

LE "LIEN" COMME MESURE DE L'INFORMATION DANS UN TABLEAU DE CONTINGENCE.

Georges Laplace.

La démarche scientifique confronte sans relâche ce qui pourrait être et ce qui est. C'est le moyen de construire une représentation du monde toujours plus proche de ce que nous appelons "la réalité".

François Jacob.

En typologie analytique le tableau de contingence et ses dérivés directs, le tableau des fréquences et les tableaux de fréquences conditionnelles, constituent le support de l'analyse structurale des données de l'analyse qualitative et quantitative des caractères. Il importe donc, pour une meilleure intelligence de la "métrique du Khi 2", fondement de la démarche de l'analyse structurale, de se familiariser avec un système de notations précis, cohérent et prégnant.

1. LE TABLEAU DE CONTINGENCE ET LES TABLEAUX DE FREQUENCES.

1.1 Le tableau de contingence.

On appelle tableau de contingence un tableau qui donne la ventilation des données d'une série ou ensemble statistique selon le croisement de deux caractères qualitatifs: le caractère I à r modalités et le caractère J à k modalités.

Dans le tableau qui suit:

a - n_{ij} exprime le nombre d'observations présentant à la fois la modalité i du caractère I et la modalité j du caractère J;

b - n_i représente l'effectif marginal de la modalité i du caractère I, c'est à dire la somme des n_{ij} le long de i, j variant de 1 à k, soit:

$$n_i = \sum_{j=1}^k n_{ij}$$

c - n_j représente l'effectif marginal de la modalité j du caractère J , c'est à dire la somme des n_{ij} le long de j , i variant de 1 à r , soit:

$$n_j = \sum_{i=1}^r n_{ij}$$

d - N représente l'effectif total de la série statistique, soit:

$$N = \sum_{j=1}^k \sum_{i=1}^r n_{ij}$$

		Caractère J (k modalités)				
		j	j'			
Caractère I (r modalités)		i	n_{ij}	$n_{ij'}$	n_i	
		i'	$n_{i'j}$	$n_{i'j'}$	$n_{i'}$	
				n_j	$n_{j'}$	N

Illustrons immédiatement le tableau de contingence théorique par un exemple concret dans lequel le caractère "ensemble industriel" s'identifie au caractère I et le caractère "catégorie typologique" s'identifie au caractère J. Nous référant aux données concernant le gisement de la Grotte Gatzarria (LAPLACE 1966), nous dressons avec les ensembles industriels des couches Cjn2 et Cjn1 du Protoaurignacien et Cbf de l'Aurignacien, classés selon l'ordre diachronique, et avec les catégories typologiques ordinales[•], classées selon l'ordre dégressif de leurs effectifs globaux, le tableau de contingence (LAPLACE 1974):

	S	A	SE	E	B	P	
Cjn2	132	93	19	0	26	0	270
Cjn1	70	25	21	3	6	0	125
Cbf	216	42	76	52	8	0	394
	418	160	116	55	40	0	789

1.2 Le tableau des fréquences.

En effectuant le quotient de l'effectif n_{ij} de chaque case et

[•] C'est-à-dire avec S pour l'ordre des Simples, A pour l'ordre des Abrupts, SE pour l'ordre des Surélevés, E pour l'ordre des Ecaillés, B pour l'ordre des Burins et P pour l'ordre des Plans.

des effectifs marginaux n_i et n_j du tableau de contingence par l'effectif de la série statistique globale N , on obtient le tableau des fréquences:

		Caractère J (k modalités)		
		j	j'	
Caractère I (r modalités)	i	f_{ij}	$f_{ij'}$	f_i
	i'	$f_{i'j}$	$f_{i'j'}$	$f_{i'}$
		f_j	$f_{j'}$	1

dans lequel $f_{ij} = \frac{n_{ij}}{N}$ $f_i = \frac{n_i}{N}$ $f_j = \frac{n_j}{N}$ $\sum_j f_i = 1$ $\sum_i f_j = 1$

Pour les ensembles industriels aurignacoïdes et les catégories typologiques ordinales de Gatzarria on obtient le tableau des fréquences:

	S	A	SE	E	B	P	
Cjn2	.167	.118	.024033342
Cjn1	.089	.032	.027	.004	.008158
Cbf	.274	.053	.096	.066	.010373
	.530	.203	.147	.070	.051	1.000

1.3 Les tableaux des fréquences conditionnelles.

A partir du tableau de contingence, on peut construire deux tableaux de fréquences conditionnelles mettant en lumière soit les profils des r modalités i du caractère I, soit les profils des k modalités j du caractère J:

		Caractère J (k modalités)		
		j	j'	
Caractère I (r modalités)	i	f_{ij}^i	$f_{ij'}^i$	1
	i'	$f_{i'j}^i$	$f_{i'j'}^i$	1
		f_j	$f_{j'}$	1

dans lequel $f_j^i = \frac{n_{ij}}{n_i} = \frac{f_{ij}}{f_i}$ $\sum_j f_j^i = 1$

le vecteur f_J^i représentant le profil des k modalités j de J dans la modalité i de I;
 et, symétriquement:

		Caractère J (k modalités)			
		j	j'		
Caractère I (r modalités)	i	$f_{i,j}^j$	$f_{i,j'}^{j'}$	$f_{i,j}^j + f_{i,j'}^{j'}$	
	i'	$f_{i',j}^j$	$f_{i',j'}^{j'}$	$f_{i',j}^j + f_{i',j'}^{j'}$	
		1	1	1	

dans lequel $f_{i,j}^j = \frac{n_{ij}}{n_j} = \frac{f_{ij}}{f_j} \quad \sum_i f_{i,j}^j = 1$

le vecteur f_I^j représentant le profil des r modