression (1) de l'indice brut comme suit :

$$\sum_{1 \leqq j \leqq J} \sum_{i \in I_j} v(i)\, w(i). \tag{3'}$$

Relativement à l'une ou à l'autre des deux v. a. duales [*cf.* (2')], on peut par exemple prendre la première, σ doit être choisie comme une suite

$(\sigma^1, \sigma^2, \ldots, \sigma^j, \ldots, \sigma^J)$ de J permutations aléatoires indépendantes, où σ^j est un élément aléatoire dans l'ensemble, muni d'une probabilité uniforme, de toutes les permutations sur I_j, $1 \leqq j \leqq J$. Ainsi, la première v. a. de (2′) se met sous la forme :

$$\sum_{1 \leqq j \leqq J} \sum_{i \in I_j} v(i)\, w(\sigma^j(i)). \tag{4′}$$

C'est sans aucune difficulté qu'on voit que la distribution de la deuxième v. a. :

$$\sum_{1 \leqq j \leqq J} \sum_{i \in I_j} v(\tau^j(i))\, w(i) \tag{5′}$$

est, avec une définition correspondante de $(\tau^1, \tau^2, \ldots, \tau^j, \ldots, \tau^J)$, la même que celle de la première.

Les calculs de la moyenne et de la variance de la somme (4′) de v. a. indépendantes sont en tout point analogues à ceux où la variable à neutraliser est un attribut...

IV. ASSOCIATION ENTRE DEUX VARIABLES QUALITATIVES NOMINALES, PARTIELLE RELATIVEMENT A UNE TROISIÈME VARIABLE QUALITATIVE NOMINALE

IV. 1. Cas d'un indice d'association conforme à la statistique du χ^2

Nous désignerons par β et γ les deux variables partitions dont il s'agit de mesurer le lien en neutralisant l'effet d'une troisième variable partition α. Les partitions P, Q et R respectivement induites par α, β et γ sur l'ensemble E des objets de cardinal n (défini par l'échantillon étudié) seront désignées comme suit :

$$\left.\begin{aligned} P &= \{ E_1, E_2, \ldots, E_i, \ldots, E_I \} \\ Q &= \{ F_1, F_2, \ldots, F_j, \ldots, F_J \} \\ \text{et :} \quad R &= \{ G_1, G_2, \ldots, G_k, \ldots, G_K \} \end{aligned}\right\} \tag{1}$$

Nous noterons également :

$$n(i) = \mathrm{card}(E_i), \qquad n(j) = \mathrm{card}(F_j), \qquad n(k) = \mathrm{card}(G_k),$$
$$p(i) = n(i)/n, \qquad p(j) = n(j)/n \qquad \text{et} \qquad p(k) = n(k)/n,$$
$$1 \leqq i \leqq I, \qquad 1 \leqq j \leqq J \qquad \text{et} \qquad 1 \leqq k \leqq K.$$

La position relative entre deux partitions est déterminée par le tableau de contingence de leur croisement. Dans ces conditions, on aura à considérer les trois tableaux de contingence suivants, respectivement associés à $P \wedge Q$, à $P \wedge R$ et à $Q \wedge R$:

$$\left.\begin{aligned} n(P \wedge Q) &= \{n(i \wedge j)/1 \leqq i \leqq I,\ 1 \leqq j \leqq J\} \\ n(P \wedge R) &= \{n(i \wedge k)/1 \leqq i \leqq I,\ 1 \leqq k \leqq K\} \\ n(Q \wedge R) &= \{n(j \wedge k)/1 \leqq j \leqq J,\ 1 \leqq k \leqq K\} \end{aligned}\right\} \quad (2)$$

où :

$$n(i \wedge j) = \mathrm{card}(E_i \cap F_j), \qquad n(i \wedge k) = \mathrm{card}(E_i \cap G_k)$$

et :

$$n(j \wedge k) = \mathrm{card}(F_j \cap G_k), \qquad 1 \leqq i \leqq I, \quad 1 \leqq j \leqq J, \quad 1 \leqq k \leqq K.$$

On aura également à nous exprimer par rapport aux tableaux des proportions :

$$\left.\begin{aligned} p(P \wedge Q) &= \{p(i \wedge j)/1 \leqq i \leqq I,\ 1 \leqq j \leqq J\} \\ p(P \wedge R) &= \{p(i \wedge k)/1 \leqq i \leqq I,\ 1 \leqq k \leqq K\} \end{aligned}\right.$$

et :

$$p(Q \wedge R) = \{p(j \wedge k)/1 \leqq j \leqq J,\ 1 \leqq k \leqq K\}, \quad (3)$$

où :

$$p(i \wedge j) = n(i \wedge j)/n, \qquad p(i \wedge k) = n(i \wedge k)/n$$

et :

$$p(j \wedge k) = n(j \wedge k)/n; \qquad 1 \leqq i \leqq I, \quad 1 \leqq i \leqq J \quad \text{et} \quad 1 \leqq k \leqq K.$$

Nous noterons enfin $\chi^2_{\beta\gamma.\alpha}$ l'indice d'association partielle cherché que nous allons obtenir en procédant par analogie formelle et statistique avec une certaine interprétation du cas total dans la comparaison de deux variables partitions.

Dans le cas partiel qui nous intéresse, la construction va se jouer sur la manière de choisir le couple de parties aléatoires (F'_j, G'_k), $1 \leqq j \leqq J$, $1 \leqq k \leqq K$. Ce choix doit se faire de telle façon que les positions relatives entre P et Q d'une part, P et R d'autre part, soient préservées lorsqu'on remplace la suite des parties $(F_1, F_2, \ldots, F_j, \ldots, F_J)$ [resp. $G_1, G_2, \ldots, G_k, \ldots, G_K$)] par celle $(F'_1, F'_2, \ldots, F'_j, \ldots, F'_J)$ [resp. $(G'_1, G'_2, \ldots, G'_k, \ldots, G'_K)$].

En se référant aux deux décompositions suivantes :

$$F_j = \sum_{1 \leqq i \leqq I} F_j \cap E_i \qquad \text{et} \qquad G_k = \sum_{1 \leqq i \leqq I} G_k \cap E_i, \quad (4)$$

$1 \leqq j \leqq J$, $1 \leqq k \leqq K$, on se rend compte que dans le cas partiel, le choix global de F'_j (resp. G'_k) se fera à partir d'une suite $\{F'_{ji}/1 \leqq i \leqq I\}$ (resp. $\{G'_{ki}/1 \leqq i \leqq I\}$) de parties aléatoires indépendantes conformément à l'h. a. l.. N_3. Les paramètres du choix de F'_{ji} (resp. G'_{ki}) sont $n(i)$ et $p(i \wedge j)/p(i)$ [resp. $n(i)$ et $p(i \wedge k)p(i)$], $1 \leqq i \leqq I$, $1 \leqq j \leqq J$ et $1 \leqq k \leqq K$.

Dans ces conditions, la v. a. $\operatorname{card}(F'_j \cap G'_k)$ se met sous la forme :

$$\operatorname{card}(F'_j \cap G'_k) = \operatorname{card}((\sum_i F'_{ji}) \cap (\sum_i G'_{ki})) \quad \text{(sommes ensemblistes)}$$

$$= \sum_{1 \leqq i \leqq I} \operatorname{card}(F'_{ji} \cap G'_{ki}) \quad (5)$$

puisque le choix de F'_{ji} (resp. G'_{ki}) se fait relativement à E_i et que les différents E_i sont mutuellement disjoints.

Chacune des v. a. de la forme $\operatorname{card}(F'_{ji} \cap G'_{ki})$ est une v. a. de Poisson de paramètre :

$$\frac{n(i)\, p(i \wedge j)\, p(i \wedge k)}{(p(i))^2} = \frac{np(i \wedge j)\, p(i \wedge k)}{p(i)}. \quad (6)$$

L'expression (5) se présente comme une somme de v. a. indépendantes de Poisson. Il s'agit donc d'une v. a. de Poisson dont le paramètre est égal à la somme des paramètres des v. a. composantes.

Dans ces conditions, l'indice « centré réduit » associé à la case (j, k) du croisement entre β et γ dans l'h. a. l. du cas partiel où l'effet de la variable α est neutralisé, se met sous la forme :

$$\left.\begin{array}{c} \dfrac{n(j \wedge k) - \sum\limits_{1 \leqq i \leqq I} (np(i \wedge j)\, p(i \wedge k)/p(i)}{\sqrt{\sum\limits_{1 \leqq i \leqq I} (np(i \wedge j)\, p(i \wedge k)/p(i))}}, \\ 1 \leqq j \leqq J, \quad 1 \leqq k \leqq K, \end{array}\right\} \quad (7)$$

de sorte que la somme pour $1 \leqq j \leqq J$, $1 \leqq k \leqq K$, des carrés des expressions (7), peut être considérée dans l'h. a. l. qu'on vient d'exprimer, comme une réalisation d'une v. a. du χ^2 à $(J-1)(K-1)$ degrés de liberté (vérification laissée au lecteur). Dans ces conditions, l'indice $\chi^2_{\beta\gamma.\alpha}$ aura pour expression :

$$\sum \left\{ \frac{(n(j \wedge k) - \sum\limits_{1 \leqq i \leqq I} (np(i \wedge j)\, p(i \wedge k)/p(i)))^2}{\sum\limits_{1 \leqq i \leqq I} (np(i \wedge j)\, p(i \wedge k)/p(i))} \Big/ 1 \leqq j \leqq J,\ 1 \leqq k \leqq K \right\} \quad (8)$$

L'indépendance conditionnelle entre β et γ, relativement à α se trouve caractérisée par les relations suivantes :

$$p(j \wedge k/i) = p(j/i)\,p(k/i) \quad \text{pour tout } 1 \leqq i \leqq I, \quad 1 \leqq j \leqq J \quad \text{et} \quad 1 \leqq k \leqq K. \tag{9}$$

Ces relations se mettent aussi sous la forme :

$$p(j \wedge k) = p(i \wedge j)\,p(i \wedge k)/p(i) \quad \text{pour tout } 1 \leqq i \leqq I, \quad 1 \leqq j \leqq J \quad \text{et} \quad 1 \leqq k \leqq K. \tag{10}$$

On voit immédiatement que l'indice (8) est nul dans le cas de l'indépendance conditionnelle.

Réciproquement si $\chi^2_{\beta\gamma.\alpha}$ est nul, c'est pour tout (j, k), en moyenne par rapport à la distribution $\{p(i)/1 \leqq i \leqq I\}$, qu'on a la relation :

$$p(j \wedge k/i) = p(j/i)\,p(k/i). \tag{11}$$

Nous laissons le soin au lecteur de récapituler les résultats du présent paragraphe au moyen de l'énoncé d'un théorème.

IV.2. Cas d'un indice d'association partielle conforme a l'indice de Lerman dans le cadre d'une h. a. l. a caractère local

1. INDICE BRUT ET H. A. L.

L'indice brut est le même que dans le cas total [Lerman (1973) *cf.* dans (1981)] il prend la forme suivante :

$$s(\beta, \gamma) = \mathrm{card}(\mathscr{R}(Q) \cap \mathscr{R}(R)) = \sum \left\{ \frac{1}{2}\, n(j \wedge k)(n(j \wedge k) - 1)/1 \leqq j \leqq J,\ 1 \leqq k \leqq K \right\} \tag{1}$$

où $\mathscr{R}(Q)$ [resp. $\mathscr{R}(R)$] est l'ensemble des paires d'objets réunies par la partition Q (resp. R) et où, rappelons-le, $n(j \wedge k)$ est le cardinal de l'intersection de la j-ième classe F_j de la partition Q et de la k-ième classe G_k de la partition R.

Comme dans le cas total, on fixe l'une des deux partitions Q ou R et on associe à l'autre une partition aléatoire en montrant que la distribution de la v. a. associée à $s(\beta, \gamma)$ ne dépend pas de celle des deux partitions fixées. On travaillera avec des partitions en classes étiquetées.

Nous allons fixer la partition Q et associer à la partition R, une partition aléatoire R' notée $\{G'_k/1 \leqq k \leqq K\}$ où G'_k est la partie aléatoire associée à G_k, $1 \leqq k \leqq K$, dans une h. a. l. qui préserve la position relative par rapport aux classes de la partition P. Dans ces conditions, G'_k sera défini sous la forme de la somme ensembliste :

$$G'_k = \sum_{1 \leqq i \leqq I} G'_{ki} \tag{2}$$

pour tout $k = 1, 2, \ldots, K$, où $(G'_{k1}, G'_{k2}, \ldots, G'_{ki}, \ldots, G'_{kI})$ est une suite de parties aléatoires indépendantes, G'_{ki} étant choisi de cardinal $n(i \wedge k)$, uniformément au hasard dans E_i; en d'autres termes, G'_{ki} est un élément aléatoire dans l'ensemble, muni d'une probabilité uniforme, des parties de E_i de même cardinal $n(i \wedge k)$.

Désignons par φ (resp. ψ et ψ') la fonction indicatrice de $\mathscr{R}(Q)$ [resp. $\mathscr{R}(R)$ et $\mathscr{R}(R')$]. Avec ces notations, la v. a. $S(\beta, \gamma')$ s'exprime comme suit :

$$S(\beta, \gamma') = \mathrm{card}(\mathscr{R}(Q) \cap \mathscr{R}(R')) = \sum \{ \varphi(p)\, \psi'(p) / p \in P_2(E) \}. \tag{3}$$

2. CALCUL DE LA MOYENNE DE LA V. A. $S(\beta, \gamma'; \alpha)$

Le calcul de la moyenne (resp. variance ou de n'importe quel moment) va devoir se décomposer relativement à une partition de l'ensemble $P_2(E)$ des paires (i. e. des parties à deux éléments) de E, conforme à la partition $P = \{ E_1, E_2, \ldots, E_i, \ldots, E_I \}$. Cette partition de $P_2(E)$, se met sous la forme :

$$\{ \{ P_2(E_i) / 1 \leqq i \leqq I \}, \{ E_i * E_{i'} / 1 \leqq i < i' \leqq I \} \}, \tag{4}$$

où $E_i * E_{i'}$ désigne l'ensemble des parties à deux éléments dont l'une des composantes appartient à E_i et l'autre à $E_{i'}$.

Ainsi, la v. a. $S(\beta, \gamma'; \alpha)$ se met sous la forme :

$$S(\beta, \gamma'; \alpha) = \sum_{1 \leqq i \leqq I} \sum \{ \varphi(p)\, \psi'(p) / p \in P_2(E_i) \} + \sum_{1 \leqq i < i' \leqq I} \{ \psi(p)\, \psi'(p) / p \in E_i * E_{i'} \} \tag{5}$$

Le calcul de l'espérance mathématique de (5) suppose l'évaluation des expressions suivantes :

$$\Pr \{ \psi'(p) = 1 / p \in P_2(E_i) \} \, [\text{resp.} \sum \{ \varphi(p) / p \in P_2(E_i) \}] \tag{6}$$

et :

$$\Pr \{ \psi'(p) = 1 / p \in E_i * E_{i'} \} \, [\text{resp.} \sum \{ \varphi(p) / p \in E_i * E_{i'} \}]. \tag{7}$$

Les expressions de gauche représentent des proportions, celles de droite des cardinaux de même forme que les numérateurs des rapports qui définissent les proportions.

Désignant par R_i (resp. R'_i) la restriction de la partition R (resp. aléatoire R') à E_i; ainsi R'_i se trouve définie par :

$$R'_i = \{ G'_{1i}, G'_{2i}, \ldots, G'_{Ki} \} \tag{8}$$

de type $\{ n(i \wedge k)/1 \leqq k \leqq K \}$.

(6) est défini par la proportion de partitions R'_i de type fixé qui préservent les deux composantes de la paire p dans une même classe G'_{ki}, $1 \leqq k \leqq K$. Nous ne reprendrons pas ici ce calcul (*cf.* [Lerman (1973)]), pour lequel on obtient :

$$\Pr\{ \psi'(p) = 1/p \in P_2(E_i) \} = \sum \left\{ \frac{n(i \wedge k)[n(i \wedge k) - 1]}{n(i)[n(i) - 1]} \middle/ 1 \leqq k \leqq K \right\} \tag{9}$$

d'autre part :

$$\sum \{ \varphi(p)/p \in P_2(E_i) \}$$
$$= \sum \left\{ \frac{1}{2} n(i \wedge j)[n(i \wedge j) - 1]/1 \leqq j \leqq J \right\} \tag{10}$$

Cette dernière expression représente en effet le nombre de paires réunies par la partition $Q_i = \{ F_{ji}/1 \leqq j \leqq J \}$ qui est la restriction de Q sur E_i et dont le type est défini par $\{ n(i \wedge j)/1 \leqq j \leqq J \}$.

De même, on a :

$$\Pr\{ \psi'(p) = 1/p \in E_i * E_{i'} \} = \sum \left\{ \frac{n(i \wedge k)}{n(i)} \times \frac{n(i' \wedge k)}{n(i')} \middle/ 1 \leqq k \leqq K \right\}. \tag{11}$$

Il s'agit en effet de la proportion de couple de parties $(G'_{ki}, G'_{ki'})$ telles que G'_{ki} (resp. $G'_{ki'}$) comprenne la composante de p qui se trouve dans E_i (resp. $E_{i'}$).

Enfin :

$$\sum \{ \varphi(p)/p \in E_i * E_{i'} \} = \sum \{ n(i \wedge j)\, n(i' \wedge j)/1 \leqq j \leqq J \}; \tag{12}$$

de sorte que, compte tenu de (5) :

$$\mathscr{E}(S(\beta, \gamma'; \alpha))$$
$$= \sum_{1 \leqq i \leqq I} \sum \left\{ \frac{n(i \wedge j)[n(i \wedge j) - 1]\, n(i \wedge k)[n(i \wedge k) - 1]}{2\, n(i)[n(i) - 1]} \middle/ 1 \leqq j \leqq J, 1 \leqq k \leqq K \right\}$$
$$+ \sum_{1 \leqq i < i' \leqq I} \sum \left\{ \frac{n(i \wedge j)\, n(i' \wedge j)\, n(i \wedge k)\, n(i' \wedge k)}{n(i)\, n(i')} \middle/ 1 \leqq j \leqq J, 1 \leqq k \leqq K \right\}. \tag{13}$$

3. CALCUL DE LA VARIANCE DE LA V.A. $S(\beta, \gamma'; \alpha)$

La difficulté réside dans l'évaluation du moment absolu d'ordre 2 de $S(\beta, \gamma'; \alpha)$ qui nécessite le calcul, pour (p, q) fixé, d'expressions de la forme $\mathscr{E}[\psi'(p)\,\psi'(q)]$. Or le calcul d'une telle expression dépend :

— de la structure propre du couple de paires $(p, q) = (\{x, y\}, \{x', y'\})$ selon que les composantes de p se répètent ou non dans celles de q.

— de la structure propre du couple de paires d'indices $(\{i, i'\}, \{i'', i'''\})$ pour lesquels $(\{x, y\}, \{x', y'\})$ appartient à $(E_i * E_{i'}) \times (E_{i''} * E_{i'''})$ où on admet que i' (resp. i''') peut être égal à i (resp. i''); auquel cas $E_i * E_{i'}$ (resp. $E_{i''} * E_{i'''}$) devient $P_2(E_i)$ [resp. $P_2(E_{i''})$].

Ainsi, le calcul du carré de $S(\beta, \gamma'; \alpha)$ va devoir se décomposer selon le croisement de deux partitions de $P_2(E) \times P_2(E)$ où la première est définie à partir de la structure de (p, q) et où la deuxième correspond au carré cartésien de (4) ci-dessus. Dans ces conditions, on verra directement que sur une même classe de la partition croisée $\sum \varphi(p)\,\varphi(q)$ aura une forme analytique tout à fait analogue au numérateur de la proportion qui définit $\mathscr{E}[\psi'(p)\,\psi'(q)]$. Précisons à présent le détail de chacune des deux partitions :

(*a*) Nous avons déjà vu dans le cas de la comparaison totale entre partitions qu'un couple de paires (p, q) peut avoir trois formes différentes selon que :

— les deux composantes de p se répètent dans q;

— une des deux composantes de p se répète dans q;

— aucune des composantes de p ne se répète dans q.

Dans ces conditions, en indiquant par des lettres différentes des objets différents, désignons respectivement par Δ, G et H l'ensemble des couples de paires de la forme $(\{x, y\}, \{x, y\})$, $(\{x, y\}, \{x, t\})$ et $(\{x, y\}, \{z, t\})$. On a :

$$P_2(E) \times P_2(E) = \Delta + G + H \text{ (somme ensembliste).} \tag{14}$$

D'autre part :

$$\operatorname{card}(\Delta) = n(n-1)/2, \qquad \operatorname{card}(G) = n(n-1)(n-2)$$

et :

$$\operatorname{card}(H) = n(n-1)(n-2)(n-3)/4. \tag{15}$$

(*b*) Conformément à ce que nous venons d'exprimer ci-dessus, nous commencerons par la décomposition suivante :

$P_2(E) \times P_2(E) = A + B + C + D$

où :

$$\left.\begin{aligned}
A&=\Big[\sum_{1\leqq i\leqq I} P_2(E_i)\Big]\times\Big[\sum_{1\leqq i'\leqq I} P_2(E_{i'})\Big]\\
B&=\Big[\sum_{1\leqq i\leqq I} P_2(E_i)\Big]\times\Big[\sum_{1\leqq i'<i''\leqq I} E_{i'}*E_{i''}\Big]\\
G&=\Big[\sum_{1\leqq i'<i''\leqq I} E_{i'}*E_{i''}\Big]\times\Big[\sum_{1\leqq i\leqq I} P_2(E_i)\Big]\\
D&=\Big[\sum_{1\leqq i<i'\leqq I} E_i*E_{i'}\Big]\times\Big[\sum_{1\leqq i''<i'''\leqq I} E_{i''}*E_{i'''}\Big]
\end{aligned}\right\}\qquad(16)$$

et nous allons affiner cette décomposition de telle sorte que sur l'intersection de chacun des sous-ensembles de la décomposition avec G ou H, la valeur de $\mathscr{E}[\psi'(p)\,\psi'(q)]=\Pr\{\psi'(p)\,\psi'(q)=1\}$ reste constante. On posera :

$$\left.\begin{aligned}
&A=A_1+A_2,\\
&\text{où :}\\
&A_1=\sum_{1\leqq i\leqq I} P_2(E_i)\times P_2(E_i),\\
&A_2=\sum_{1\leqq i\neq i' I} P_2(E_i)\times P_2(E_{i'}),\\
&B=B_1+B_2,\\
&\text{où :}\\
&B_1=\sum_{1\leqq i\neq i' I} P_2(E_i)\times(E_i*E_{i'}),\\
&B_2=\sum_{1\leqq i\leqq I} P_2(E_i)\sum\{E_{i'}*E_{i''}/1\leqq i'<i''\leqq I,\ i'\neq i,\ i''\neq i\},\\
&C=C_1+C_2,\\
&\text{où :}\\
&C_1=\sum_{1\leqq i\neq i'\leqq I}(E_i*E_{i'})\times P_2(E_{i'})\\
&C_2=\sum_{1\leqq i'<i''\leqq I}(E_{i'}*E_{i''})\times\sum\{P_2(E_i)/1\leqq i\leqq I,\ i\neq i',\ i\neq i''\},\\
&D=D_1+D_2,\\
&\text{où :}\\
&D_1=\sum\{(E_i*E_{i'})\times(E_i*E_{i''})/(\{i,i'\},\{i,i''\})\in G(I)\},\\
&D_2=\sum\{(E_i*E_{i'})\times(E_{i''}*E_{i'''})/(\{i,i'\},\{i'',i'''\})\in H(I)\}
\end{aligned}\right\}\qquad(17)$$

où nous avons noté $G(I)$ [resp. $H(I)$] l'ensemble des couples de paires d'indices avec (resp. sans) une composante commune.

On n'aura guère à se préoccuper de la décomposition de Δ selon la partition ci-dessus; en effet, en commençant par décomposer $S^2(\beta, \gamma'; \alpha)$, conformément à la partition $\{\Delta, G, H\}$, on obtient :

$$S^2(\beta, \gamma'; \alpha) = \sum \{\varphi(p)\,\psi'(p)/p \in P_2(E)\} + \sum \{\varphi(p)\,\varphi(q)\,\psi'(p)\,\psi'(q)/(p, q) \in G\} + \sum \{\varphi(p)\,\varphi(q)\,\psi'(p)\,\psi'(q)/(p, q) \in H\}. \quad (18)$$

Pour ce qui est de la décomposition de G, on notera que les intersections de G avec certaines des classes de la partition (17), sont vides; c'est exactement le cas pour :

$$G \cap A_2, \quad G \cap B_2, \quad G \cap C_2 \quad \text{et} \quad G \cap D_2.$$

Il reste donc à effectuer les calculs suivants :

3.*g*.*a*. 1. *calcul de* $\mathscr{E}[\psi'(p)\,\psi'(q)/(p, q) \in G \cap A_1]$

(p, q) est un couple de paires de la forme $(\{x, y\}, \{x, z\})$ où les trois objets distincts x, y et z sont dans une même classe E_i. *Connaissant* i, l'espérance cherchée représente la proportion de partitions $\{G'_{ki}/1 \leqq k \leqq K\}$ pour lesquelles les trois objets restent réunis dans une même classe G'_{ki}. Cette proportion vaut :

$$\sum_{1 \leqq k \leqq K} \frac{n(i \wedge k)[n(i \wedge k) - 1][n(i \wedge k) - 2]}{n(i)[n(i) - 1][n(i) - 2]}. \quad (19)$$

3.*g*.*a*. 1'. *calcul de* $\sum \{\varphi(p)\,\varphi(q)/(p, q) \in G \cap A_1\}$

Il s'agit d'une énumération; en se cantonnant à une même classe E_i, la quantité cherchée représente le nombre de couples de paires de la forme $(\{x, y\}, \{x, z\})$ pour lesquels x, y et z se trouvent réunis dans une même classe F_{ji} de la partition Q_i qui représente la restriction de la partition Q à E_i. Le nombre cherché s'écrit :

$$\sum_{1 \leqq j \leqq J} n(i \wedge j)[n(i \wedge j) - 1][n(i \wedge j) - 2]. \quad (19')$$

3.g.b.1. calcul de $\mathscr{E}[\psi'(p)\,\psi'(q)/(p, q) \in G \cap B_1\}$

Pour (i, i') fixé et relativement au couple de paires $(p, q) = (\{x, y\}, \{x, z\})$, il s'agit de la proportion de couples de partitions $(\{G'_{ki}/1 \leqq k \leqq K\}, \{G'_{ki'}/1 \leqq k \leqq K\})$ qui d'une part, préservent x et y dans la

k-ième classe de E_i et qui d'autre part, préservent z dans la k-ième classe de $E_{i'}$, $1 \leqq k \leqq K$; cette proportion vaut par conséquent :

$$\sum_{1 \leqq k \leqq K} \frac{n(i \wedge k)[n(i \wedge k)-1]\, n(i' \wedge k)}{n(i)[n(i)-1]\, n(i')}. \tag{20}$$

3.g.b.1'. *calcul de* $\sum \{ \varphi(p)\, \varphi(q)/(p, q) \in G \cap B_1 \}$

Pour (i, i') fixé $(i \neq i')$, il s'agit du nombre de couples de paires de la forme $(\{x, y\}, \{x, z\})$ tels que d'une part, x et y appartiennent à une même classe F_{ji} et d'autre part, z appartienne à la classe $F_{ji'}$. Ce cardinal est égal à :

$$\sum_{1 \leqq j \leqq J} n(i \wedge j)[n(i \wedge j)-1]\, n(i' \wedge j). \tag{21}$$

3.g.c.1. *calcul de* $\mathscr{E}[\psi'(p)\, \psi'(q)/(p, q) \in G \cap C_1]$

Ce calcul est analogue en tout point à celui ci-dessus de 3.g.b.1. Le résultat est donc pour i et i' fixés $(i \neq i')$:

$$\sum_{1 \leqq k \leqq K} \frac{n(i \wedge k)\, n(i' \wedge k)[n(i' \wedge k)-1]}{n(i)\, n(i')[n(i')-1]}. \tag{22}$$

3.g.c.1'. *calcul de* $\sum \{ \varphi(p)\, \varphi(q)/(p, q) \in G \cap C_1 \}$

Ce calcul est analogue à celui 3.g.b.1' ci-dessus. On obtient pour (i, i') fixé $(i' \neq i)$:

$$\sum_{1 \leqq j \leqq J} n(i \wedge j)\, n(i' \wedge j)[n(i' \wedge j)-1]. \tag{23}$$

3.g.d.1. *calcul de* $\mathscr{E}[\psi'(p)\, \psi'(q)/(p, q) \in G \cap D_1]$

Pour i, i' et i'' fixés, il s'agit relativement au couple de paires $(\{x, y\}, \{x, z\})$, de la proportion de triplets de partitions $G'_{ki}/1 \leqq k \leqq K\}$, $\{G'_{ki'}/1 \leqq k \leqq K\}$, $\{G'_{ki''}/1 \leqq k \leqq K\})$ qui, respectivement, préservent les objets x, y et z dans leur k-ième classe, $1 \leqq k \leqq K$; cette proportion vaut par conséquent :

$$\sum_{1 \leqq k \leqq K} \frac{n(i \wedge k)\, n(i' \wedge k)\, n(i'' \wedge k)}{n(i)\, n(i')\, n(i'')}. \tag{24}$$

3.g.d.1'. *calcul de* $\sum \{ \varphi(p)\, \varphi(q)/(p, q) \in G \cap D_1 \}$

i, i' et i'' étant fixés, il s'agit du nombre de couples de paires $(\{x, y\}, \{x, z\})$ tels que pour un même j, $x \in F_{ji}$, $y \in F_{ji'}$ et $z \in F_{ji''}$. Le cardinal cherché est donc égal à :

$$\sum_{1 \leq j \leq J} n(i \wedge j)\, n(i' \wedge j)\, n(i'' \wedge j). \tag{25}$$

3.*h*.*a*.1. *calcul de* $\mathscr{E}[\psi'(p)\, \psi'(q)/(p, q) \in H \cap A_1]$

Pour i fixé et relativement à un couple de paires de la forme $(\{x, y\}, \{x, t\})$, il s'agit de la somme de deux proportions de partitions $\{G'_{ki}/1 \leq k \leq K\}$; celles qui regroupent dans une même classe les quatre objets distincts x, y, z et t et celles qui regroupent les deux premiers objets x et y dans une classe k et les deux derniers objets z et t dans une classe $k' \neq k$. La proportion globale vaut :

$$\Big\{ \sum_{1 \leq k \leq K} n(i \wedge k)[n(i \wedge k)-1][n(i \wedge k)-2][n(i \wedge k)-3] + \sum_{1 \leq k \neq k' \leq K} n(i \wedge k)[n(i \wedge k)-1]\, n(i \wedge k')[n(i \wedge k')-1] \Big\} \Big/ n(i)[n(i)-1][n(i)-2][n(i)-3]. \tag{26}$$

3.*h*.*a*.1'. *calcul de* $\sum \{\varphi(p)\, \varphi(q)/(p, q) \in H \cap A_1\}$

En se cantonnant à une même classe E_i, le cardinal cherché est la somme de deux nombres; celui de couples de paires de la forme $(\{x, y\}, \{z, t\})$ pour lesquels les quatre objets distincts composants se retrouvent dans une même classe F_{ji} de la partition $\{F_{ji}/1 \leq j \leq J\}$ et celui pour lesquels $\{x, y\}$ est incluse dans une classe F_{ji} et $\{z, t\}$, dans une classe $F_{j'i}$ pour $j' \neq j$. Ce cardinal vaut :

$$\frac{1}{4}\Big\{ \sum_{1 \leq j \leq J} n(i \wedge j)[n(i \wedge j)-1][n(i \wedge j)-2][n(i \wedge j)-3] + \sum_{1 \leq j \neq j' \leq J} n(i \wedge j)[n(i \wedge j)-1]\, n(i \wedge j')[n(i \wedge j')-1] \Big\}. \tag{27}$$

3.*h*.*a*.2. *calcul de* $\mathscr{E}[\psi'(p)\, \psi'(q)/(p, q) \in H \cap A_2]$

Pour (i, i') fixé et relativement à un couple de paires de la forme $(\{x, y\}, \{z, t\})$, il s'agit de la proportion de couples de partitions $(\{G'_{ki}/1 \leq k \leq K\}, \{G'_{ki'}/1 \leq k \leq K\})$ qui, respectivement, préservent la paire $\{x, y\}$ pour l'une, $\{z, t\}$ pour l'autre, dans une même classe : $\{x, y\} \subset G'_{ki}$, $\{z, t\} \subset G'_{k'i'}$, $1 \leq k$, $k' \leq K$. Cette proportion vaut :

$$\sum_{1 \leq k,\, k' \leq K} \frac{n(i \wedge k)[n(i \wedge k)-1]\, n(i' \wedge k')[n(i' \wedge k')-1]}{n(i)[n(i)-1]\, n(i')[n(i')-1]}. \tag{28}$$

3.*h*.*a*.2'.. calcul de $\sum \{\varphi(p)\,\varphi(q)/(p,q)\in H \cap A_2\}$

Pour (i, i') fixé, ce cardinal est égal au nombre de paires réunies par la partition $\{F_{ji}/1 \leqq j \leqq J\}$ de E_i, multiplié par le nombre de paires réunies par la partition $\{F_{ji'}/1 \leqq j \leqq J\}$ de $E_{i'}$; sa valeur est alors :

$$\frac{1}{4}\sum_{1\leqq j,j'\leqq J} n(i\wedge j)[n(i\wedge j)-1]\,n(i'\wedge j')[n(i'\wedge j')-1]. \tag{29}$$

3.*h*.*b*.1. *calcul de* $\mathscr{E}[\psi'(p)\,\psi'(q)/(p,q)\in H\cap B_1]$

Pour (i, i') fixé, relativement à un couple de paires $(\{x, y\}, \{z, t\})$, il s'agit de la proportion de couples de partitions $(\{G'_{ki}/1\leqq k\leqq K\}, \{G'_{ki'}/1\leqq k\leqq K)$ pour lesquelles la paire $\{x, y\}$ se trouve réunie dans une même classe G'_{ki}, les objets z et t appartiennent respectivement à deux classes $G'_{k'i}$ et $G'_{k'i'}$ de même indice. Compte tenu du fait que k' peut être égal à k ou différent de k, on obtient :

$$\sum_{1\leqq k\leqq K}\frac{n(i\wedge k)[n(i\wedge k)-1][n(i\wedge k)-2]\,n(i'\wedge k)}{n(i)[n(i)-1][n(i)-2]\,n(i')} + \sum_{1\leqq k\neq k'\leqq K}\frac{n(i\wedge k)[n(i\wedge k)-1]\,n(i\wedge k')\,n(i'\wedge k')}{n(i)[n(i)-1][n(i)-2]\,n(i')}. \tag{30}$$

3.*h*.*b*.1'. *calcul de* $\sum\{\varphi(p)\,\varphi(q)/(p,q)\in H\cap B_1\}$

Il s'agit du nombre de couples de paires de la forme $(\{x, y\}, \{z, t\})$ tels que $\{x, y\}\subset F_{ji}$ pour un $j=1, 2, \ldots, J$ et $\{z, t\}\in F_{j'i} \star F_{j'i'}$ pour un $j'=1, 2, \ldots, J$; i et i' étant fixés. L'énumération doit tenir compte du cas où $j'=j$ et du cas où $j'\neq j$. Dans ces conditions, on obtient pour le cardinal cherché l'expression suivante :

$$\frac{1}{2}\sum_{1\leqq j\leqq J} n(i\wedge j)[n(i\wedge j)-1][n(i\wedge j)-2]\,n(i'\wedge j) + \frac{1}{2}\sum_{1\leqq j\neq j'\leqq J} n(i\wedge j)[n(i\wedge j)-1]\,n(i\wedge j')\,n(i'\wedge j'). \tag{31}$$

3.*h*.*b*.2. *calcul de* $\mathscr{E}[\psi'(p)\,\psi'(q)/(p,q)\in H\cap B_2]$

Cette espérance se réduit, pour i, i' et i'' fixés (mutuellement distincts), à la proportion de triplets de partitions $(\{G'_{ki}/1\leqq k\leqq K\}, \{G'_{ki'}/1\leqq k\leqq K\}, \{G'_{ki''}/1\leqq k\leqq K\})$ pour lesquelles, relativement à un couple de paires $(\{x, y\}, \{z, t\})$ donné, $\{x, y\}$ se trouve incluse dans une même classe

G'_{ki}, z et t appartiennent respectivement à deux classes de même étiquette $G'_{k'i'}$ et $G'_{k'i''}$. On obtient dans ces conditions pour la proportion cherchée :

$$\sum_{1 \leqq k, k' \leqq K} \frac{n(i \wedge k)[n(i \wedge k)-1]\, n(i' \wedge k')\, n(i'' \wedge k')}{n(i)[n(i)-1]\, n(i')\, n(i'')}. \tag{32}$$

3.*h*.*b*.2′. *calcul de* $\sum \{ \varphi(p)\, \varphi(q)/(p, q) \in H \cap B_2 \}$

i, i' et i'' étant fixés mutuellement disjoints, il s'agit du nombre de couples de paires de la forme $(\{x, y\}, \{z, t\})$ où $\{x, y\}$ se trouve réunie dans une même classe de la partition $\{F_{ij}/1 \leqq j \leqq J\}$ et où, respectivement, z et t appartiennent à deux classes de même étiquette j' des deux partitions $\{F_{ji'}/1 \leqq j \leqq J\}$ et $\{F_{ji''}/1 \leqq j \leqq J\}$. Ce cardinal s'exprime donc par la formule :

$$\frac{1}{2} \sum_{1 \leqq j, j' \leqq J} n(i \wedge j)[n(i \wedge j)-1]\, n(i' \wedge j')\, n(i'' \wedge j'). \tag{33}$$

3.*h*.*c*.1. *calcul de* $\mathscr{E}[\psi'(p)\, \psi'(q)/(p, q) \in H \cap C_1]$

Il s'agit d'une situation analogue à celle considérée au paragraphe 3.*h*.*b*.1 ci-dessus à cela près que les rôles de i et de i' sont inversés. Pour (i, i') fixé et relativement à un couple de paires $(\{x, y\}, \{z, t\})$, il s'agit de la proportion de couples de partitions $(\{G'_{ki}/1 \leqq k \leqq K\}, \{G'_{k'i'}/1 \leqq k' \leqq K\})$ pour lesquelles d'une part, x et y appartiennent respectivement à deux classes de même étiquette des deux partitions et d'autre part, $\{z, t\}$ se trouve incluse dans une même classe de la deuxième partition. Cette proportion vaut :

$$\sum_{1 \leqq k \leqq K} \frac{n(i \wedge k)\, n(i' \wedge k)[n(i' \wedge k)-1][n(i' \wedge k)-2]}{n(i)\, n(i')[n(i')-1][n(i')-2]}$$
$$+ \sum_{1 \leqq k \neq k' \leqq K} \frac{n(i \wedge k)\, n(i' \wedge k)\, n(i' \wedge k')[n(i' \wedge k')-1]}{n(i)\, n(i')[n(i')-1][n(i')-2]}. \tag{34}$$

3.*h*.*c*.1′. *calcul de* $\sum \{ \varphi(p)\, \psi(q)/(p, q) \in H \cap G_1 \}$

Ce calcul est analogue à celui 3.*h*.*b*.1′ ci-dessus. On obtient pour ce cardinal la valeur suivante :

$$\frac{1}{2} \sum_{1 \leqq j \leqq J} n(i \wedge j)\, n(i' \wedge j)[n(i' \wedge j)-1][n(i' \wedge j)-2]$$
$$+ \frac{1}{2} \sum_{1 \leqq j \neq j' \leqq J} n(i \wedge j)\, n(i' \wedge j)\, n(i' \wedge j')[n(i' \wedge j')-1]. \tag{35}$$

3 . *h* . *c* . 2. *calcul de* $\mathscr{E}[\psi'(p)\,\psi'(q)/(p,q)\in H\cap C_2]$

Ce calcul est analogue à celui du paragraphe 3.*h*.*b*.2, ci-dessus. Relativement à un couple de paires de la forme $(\{x,y\},\{z,t\})$ et pour i', i'' et i fixés mutuellement distincts, on a pour la valeur de la proportion impliquée ici :

$$\sum_{1\leq k,\,k'\leq K}\frac{n(i'\wedge k)\,n(i''\wedge k)\,n(i\wedge k')[n(i\wedge k')-1]}{n(i')\,n(i'')\,n(i)[n(i)-1]}. \tag{36}$$

3 . *h* . *c* . 2'. *calcul de* $\sum\{\varphi(p)\,\varphi(q)/(p,q)\in H\cap C_2\}$

De la même façon que ci-dessus, ce calcul est analogue à celui du paragraphe 3.*h*.*b*.2'. Le cardinal cherché se met sous la forme :

$$\frac{1}{2}\sum_{1\leq j,\,j'\leq J} n(i'\wedge j)\,n(i''\wedge j)\,n(i\wedge j')[n(i\wedge j')-1]. \tag{37}$$

3 . *h* . *d* . 1. *calcul de* $\mathscr{E}[\psi'(p)\,\psi'(q)/(p,q)\in H\cap D_1]$

On fixe i, i' et i'' mutuellement distincts. Relativement à un couple de paires de la forme $(\{x,y\},\{z,t\})$ et à une attribution de x à E_i, y à $E_{i'}$, z à E_i et t à $E_{i''}$, il s'agit de la proportion de tri-uples de partitions $(\{G'_{ki}/1\leq k\leq K\}$, $\{G'_{ki'}/1\leq k\leq K\}$, $\{G'_{ki''}/1\leq k\leq K\})$ pour lesquelles :

— x et y se retrouvent respectivement dans deux classes de même étiquette des deux partitions $\{G'_{ki}/1\leq k\leq K\}$ et $\{G'_{ki'}/1\leq k\leq K\}$.

— z et t se retrouvent respectivement dans deux classes de même étiquette des deux partitions $\{G'_{ki}/1\leq k\leq K\}$ et $\{G'_{ki''}/1\leq k\leq K\}$.

On aura à distinguer deux cas selon que les deux étiquettes sont les mêmes ou non. On obtient pour la proportion cherchée :

$$\sum_{1\leq k\leq K}\frac{n(i\wedge k)\,n(i'\wedge k)[n(i\wedge k)-1]\,n(i''\wedge k)}{n(i)[n(i)-1]\,n(i')\,n(i'')} + \sum_{1\leq k\neq k'\leq K}\frac{n(i\wedge k)\,n(i'\wedge k)\,n(i\wedge k')\,n(i''\wedge k')}{n(i)[n(i)-1]\,n(i')\,n(i'')}. \tag{38}$$

3 . *h* . *d* . 1'. *calcul de* $\sum\{\varphi(p)\,\varphi(q)/(p,q)\in H\cap D_1\}$

On fixe i, i' et i'' mutuellement distincts. Il s'agit du nombre de couples de paires de la forme $(p,q)=(\{x,y\},\{z,t\})$ tels que d'une part, x et y appartiennent respectivement à deux classes de même étiquette F_{ji} et $F_{ji'}$ et d'autre part, z et t appartiennent à deux classes de même étiquette $F_{j'i}$ et

$F_{j'i''}$. Il y a comme ci-dessus lieu de distinguer selon que $j'=j$ ou $j'\neq j$. Le cardinal cherché se met dans ces conditions sous la forme :

$$\sum_{1\leqq j\leqq J} n(i\wedge j)\,n(i'\wedge j)\,[n(i\wedge j)-1]\,n(i''\wedge j)$$
$$+\sum_{1\leqq j\neq j'\leqq J} n(i\wedge j)\,n(i'\wedge j)\,n(i\wedge j')\,n(i''\wedge j'). \quad (39)$$

3.*h*.*d*.2. *calcul de* $\mathscr{E}[\psi'(p)\,\psi'(q)/(p,q)\in H\cap D_2]$

Les indices i, i', i'' et i''' étant fixés mutuellement distincts, il s'agit relativement à un couple de paires de la forme $(\{x,y\},\{z,t\})$, de la proportion de quatre-uples de partitions $(\{G'_{ki}/1\leqq k\leqq K\}$, $\{G'_{ki'}/1\leqq k\leqq K\}$, $\{G'_{ki''}/1\leqq k\leqq K\}$, $\{G'_{ki'''}/1\leqq k\leqq K\})$ pour lesquelles d'une part, x et y se retrouvent respectivement dans deux classes de même étiquette des deux premières partitions et d'autre part, z et t se retrouvent respectivement dans deux classes de même étiquette des deux dernières partitions. Cette proportion est égale à :

$$\sum_{1\leqq k,\,k'\leqq K}\frac{n(i\wedge k)\,n(i'\wedge k)\,n(i''\wedge k')\,n(i'''\wedge k')}{n(i)\,n(i')\,n(i'')\,n(i''')}. \quad (40)$$

Il faut noter que ce calcul, comme ceux qui précèdent est effectué pour *une* attribution des objets du couple de paires aux différentes classes E_i, $1\leqq i\leqq I$.

3.*h*.*d*.2′. *calcul de* $\sum\{\varphi(p)\,\varphi(q)/(p,q)\in H\cap D_2\}$

Il s'agit du nombre de couples de paires de la forme $(\{x,y\},\{z,t\})$ pour lesquelles d'une part, les deux objets x et y se retrouvent respectivement dans deux classes de même étiquette des deux partitions $\{F_{ji}/1\leqq j\leqq J\}$ et $\{F_{ji'}/1\leqq j\leqq J\}$ et d'autre part, les deux objets z et t se retrouvent respectivement dans deux classes de même étiquette des deux partitions $\{F_{ji''}/1\leqq j\leqq J\}$ et $\{F_{ji'''}/1\leqq j\leqq J\}$. Ce nombre est égal par conséquent à :

$$\sum_{1\leqq j,\,j'\leqq J} n(i\wedge j)\,n(i'\wedge j)\,n(i''\wedge j')\,n(i'''\wedge j'). \quad (40')$$

3.*g*. *calcul de* $\mathscr{E}(\sum\{\varphi(p)\,\varphi(q)\,\psi'(p)\,\psi'(q)/(p,q)\in G\})$

On commencera par décomposer, conformément à (18), la somme sous le signe espérance en huit morceaux respectivement correspondants à A_1, A_2, B_1, B_2, C_1, C_2, D_1 et D_2. Pour fixer les idées, donnons l'exemple du morceau correspondant à B_2 :

$$\sum\{\sum\{\varphi(p)\varphi(q)\psi'(p)\psi'(q)/(p,q)\in G\cap P_2(E_i)\times E_{i'} * E_{i''}\}/$$
$$1\leqq i\leqq I,\ 1\leqq i'<i''\leqq I,\ i'\neq i,\ i''\neq i\}. \quad (41)$$

Compte-tenu des calculs préliminaires précédents [formules (19) à (25)], on obtient :

$$\sum_{1\leqq i\leqq I}\ \sum_{1\leqq j\leqq J}\ \sum_{1\leqq k\leqq K}$$
$$n(i\wedge j)[n(i\wedge j)-1][n(i\wedge j)-2]\,n(i\wedge k)[n(i\wedge k)-1]$$
$$\times[n(i\wedge k)-2]/n(i)[n(i)-1][n(i)-2]$$
$$+\sum_{1\leqq i\neq i'\leqq I}\ \sum_{1\leqq j\leqq J}\ \sum_{1\leqq k\leqq K}$$
$$n(i\wedge j)[n(i\wedge j)-1]\,n(i'\wedge j)\,n(i\wedge k)$$
$$\times[n(i\wedge k)-1]\,n(i'\wedge k)/n(i)[n(i)-1]\,n(i')$$
$$+\sum_{1\leqq i\neq i'\leqq I}\ \sum_{1\leqq j\leqq J}\ \sum_{1\leqq k\leqq K}$$
$$n(i\wedge j)\,n(i'\wedge j)[n(i'\wedge j)-1]\,n(i\wedge k)\,n(i'\wedge k)$$
$$\times[n(i'\wedge k)-1]/n(i)\,n(i')[n(i')-1]$$
$$+\sum_{G(I)}\{\sum_{1\leqq j\leqq J}\ \sum_{1\leqq k\leqq K}$$
$$n(i\wedge j)\,n(i'\wedge j)\,n(i''\wedge j)\,n(i\wedge k)\,n(i'\wedge k)$$
$$\times n(i''\wedge k)/n(i)\,n(i')\,n(i'')\}. \quad (42)$$

où, rappelons le [*cf.* formule (18)], $G(I)$ est l'ensemble des couples de paires de la forme $(\{i,i'\},\{i,i''\})$ où des symboles distincts représentent des indices différents.

Nous noterons $\gamma(G)$ l'expression (42) précédente.

3.*h. calcul de* $\mathscr{E}(\sum\{\varphi(p)\varphi(q)\psi'(p)\psi'(q)/(p,q)\in H\})$

Ici encore, on aura à décomposer, conformément à (18), la somme sous le signe espérance en huit morceaux respectivement correspondants à A_1, A_2, B_1, B_2, C_1, C_2, D_1 et D_2. Compte-tenu des calculs préliminaires précédents [formules (26) à (41)] on obtient :

$$+\sum_{1\leqq i\neq i'\leqq I}\{\sum_{1\leqq j\leqq J} n(i\wedge j)\,n(i'\wedge j)[n(i'\wedge j)-1][n(i'\wedge j)-2]$$

$$\times\Big[\sum_{1\leqq k\leqq K} n(i\wedge k)[n(i\wedge k)-1][n(i\wedge k)-2][n(i\wedge k)-3]\Big]/$$

$$4\,n(i)[n(i)-1][n(i)-2][n(i)-3]$$

$$+\sum_{1\leqq i\neq i'I}\Big[\sum_{1\leqq j,j'\leqq J} n(i\wedge j)[n(i\wedge j)-1]$$

$$\times n(i'\wedge j')[n(i'\wedge j')-1]\Big]$$

$$\times\Big[\sum_{1\leqq k,k'\leqq K} n(i\wedge k)[n(i\wedge k)-1]\,n(i'\wedge k')[n(i'\wedge k')-1]\Big]/$$

$$4\,n(i)[n(i)-1]\,n(i')[n(i')-1]$$

$$+\sum_{1\leqq i\neq i'I}\Big\{\sum_{1\leqq j\leqq J} n(i\wedge j)[n(i\wedge j)-1][n(i\wedge j)-2]$$

$$+\sum_{1\leqq j\neq j'\leqq J} n(i\wedge j)[n(i\wedge j)-1]\,n(i\wedge j')\,n(i'\wedge j')\Big\}$$

$$\times\sum_{1\leqq k\leqq K} n(i\wedge k)[n(i\wedge k)-1][n(i\wedge k)-2]$$

$$+\sum_{1\leqq k\neq k'\leqq K} n(i\wedge k)[n(i\wedge k)-1]\,n(i\wedge k')\,n(i'\wedge k')\Big\}/$$

$$2\,n(i)[n(i)-1][n(i)-2]\,n(i')$$

$$+\sum\Big\{\Big[\sum_{1\leqq j,j'\leqq J} n(i\wedge j)[n(i\wedge j)-1]\,n(i'\wedge j')\,n(i''\wedge j')$$

$$\times\sum_{1\leqq k,k'\leqq K} n(i\wedge k)[n(i\wedge k)-1]\,n(i'\wedge k')\,n(i''\wedge k')\Big]/$$

$$2\,n(i)[n(i)-1]\,n(i')\,n(i'')\,\Big|\,1\leqq i\leqq I,\ 1\leqq i'<i''\leqq I,\ i'\neq i,\ i''\neq i\Big\}$$

$$+\sum_{1\leqq i\neq i'\leqq I}\Big\{\sum_{1\leqq j\leqq J} n(i\wedge j)\,n(i'\wedge j)[n(i'\wedge j)-1][n(i'\wedge j)-2]$$

$$+\sum_{1\leqq j\neq j'\leqq J} n(i\wedge j)\,n(i'\wedge j)\,n(i'\wedge j')[n(i'\wedge j')-1]\Big\}$$

$$\times\sum_{1\leqq k\leqq K} n(i\wedge k)\,n(i'\wedge k)[n(i'\wedge k)-1][n(i'\wedge k)-2]$$

$$+\sum_{1\leqq k\neq k'\leqq K} n(i\wedge k)\,n(i'\wedge k)\,n(i'\wedge k')[n(i'\wedge k')-1]\Big\}/$$

$$2\,n(i)\,n(i')\,[n(i')-1]\,[n(i')-2]$$

$$+\sum\{\{\sum_{1\leq j,j'\leq J} n(i'\wedge j)\,n(i''\wedge j)\,n(i\wedge j')\,[n(i\wedge j')-1]\}$$

$$\times\{\sum_{1\leq k,k'\leq K} n(i'\wedge k)\,n(i''\wedge k)\,n(i\wedge k')\,[n(i\wedge k')-1]\}/$$

$$2\,n(i')\,n(i'')\,n(i)\,[n(i)-1]\,|\,1\leq i'<i''\leq I,$$

$$1\leq i\leq I,\ i\neq i',\ i\neq i''\}$$

$$+\sum_{G(I)}\{\sum_{1\leq j\leq J} n(i\wedge j)\,n(i'\wedge j)\,[n(i\wedge j)-1]\,n(i''\wedge j)$$

$$+\sum_{1\leq j\neq j'\leq J} n(i\wedge j)\,n(i'\wedge j)\,n(i\wedge j')\,n(i''\wedge j')\}$$

$$\times\sum_{1\leq k\leq K} n(i\wedge k)\,n(i'\wedge k)\,[n(i\wedge k)-1]\,n(i''\wedge k)$$

$$+\sum_{1\leq k\neq k'\leq K} n(i\wedge k)\,n(i'\wedge k)\,n(i\wedge k')\,n(i''\wedge k')\}/$$

$$n(i)\,[n(i)-1]\,n(i')\,n(i'')$$

$$+\sum_{H(I)}\{\sum_{1\leq j,j'\leq J} n(i\wedge j)\,n(i'\wedge j)\,n(i''\wedge j')\,n(i'''\wedge j')\}$$

$$\times\{\sum_{1\leq k,k'\leq K} n(i\wedge k)\,n(i'\wedge k)\,n(i''\wedge k')\,n(i'''\wedge k')\}/$$

$$n(i)\,n(i')\,n(i'')\,n(i'''). \qquad (43)$$

où, rappelons-le encore une fois $G(I)$ [resp. $H(I)$] est l'ensemble des couples de paires d'indices avec (resp. sans) composante commune.

Nous noterons $\eta(H)$ l'expression (43) précédente.

THÉORÈME 1 : *La moyenne* $\mu(\beta,\gamma';\alpha)$ *de la v. a.* $S(\beta,\gamma';\alpha)$ *est donnée par l'expression* (14) *ci-dessus. La variance* $\mathrm{var}(\beta,\gamma';\alpha)$ *de cette v. a. est donnée par la formule :*

$$\mathrm{var}(\beta,\gamma';\alpha)=\mu(\beta,\gamma';\alpha)+\gamma(G)+\eta(H)-[\mu(\beta,\gamma';\alpha)]^2, \qquad (44)$$

où $\gamma(G)$ [resp. $\eta(H)$] *est donné par la formule* (42) [resp. (43)] *ci-dessus. L'indice d'association partielle* $\rho_{\beta\gamma.\alpha}$ *est dans ces conditions donné par la formule*

$$\rho_{\beta\gamma.\alpha}=[s(\beta,\gamma)-\mu(\beta,\gamma';\alpha)]/\sqrt{\mathrm{var}(\beta,\gamma';\alpha)}, \qquad (45)$$

où $s(\beta,\gamma)$ *est l'indice brut défini par la formule* (1) *ci-dessus.*

Les expressions rentrant dans la composition de la formule (44) peuvent paraître complexes d'un point de vue analytique, toutefois leur programmation sur ordinateur, bien que délicate, ne présente pas de difficulté majeure.

On a pu déjà remarquer la symétrie en j et k, $1 \leqq j \leqq J$, $1 \leqq k \leqq K$, des expressions de $\mu(\beta, \gamma'; \alpha)$ et de $\operatorname{var}(\beta, \gamma'; \alpha)$. Cette symétrie demeure pour n'importe quel moment de la distribution de $S(\beta, \gamma'; \alpha)$ car comme dans le cas total et comme nous allons chercher à nous en persuader, la distribution de la v. a. $S(\beta, \gamma'; \alpha)$ est la même que celle de la v. a. duale $S(\beta', \gamma; \alpha)$. La partition aléatoire Q' associée à Q peut être notée $\{F'_j/1 \leqq j \leqq J\}$ où F'_j est la partie aléatoire de E associée à la classe F_j de la partition Q, $1 \leqq j \leqq J$, dans une h. a. l. qui préserve la position relative par rapport aux classes E_i de la partition P, $1 \leqq i \leqq I$. Par conséquent, F'_j sera défini sous la forme de la somme ensembliste :

$$F'_j = \sum_{1 \leqq i \leqq I} F'_{ji} \tag{46}$$

pour tout $j = 1, 2, \ldots, J$, où $(F'_{j1}, F'_{j2}, \ldots, F'_{ji}, \ldots, F'_{jI})$ est une suite de parties aléatoires indépendantes, F'_{ji} étant choisi de cardinal $n(i \wedge j)$, uniformément au hasard dans E_i; en d'autres termes F'_{ji} est un élément aléatoire dans l'ensemble, muni d'une probabilité uniforme, des parties de E_i de même cardinal $n(i \wedge j)$. En se reférant à des notations que l'on comprend, si la v. a. $S(\beta, \gamma'; \alpha)$ pouvait se mettre sous la forme [*cf.* (4) et (5)] :

$$\begin{aligned} S(\beta, \gamma'; \alpha) = & \sum_{1 \leqq i \leqq I} \sum \{\varphi(p)\, \psi'(p)/p \in P_2(E_i)\} \\ & + \sum_{1 \leqq i < i' \leqq I} \sum \{\varphi(p)\, \psi'(q)/(p, q) \in E_i * E_{i'}\}, \end{aligned} \tag{47}$$

celle $S(\beta', \gamma; \alpha)$ peut se mettre aussi sous la forme :

$$\begin{aligned} S(\beta', \gamma; \alpha) = & \sum_{1 \leqq i \leqq I} \sum \{\varphi'(p)\, \psi(p)/p \in P_2(E_i)\} \\ & + \sum_{1 \leqq i < i' \leqq I} \sum \{\varphi'(p)\, \psi(q)/(p, q) \in E_i * E_{i'}\}. \end{aligned} \tag{48}$$

En vertu des propriétés de dualité que nous avions établies lors de l'étude du cas total (*cf.* [Lerman (1973)]), chacune des v.a. de la forme $\sum \{\varphi(p)\, \psi'(p)/p \in P_2(E_i)\}$ a la même distribution que celle $\sum \{\varphi' p)\, \psi(p)/p \in P_2(E_i)\}$, $1 \leqq i \leqq I$. De même, on peut montrer que pour tout entier r, le moment d'ordre r de la distribution de $\sum \{\varphi(p)\, \psi'(q)/(p, q) \in E_i * E_{i'}\}$ est identique au moment d'ordre r de la distribution de $\sum \{ \varphi'(p) \psi(q)/(p, q) \in E_i * E_{i'} \}$, $1 \leqq i < i' \leqq I$. Enfin le système des relations stochastiques entre ces différentes v. a. qui rentrent dans la composition de la somme (47) est, pour des raisons d'identification formelle, exactement le même que celui qui régit les différentes v. a. associées dualement de la somme (48). Il en résulte la propriété suivante :

THÉORÈME 2 : *Dans le cadre des h. a. l. à caractère partiel* [cf. *formule* (2) *pour la v. a.* $S(\beta, \gamma'; \alpha)$ *et formule* (46) *pour la v. a.* $S(\beta', \gamma; \alpha)$], *la distribution de la v. a.* $S(\beta, \gamma'; \alpha)$ *est la même que celle de la v. a.* $S(\beta', \gamma; \alpha)$.

4. CAS DE L'INDÉPENDANCE CONDITIONNELLE

La notion d'indépendance sous jacente à notre indice de comparaison entre partitions s'exprime au niveau de l'ensemble des paires d'objets de l'échantillon étudié. Pour le voir, revenons au cas total de la comparaison entre les deux variables partitions β et γ (*cf.* [Lerman (1973)]). Dans ce cas, le numérateur de l'indice de comparaison se met sous la forme :

$$\sum \left\{ \frac{n(j \wedge k)[n(j \wedge k) - 1]}{2} - \frac{n(j)[n(j) - 1]\, n(k)[n(k) - 1]}{2\, n(n-1)} \Big/ 1 \leqq j \leqq J,\ 1 \leqq k \leqq K \right\}. \quad (49)$$

La relation définissant l'indépendance entre les variables partitions β et γ s'exprime ici comme suit. Pour tout (j, k), $1 \leqq j \leqq J$, $1 \leqq k \leqq K$, la proportion de paires réunies dans la classe (j, k) de la partition croisée $Q \wedge R$ est égale au produit de la proportion de paires réunies dans la classe j par la proportion de paires réunies dans la classe k; soit :

$$\frac{n(j \wedge k)[n(j \wedge k) - 1]}{n(n-1)} = \frac{n(j)[n(j) - 1]}{n(n-1)} \times \frac{n(k)[n(k) - 1]}{n(n-1)}$$

$$\text{pour tout } (j, k), \quad 1 \leqq j \leqq J, \quad 1 \leqq k \leqq K. \quad (50)$$

Il est clair que les relations (50) entraînent la nullité de l'expression (49).

Dans ces conditions, l'indépendance conditionnelle entre les variables partitions β et γ relativement à la variable partition α, s'exprime dans ce cadre comme suit :

— Pour tout i, $1 \leqq i \leqq I$ et tout (j, k), $1 \leqq j \leqq J$, $1 \leqq k \leqq K$, la proportion de paires réunies dans la classe (i, j, k) de $P \wedge Q \wedge R$, est égale au produit de la proportion des paires réunies dans la classe (i, j) de $P \wedge Q$ par la proportion de paires réunies dans la classe (i, k) de $P \wedge R$; ces différentes proportions étant conditionnelles et relatives à l'ensemble des paires de la classe E_i :

$$(\forall i, 1 \leqq i \leqq I) \quad \text{et} \quad (\forall (j, k), 1 \leqq j \leqq J, 1 \leqq k \leqq K),$$

$$\frac{n(i \wedge j \wedge k)[n(i \wedge j \wedge k)-1]}{n(i)[n(i)-1]} = \frac{n(i \wedge j)[n(i \wedge j)-1]}{n(i)[n(i)-1]} \times \frac{n(i \wedge k)[n(i \wedge k)-1]}{n(i)[n(i)-1]}. \quad (51)$$

– Pour tout (i, i'), $1 \leqq i < i' \leqq I$ et tout (j, k), $1 \leqq j \leqq J$, $1 \leqq k \leqq K$, la proportion relative de paires de la classe (j, k) de $Q \wedge R$ dont les deux composantes appartiennent respectivement aux classes (i, j, k) et (i', j, k) de $P \wedge Q \wedge R$ est égale au produit de la proportion relative des paires de la classe j dont les deux composantes appartiennent respectivement aux classes (i, j) et (i', j) de $P \wedge Q$, par la proportion relative des paires de la classe k dont les deux composantes appartiennent respectivement aux classes (i, k) et (i', k) de $P \wedge R$; ces différentes proportions étant conditionnelles et relatives à l'ensemble des paires dont l'une des composantes appartient à E_i et l'autre composante à $E_{i'}$:

$$(\forall (i, i'), 1 \leqq i < i' \leqq I) \quad \text{et} \quad (\forall (j, k), 1 \leqq j \leqq J, 1 \leqq k \leqq K),$$

$$\frac{n(i \wedge j \wedge k)\, n(i' \wedge j \wedge k)}{n(i)\, n(i')} = \frac{n(i \wedge j)\, n(i' \wedge j)}{n(i)\, n(i')} \times \frac{n(i \wedge k)\, n(i' \wedge k)}{n(i)\, n(i')}. \quad (52)$$

Rappelons l'expression explicite de l'indice centré, numérateur du coefficient $\rho_{\beta\gamma.\alpha}$:

$$\begin{aligned}\sum \{ n(j \wedge k)[n(j \wedge k)-1]/2 - (& \sum_{1 \leqq i \leqq I} n(i \wedge j)[n(i \wedge j)-1] \\ & \times n(i \wedge k)[n(i \wedge k)-1]/2\, n(i)[n(i)-1] \\ & + \sum_{1 \leqq i < i' \leqq I} n(i \wedge j)\, n(i' \wedge j)\, n(i \wedge k)\, n(i' \wedge k)/n(i)\, n(i')) \\ & \qquad \mid 1 \leqq j \leqq J, 1 \leqq k \leqq K \}. \quad (53)\end{aligned}$$

Dans le cas de l'hypothèse de l'indépendance conditionnelle caractérisée par les relations (51) et (52) ci-dessus, on peut remplacer le contenu de la parenthèse de l'expression (53) (qui représente $\mu(\beta, \gamma'; \alpha)$ par :

$$\sum_{1 \leqq i \leqq I} n(i \wedge j \wedge k)[n(i \wedge j \wedge k)-1]/2 + \sum_{1 \leqq i < i' \leqq I} n(i \wedge j \wedge k)\, n(i' \wedge j \wedge k). \quad (54)$$

Ce cardinal est égal à $n(j \wedge k)[n(j \wedge k)-1]/2$; en effet la somme (54) représente la décomposition du nombre de paires réunies dans la classe $(j \wedge k)$ de $Q \wedge R$ conformément à la partition (5) (*voir* § 1 ci-dessus) de $P_2(E)$. Il en résulte la propriété suivante :

PROPRIÉTÉ. — L'indice de corrélation partielle $\rho_{\beta\gamma.\alpha}$ est nul dans le cas de l'indépendance conditionnelle caractérisée par les relations (51) et (52) ci-dessus.

IV.3. Cas d'un indice d'association partielle dans le cadre d'une h. a. l. globale

1. EXPRESSION DE L'H. A. L.

Les notations restent les mêmes que celles du paragraphe précédent, l'indice cherché ici sera désigné par $r_{\beta\gamma.\alpha}$.

L'indice brut entre les deux variables partition β et γ, $s(\beta, \gamma)$ reste toujours défini par la relation (1) du paragraphe précédent. Ce qui va changer est la nature des calculs conditionnels qui auront un caractère plus global. La v. a. $S_1(\beta, \gamma'; \alpha)$ [resp. $S_1(\beta', \gamma; \alpha)$] associée à $s(\beta, \gamma)$ sera en fait définie à partir de l'expression de ses moments calculés d'une certaine façon dans le cadre de la même h. a. l. que ci-dessus. De façon précise, les calculs conditionnels qui continueront à être effectués au niveau de E et non de $F=P_2(E)$, le seront avec un conditionnement plus global, « en moyenne », en ne retenant d'une paire d'objets $\{x, y\}$ que son appartenance à $\mathcal{R}(P)$ où à $\mathcal{S}(P)$ (ensemble des paires séparées par P) et en ignorant la position relative des deux composantes de la paire par rapport aux diverses classes E_i, $1 \leqq i \leqq I$, de la partition P. De même, relativement à un couple de paires $(\{x, y\}, \{x, z\})$ de G [resp. $(\{x, y\}, \{z, t\})$ de H)], on ne retiendra de sa structure que son appartenance à $\mathcal{R}(P) \times \mathcal{R}(P)$, à $\mathcal{R}(P) \times \mathcal{S}(P)$, à $\mathcal{S}(P) \times \mathcal{R}(P)$ ou à $\mathcal{S}(P) \times \mathcal{S}(P)$, en ignorant la position relative par rapport aux diverses classes E_i, $1 \leqq i \leqq I$, des objets entrant dans la composition du couple de paires.

Cette approche se justifie à différents égards. Il s'agit d'abord de l'attitude statistique générale devant l'ignorance. D'autre part, dans le cas où les cardinaux des classes de la partition croisée $P \wedge R$ sont identiques, la moyenne de la v. a. $S_1(\beta, \gamma'; \alpha)$ est identique à celle de $S(\beta, \gamma'; \alpha)$; de sorte qu'on obtient le même numérateur du coefficient d'association partielle. Par ailleurs, toujours dans la même situation, les expressions élémentaires servant au calcul de la variance de $S(\beta, \gamma'; \alpha)$ se rapprochent formellement de celles correspondantes servant au calcul de la variance de $S_1(\beta, \gamma', \alpha)$. Enfin, si nous accentuons cette démarche jusqu'à ne distinguer, relativement à un couple de paires (p, q), que $p=q$ et $p \neq q$, on se retrouve dans la même situation que celle de la comparaison partielle d'attributs (*cf.* § II) où on aurait ici à comparer les deux parties $\mathscr{R}(Q)$ et $\mathscr{R}(R)$ de F, en cherchant à neutraliser l'influence de la partie $\mathscr{R}(P)$. L'analyse du cas ordinal [Lerman (1983)] montre que c'est exactement ce type d'approche qui correspond à la définition du coefficient d'association partielle entre variables « rang » de M. G. Kendall.

Ainsi en tenant compte de la structure propre du couple de paires d'objets, nous fournissons un calcul de la variance plus précis que le dernier mentionné.

2. Calcul de la moyenne de la v.a. $S_1(\beta, \gamma'; \alpha)$

Nous continuons à noter φ (resp. ψ') la fonction indicatrice de $\mathscr{R}(Q)$ [resp. $\mathscr{R}(R')$]. La v. a. $S_1(\beta, \gamma'; \alpha)$ peut s'exprimer par la formule suivante :

$$S_1(\beta, \gamma'; \alpha) = \sum \{ \varphi(p)\,\psi'(p) / p \in \mathscr{R}(P) \} + \sum \{ \varphi(p)\,\psi'(p) / p \in \mathscr{S}(P) \}. \tag{1}$$

Dans ces conditions, le calcul de $\mathscr{E}[S_1(\beta, \gamma'; \alpha)]$ suppose le calcul des deux expressions suivantes :

$$\Pr\{ \psi'(p) = 1 / p \in \mathscr{R}(P) \} \tag{2}$$

et :

$$\Pr\{ \psi'(p) = 1 / p \in \mathscr{S}(P) \}. \tag{3}$$

Nous désignerons par $r[(n; t); P]$ le cardinal de l'ensemble des partitions R_1 en classes étiquetées de type t pour lesquelles le type de la partition croisée $P \wedge R_1$ est celui de $P \wedge R$; soit $\{ n(i \wedge k) / 1 \leqq i \leqq I,\ 1 \leqq k \leqq K \}$. Nous n'aurons pas besoin d'expliciter ce cardinal dans les raisonnements par symétrie que nous aurons à conduire ci-dessous.

Nous désignerons par $r[p/p \in \mathscr{R}(P)]$ (resp. $r[p/p \in \mathscr{S}(P)]$) le nombre *moyen* de fois où les composantes de la paire p se trouvent réunies dans une même classe de la partition R_1, lorsque R_1 décrit le sous-ensemble de $\mathscr{P}(n; t)$ caractérisé par le type de $P \wedge R_1$.

Dans ces conditions, les expressions (2) et (3) ci-dessus peuvent se mettre sous la forme :

$$\Pr\{ \psi'(p) = 1 / p \in \mathscr{R}(P) \} = \frac{r[p/p \in \mathscr{R}(P)]}{r[(n; t); P]}, \tag{4}$$

$$\Pr\{ \psi'(p) = 1 / p \in \mathscr{S}(P) \} = \frac{r[p/p \in \mathscr{S}(P)]}{r[(n; t); P]}. \tag{5}$$

Pour avoir le premier des deux rapports, nous allons compter de deux façons différentes le nombre de réunions de paires de $\mathscr{R}(P)$ qui ont lieu par la partition R_1, lorsque R_1 décrit son ensemble d'évolution dont le cardinal vient d'être noté $r[(n; t); P]$. Ce nombre de réunions s'exprime d'une part sous la forme :

$$r[p/p \in \mathscr{R}(P)] \times \mathrm{card}[\mathscr{R}(P)]$$

et d'autre part, sous la forme :

$$r[(n; t); P] \times \mathrm{card}[\mathscr{R}(P) \cap \mathscr{R}(R)].$$

Il en résulte que :

$$\Pr\{\psi'(p) = 1/p \in \mathscr{R}(P)\} = \frac{\mathrm{card}[\mathscr{R}(P) \cap \mathscr{R}(R)]}{\mathrm{card}[\mathscr{R}(P)]}. \tag{6}$$

C'est de la même manière qu'on évalue le deuxième des deux rapports (5) ci-dessus. On compte de deux façons différentes le nombre de réunions de paires de $\mathscr{S}(P)$ qui s'opèrent par la partition R_1, lorsque R_1 décrit son espace d'évolution. Ce nombre de réunions de paires s'exprime d'une part sous la forme :

$$r[p/p \in \mathscr{S}(P)] \times \mathrm{card}[\mathscr{S}(P)]$$

et d'autre part, sous la forme :

$$r[(n; t); P] \times \mathrm{card}[\mathscr{S}(P) \cap \mathscr{R}(R)].$$

Il en résulte que :

$$\Pr\{\psi'(p) = 1/p \in \mathscr{S}(P)\} = \frac{\mathrm{card}[\mathscr{S}(P) \cap \mathscr{R}(R)]}{\mathrm{card}[\mathscr{S}(P)]}. \tag{7}$$

Dans ces conditions, on obtient :

$$\mathscr{E}[S_1(\beta, \gamma'; \alpha) = \frac{\mathrm{card}[\mathscr{R}(P) \cap \mathscr{R}(Q)]\,\mathrm{card}[\mathscr{R}(P) \cap \mathscr{R}(R)]}{\mathrm{card}[\mathscr{R}(P)]} + \frac{\mathrm{card}[\mathscr{S}(P) \cap \mathscr{R}(Q)]\,\mathrm{card}[\mathscr{S}(P) \cap \mathscr{R}(R)]}{\mathrm{card}[\mathscr{S}(P)]}$$

puisque φ est la fonction indicatrice de $\mathscr{R}(Q)$.

En explicitant par rapport aux cardinaux des classes, l'expression (8) peut se mettre sous la forme :

$$\sum \{ n(i \wedge j)[n(i \wedge j)-1]\, n(i' \wedge k)[n(i' \wedge k)-1] / 1 \leqq i < i' \leqq I,\ 1 \leqq j \leqq J,\ 1 \leqq k \leqq K \} /$$

$$2 \sum \{ n(i)[n(i)-1] / 1 \leqq i \leqq I \} + \sum \{ n(i \wedge j)\, n(i' \wedge j)\, n(i'' \wedge k)\, n(i''' \wedge k) / 1 \leqq i < i' \leqq I,\ 1 \leqq i' < i'' \leqq I,\ 1 \leqq j \leqq J,\ 1 \leqq k \leqq K \} / \sum \{ n(i)\, n(i') / 1 \leqq i < i' \leqq I \}. \tag{9}$$

Il est naturel de chercher à comparer $\mathscr{E}[S_1(\beta, \gamma'; \alpha)]$ avec $\mathscr{E}[S(\beta, \gamma'; \alpha)]$ obtenu au paragraphe précédent dont nous allons reprendre la formule (14) qui peut se mettre sous la forme :

$$\mathscr{E}[S(\beta, \gamma'; \alpha)] = \sum_{1 \leq i \leq I} \frac{card[P_2(E_i) \cap \mathscr{R}(Q)]\, \mathrm{card}[P_2(E_i) \cap \mathscr{R}(R)]}{\mathrm{card}[P_2(E_i)]} + \sum_{1 \leq i < i' \leq I} \frac{\mathrm{card}[(E_i * E_{i'}) \cap \mathscr{R}(Q)]\, \mathrm{card}[(E_i * E_{i'}) \cap \mathscr{R}(R)]}{\mathrm{card}(E_i * E_{i'})} \tag{10}$$

Par conséquent, $\mathscr{E}[S(\beta, \gamma'; \alpha)]$ est essentiellement distinct de $\mathscr{E}[S_1(\beta, \gamma'; \alpha)]$ et on voit bien que sa structure correspond à une h. a. l. plus locale que celle qui définit $\mathscr{E}[S_1(\beta, \gamma'; \alpha)]$.

3. Définition de l'indépendance conditionnelle

La définition de la notion d'indépendance conditionnelle liée à l'expression de l'h. a. l. considérée ici (*cf.* § 1 ci-dessus) devient :

– la proportion relative de paires de la classe (j, k) de $Q \wedge R$ $(1 \leqq j \leqq J,\ 1 \leqq k \leqq K)$ dont les deux composantes se trouvent réunies dans une même classe non spécifiée de la partition P, est égale au produit de la proportion relative des paires de la classe j dont les deux composantes se trouvent réunies dans une même classe non spécifiée de la partition P, par la proportion relative des paires de la classe k dont les deux composantes se trouvent réunies dans une même classe non spécifiée de la partition P; ces proportions étant relatives à l'ensemble des paires réunies par la partition P.

– La proportion relative des paires de la classe (j, k) de $Q \wedge R$ $(1 \leqq j \leqq J,\ 1 \leqq k \leqq K)$ dont les deux composantes se trouvent séparées dans deux classes distinctes mais non spécifiées de la partition P est égale au produit de la proportion relative des paires de la classe j dont les deux composantes se trouvent séparées dans deux classes distinctes mais non spécifiées de la partition P, par la proportion relative des paires de la classe k dont les deux composantes se trouvent séparées dans deux classes distinctes mais non spécifiées de la partition P; ces proportions étant relatives à l'ensemble des paires séparées par la partition P.

La première assertion s'exprime par les relations :

$$(\forall (j, k),\ 1 \leqq j \leqq J,\ 1 \leqq k \leqq K),$$

$$\sum_{1 \leqq i \leqq I} n(i \wedge j \wedge k)[n(i \wedge j \wedge k)-1] / \sum_{1 \leqq i \leqq I} n(i)[n(i)-1]$$
$$= \{ \sum_{1 \leqq i \leqq I} n(i \wedge j)[n(i \wedge j)-1] / \sum_{1 \leqq i \leqq I} n(i)[n(i)-1] \}$$
$$\times \{ \sum_{1 \leqq i \leqq I} n(i \wedge k)[n(i \wedge k)-1] / \sum_{1 \leqq i \leqq I} n(i)[n(i)-1] \} \qquad (11)$$

La deuxième assertion s'exprime par les relations :

$$(\forall (j, k),\ 1 \leqq j \leqq J,\ 1 \leqq k \leqq K),$$

$$\sum_{1 \leqq i < i' \leqq I} n(i \wedge j \wedge k)\, n(i' \wedge j \wedge k) / \sum_{1 \leqq i < i' \leqq I} n(i)\, n(i')$$
$$= \{ \sum_{1 \leqq i < i' \leqq I} n(i \wedge j)\, n(i' \wedge j) / \sum_{1 \leqq i < i' \leqq I} n(i)\, n(i') \}$$
$$\times \sum_{1 \leqq i < i' \leqq I} n(i \wedge k)\, n(i' \wedge k) / \sum_{1 \leqq i < i' \leqq I} n(i)\, n(i') \}. \qquad (12)$$

Propriété : L'indice de corrélation partielle $r_{\beta\gamma.\alpha}$ est nul dans le cas de l'indépendance conditionnelle caractérisée par les relations (11) et (12) ci-dessus.

En effet, il suffit de reprendre la décomposition déjà utilisée [*cf.* (54), § IV.2] pour le cardinal $n(j \wedge k)[n(j \wedge k)-1]/2$:

$$n(j \wedge k)[n(j \wedge k)-1]/2$$
$$= \sum_{1 \leqq i \leqq I} n(i \wedge j \wedge k)[n(i \wedge j \wedge k)-1]/2$$
$$+ \sum_{1 \leqq i < i' \leqq I} n(i \wedge j \wedge k)\, n(i' \wedge j \wedge k). \qquad (13)$$

Dans le cadre de l'hypothèse de l'indépendance conditionnelle, la relation (11) permet d'identifier la première somme de (13) avec le premier des deux rapports dont l'expression (9) fait la somme, alors que la relation (12) permet l'identification de la deuxième somme de (13) avec le deuxième des deux rapports de (9).

4. CALCUL DE LA VARIANCE DE LA V.A. $S_1(\beta, \gamma'; \alpha)$

Avec les notations déjà introduites ci-dessus, le carré de la v. a. $S_1(\beta, \gamma'; \alpha)$ peut se mettre sous la forme suivante :

$$\begin{aligned}
&\sum \{ \varphi(p)\, \psi'(p) / p \in \mathcal{R}(P) \} \\
&\quad + \sum \{ \varphi(p)\, \varphi(q)\, \psi'(p)\, \psi'(q) / (p, q) \in (\mathcal{R}(P))^{[2]} \} \\
&\quad + \sum \{ \varphi(p)\, \psi'(p) / p \in \mathcal{S}(P) \} \\
&\quad + \sum \{ \varphi(p)\, \varphi(q)\, \psi'(p)\, \psi'(q) / (p, q) \in (\mathcal{S}(P))^{[2]} \} \\
&\quad + 2 \sum \{ \varphi(p)\, \varphi(q)\, \psi'(p)\, \psi'(q) / (p, q) \in \mathcal{R}(P) \times \mathcal{S}(P) \}. \qquad (14)
\end{aligned}$$

où on a noté $X^{[2]}$ l'ensemble des couples formé d'éléments distincts de X.

Ce calcul nécessite la décomposition selon G (ensemble des couples de paires de E ayant une composante commune) et H (ensemble des couples de paires de E sans composante commune) de chacun des ensembles $(\mathcal{R}(P))^{[2]}$, $(\mathcal{S}(P))^{[2]}$ et $\mathcal{R}(P) \times \mathcal{S}(P)$. On a besoin d'évaluer chacune des expressions suivantes :

$$\left.\begin{aligned}
\gamma(g.r.2) &= \mathcal{E}[\psi'(p)\, \psi'(q) / (p, q) \in G \cap (\mathcal{R}(P))^{[2]}] \\
c(g.r.2) &= \sum \{ \varphi(p)\, \varphi(q) / (p, q) \in G \cap (\mathcal{R}(P))^{[2]} \} \\
\gamma(h.r.2) &= \mathcal{E}[\psi'(p)\, \psi'(q) / (p, q) \in H \cap (\mathcal{R}(P))^{[2]}] \\
c(h.r.2) &= \sum \{ \varphi(p)\, \varphi(q) / (p, q) \in H \cap (\mathcal{R}(P))^{[2]} \} \\
\gamma(g.s.2) &= \mathcal{E}[\psi'(p)\, \psi'(q) / (p, q) \in G \cap (\mathcal{S}(P))^{[2]}] \\
c(g.s.2) &= \sum \{ \varphi(p)\, \varphi(q) / (p, q) \in G \cap (\mathcal{S}(P))^{[2]} \} \\
\gamma(h.s.2) &= \mathcal{E}[\psi'(p)\, \psi'(q) / (p, q) \in H \cap (\mathcal{S}(P))^{[2]}] \\
c(h.s.2) &= \sum \{ \varphi(p)\, \varphi(q) / (p, q) \in H \cap (\mathcal{S}(P))^{[2]} \} \\
\gamma(g.r.s) &= \mathcal{E}[\psi'(p)\, \psi'(q) / (p, q) \in G \cap (\mathcal{R}(P) \times \mathcal{S}(P))] \\
c(g.r.s) &= \sum \{ \varphi(p)\, \varphi(q) / (p, q) \in G \cap (\mathcal{R}(P) \times \mathcal{S}(P)) \} \\
\gamma(h.r.s) &= \mathcal{E}[\psi'(p)\, \psi'(q) / (p, q) \in H \cap (\mathcal{R}(P) \times \mathcal{S}(P))] \\
c(h.r.s) &= \sum \{ \varphi(p)\, \varphi(q) / (p, q) \in H \cap (\mathcal{R}(P) \times \mathcal{S}(P)) \}.
\end{aligned}\right\} \qquad (15)$$

Tous ces calculs ont bien été effectués dans le rapport de recherche [Lerman (1981)] dont se trouve issu cet article. A la différence de la précédente forme de l'h. a. l. (*cf.* § IV.2) où nous avons tenu à présenter le détail de tous les calculs, nous nous contenterons ici, pour limiter le volume de ce texte, de reprendre à titre d'exemple, le calcul de $\gamma(g.s.2)$ et de $c(g.s.2)$, en laissant le soin au lecteur d'effectuer selon le même principe les autres calculs où d'ailleurs il apparaîtra que les quantités c sont définies par des expressions respectivement analogues à celles des numérateurs des proportions γ.

4.1. *Calcul de* $\gamma(g.s.2)=\mathscr{E}[\psi'(p)\,\psi'(q)/(p,q)\in G\cap(\mathscr{S}(P))^{[2]}]$

R_1 ayant le sens défini au paragraphe 2 ci-dessus, nous dirons qu'un couple de paires (p,q) est « spécifié » par la partition R_1 si cette dernière réunit les deux composantes de p (resp. de q); en d'autres termes, si (p,q) appartient à $(\mathscr{R}(R_1))^{[2]}$.

Nous allons calculer de deux façons différentes le nombre total de spécifications qui s'opèrent au niveau de l'ensemble $G\cap(\mathscr{S}(P))^{[2]}$ lorsque R_1 décrit son espace d'évolution dont, rappelons-le, le cardinal a été noté $r[(n;t);P]$. Pour cela, on introduit le nombre *moyen* de fois où (p,q) de $G\cap(\mathscr{S}(P))^{[2]}$ se trouve spécifié lorsque R_1 décrit son espace d'évolution : $r[(p,q)/(p,q)\in G\cap(\mathscr{S}(P))^{[2]}]$. On a :

$$\begin{aligned} r[(p,q)/(p,q)\in G\cap(\mathscr{S}(P))^{[2]}]\times\operatorname{card}[G\cap(\mathscr{S}(P))^{[2]}] \\ =r[(n;t);P]\times\operatorname{card}[G\cap(\mathscr{R}(R_1)^{[2]}\cap\mathscr{S}(P)^{[2]})]; \end{aligned} \tag{16}$$

en effet, chaque partition R_1 spécifie le même nombre $\operatorname{card}[G\cap(\mathscr{R}(R))^{[2]}\cap(\mathscr{S}(P))^{[2]}]$ de couples de paires de $G\cap(\mathscr{S}(P))^{[2]}$.

La quantité cherchée qui représente la proportion :

$$r[(p,q)/(p,q)\in G\cap(\mathscr{S}(P))^{[2]}]/r[(n;t);P], \tag{17}$$

sera calculée au moyen de la proportion suivante :

$$\operatorname{card}[G\cap(\mathscr{R}(R)^{[2]}\cap\mathscr{S}(P)^{[2]})]/\operatorname{card}[G\cap(\mathscr{S}(P))^{[2]}]. \tag{18}$$

La détermination du premier cardinal peut se voir à partir d'un raisonnement constructif sur le nombre de façons de choisir le couple de paires de la forme $(\{x,y\},\{x,z\})$ tel que $\{x,y\}$ et $\{x,z\}$ appartiennent à $\mathscr{S}(P)$; c'est-à-dire, se trouvent séparées par la partition $P=\{E_i/1\leqq i\leqq I\}$. On obtient :

$$\begin{aligned} \sum\{n(i)\,n(i')[n(i')-1]/1\leqq i\neq i'\leqq I\}+\sum\{n(i)\,n(i')\,n(i'')/ \\ (\{i,i'\},\{i,i''\})\in G(I)\}, \end{aligned} \tag{19}$$

où, rappelons-le, $G(I)$ est l'ensemble des couples de paires d'indices ayant une composante commune. D'autre part, l'ensemble sous le premier signe card de (18) se met sous la forme :

$$G(E) \cap ([\mathscr{R}(R) \cap \mathscr{S}(P)] \times [\mathscr{R}(R) \cap \mathscr{S}(P)]), \tag{20}$$

son énumération se fera conformément à la partition $\{G_k/1 \leqq k \leqq K\}$ définie par R.

Considérons dans ces conditions la partition $P_k = \{G_{ik}/1 \leqq i \leqq I\}$ définie par la restriction de la partition P à la classe G_k; il s'agit alors de calculer :

$$\operatorname{card}[G(E) \cap (\mathscr{S}(P_k))^{[2]}] \tag{21}$$

et on se retrouve devant le même type d'énumération que ci-dessus [*cf.* formule (19)]. Ainsi, le cardinal cherché se met sous la forme :

$$\begin{aligned}\sum_{1 \leqq k \leqq K} \{ & \textstyle\sum \{ n(i \wedge k)\, n(i' \wedge k) [n(i' \wedge k) - 1]/1 \leqq i \neq i' \leqq I \} \\ & + \textstyle\sum \{ n(i \wedge k)\, n(i' \wedge k)\, n(i'' \wedge k)/(\{i, i'\}, \{i, i''\}) \in G(I) \} \}\end{aligned} \tag{22}$$

où $\{n(i \wedge k)/1 \leqq i \leqq I\}$ définit le type de la partition P_k.

La proportion cherchée est ainsi définie au moyen du rapport (22)/(19).

4.2. *Calcul de* $c(g.s.2) = \sum \{ \varphi(p)\, \varphi(q)/(p, q) \in G \cap (\mathscr{S}(P))^{[2]} \}$

Cette expression représente exactement :

$$\operatorname{card}\{ G(E) \cap ([\mathscr{R}(Q) \cap \mathscr{S}(P)] \times [\mathscr{R}(Q) \cap \mathscr{S}(P)]) \} \tag{23}$$

qui, compte tenu de ci-dessus, s'exprime par la formule suivante :

$$\begin{aligned}\sum_{1 \leqq j \leqq J} \{ & \textstyle\sum \{ n(i \wedge j)\, n(i' \wedge j) [n(i' \wedge j) - 1]/1 \leqq i \neq i' \leqq I \} \\ & + \textstyle\sum \{ n(i \wedge j)\, n(i' \wedge j)\, n(i'' \wedge j)/(\{i, i'\}, \{i, i''\}) \in G(I) \} \}.\end{aligned} \tag{24}$$

THÉORÈME 1 : *La moyenne* $\mu_1(\beta, \gamma'; \alpha)$ *de la v. a.* $S_1(\beta, \gamma'; \alpha)$ *est donnée par l'expression* (9) *ci-dessus. La variance* $\operatorname{var.}_1(\beta, \gamma'; \alpha)$ *de cette v. a. est donnée par la formule :*

$$\begin{aligned} var._1(\beta, \gamma'; \alpha) \\ = \mu_1(\beta, \gamma'; \alpha) + \gamma(g.r.2)\, c(g.r.2) + \gamma(h.r.2)\, c(h.r.2) \\ + \gamma(g.s.2)\, c(g.s.2) + \gamma(h.s.2)\, c(h.s.2) \\ + 2[\gamma(g.r.s)\, c(g.r.s) + \gamma(h.r.s)\, c(h.r.s)] \\ - (\mu_1(\beta, \gamma'; \alpha))^2. \end{aligned} \tag{25}$$

L'indice d'association partielle $r_{\beta\gamma.\alpha}$ est donné par la formule :

$$r_{\beta\gamma.\alpha}=[s(\beta,\gamma)-\mu_1(\beta,\gamma';\alpha)]/\sqrt{\mathrm{var}_1(\beta,\gamma';\alpha)} \tag{26}$$

où $s(\beta,\gamma)$ est l'indice brut déjà défini par la formule (1) du paragraphe IV.2 ci-dessus.

Les expressions rentrant dans la composition de la formule (25) peuvent paraître complexes d'un point de vue analytique, toutefois, leur programmation sur ordinateur, bien que délicate au niveau de l'indexation à gérer, ne doit pas présenter de trop grande difficulté. Il sera alors intéressant de comparer les comportements respectifs des deux indices $\rho_{\beta\gamma.\alpha}$ (*cf.* § IV.2 précédent) et $r_{\beta\gamma.\alpha}$ correspondants aux deux formes de l'h. a. l. locale et globale.

On a pu remarquer la symétrie des expressions de la moyenne $\mu_1(\beta,\gamma';\alpha)$ et de la variance $\mathrm{var}_1(\beta,\gamma';\alpha)$ par rapport aux cardinaux des classes de $P \wedge Q$ d'une part et de $P \wedge R$ d'autre part. De sorte que le résultat aurait été le même si les calculs avaient été effectués par rapport à la v. a.; définie en fait à partir de ses moments :

$$S_1(\beta',\gamma;\alpha)=\sum\{\varphi'(p)\,\psi(p)/p\in\mathscr{R}(P)\}$$
$$+\sum\{\varphi'(p)\,\psi(p)/p\in\mathscr{S}(P)\}. \tag{27}$$

En décomposant l'ensemble $(P_2(E))^{[q]}$ des q-uples de paires dont deux quelconques sont distinctes, en sous-ensembles dont chacun est formé des q-uples de paires de même configuration (*cf.* [Lerman (1973)]), on peut montrer que le moment absolu d'ordre quelconque fixé, calculé « globalement » de la même manière que l'ont été la moyenne et le moment d'ordre 2, de la v. a. $S_1(\beta,\gamma';\alpha)$ est identique à celui de la v. a. $S_1(\beta',\gamma;\alpha)$. Dans ces conditions, on peut avancer la propriété suivante :

THÉORÈME 2 : *Dans le cadre de l'h. a. l. à caractère partiel et « global »* (cf. § 1 *ci-dessus), la distribution de la v. a.* $S_1(\beta,\gamma';\alpha)$ *est la même que celle de la v. a.* $S_1(\beta',\gamma;\alpha)$, *où la partition aléatoire* Q' *est définie de façon conforme à celle de* R'.

V. CONCLUSION

Au terme de cette longue étude, dont le volet ordinal se trouve dans [Lerman (1983)]), à la croisée de l'analyse non linéaire des données et de la statistique non paramétrique, il reste encore beaucoup à faire, surtout au niveau de l'opérationnalisation et de la validation.

Une recherche algorithmique particulière est en effet nécessaire pour l'informatisation du calcul des indices d'association partielle, dans les différents cas, surtout lorsqu'il s'agit d'établir toute la table des indices entre éléments d'un ensemble V de variables descriptives, partiels, relativement à une variable w exogène ($w \notin V$).

Signalons qu'une des motivations source de ce travail a été une étude sur la répartition des dépenses d'un ensemble de ménages sur différents postes, où on n'a pu dégager par la classification une organisation cohérente et nuancée de l'ensemble des postes de dépense qu'en neutralisant la variable « revenu du ménage ».

Plus généralement, on peut se poser la question de savoir ce que deviennent les différents comportements dégagés à partir de la classification de l'ensemble V des variables descriptives si on neutralisait l'influence d'une variable w fortement discriminante ($w \notin V$). Cette approche répond à la démarche du chercheur en sciences humaines qui désire éliminer dans son analyse l'influence de facteurs bien connus. Nous espérons que ce dernier trouvera ici un outil adapté. Toutefois, pour qu'il puisse en tirer profit, la collaboration doit être plus que jamais étroite avec le statisticien qui cherche à valider sa méthode et l'informaticien appliqué, son programme.

BIBLIOGRAPHIE

H. E. Daniels, *The Relation Between Measures of Correlation in the Universe of Sample Permutations*, Biometrika, vol. 33, 1974, p. 129-135.

J. J. Daudin, *Coefficient de Tschuprow partiel et indépendance conditionnelle*, Statistique et Analyse des Données, 1979, p. 55-58.

J. J. Daudin, *Partial Association Measures and an Application to Qualitative Regression*, Biometrika, vol. 67, 1980, p. 581-590.

M. G. Kendall, *Rank Correlation Methods* 4th edition, London: Griffin, 1970.

G. Lecalve, *Problèmes d'analyse des données*, Thèse d'état, 2e partie, Université de Rennes-I, 1976.

I. C. Lerman, *Les bases de la classification automatique*, Gauthier-Villars, Paris, 1970.

I. C. Lerman, *Étude distributionnelle de statistiques de proximité entre structures finies de même type; application à la classification automatique*, Cahiers du Bureau Universitaire de Recherche Opérationnelle, série recherche, n° 19, 1973.

I. C. Lerman, *Formal Analysis of a General Notion of Proximity Between Variables*, Actes du colloque « Congrès Européen des statisticiens », Grenoble, Septembre 1976, parus chez North Holland en 1977.

I. C. Lerman, R. Gras et H. Rostam, *Élaboration et évaluation d'un indice d'implication pour des données binaires I et II*, Revue Mathématiques et Sciences humaines, n° 74, p. 5-35 et n° 75, 1981, p. 5-47.

I. C. Lerman, *Corrélation partielle dans le cas « qualitatif »*, Rapport I.R.I.S.A., n° 153, Octobre 1981, Rennes, 125 p.

I. C. LERMAN, *Classification et analyse ordinale des données*, Dunod, Paris, 1981, 760 p.

I. G. LERMAN, *Indices d'associations partielle entre variables « qualitatives ordinales »*, Publications de l'Institut de Statistique des Universités de Paris XXVIII, fasc. 1 et 2, 1983, p. 7-46.

P. A. P. MORAN, *Partial and Multiple Rank Correlation*, Biometrika, vol. 38, 1951, p. 26-32.

G. SAPORTA, *Quelques applications des opérateurs d'Escoufier au traitement des variables qualitatives*, Statistique et Analyse des Données, vol. 1, 1976, p. 38-46.

H. R. SOMMERS, *Analysis of Partial Rank Correlation Measures Based on the Product-Moment Model: part one*, Social forces, vol. 53, n° 2, décembre 1974.

G. E. NOETHER, *On a Theorem by Wald and Wolfowitz*, Annals of Mathematical Statistics, vol. 20, 1949, p. 455-458.

A. WALD et J. WOLFOWITZ, *Statistical Tests Based on Permutations of the Observations*, Annals of Mathematical Statistics, vol. 15, 1944, p. 358-372.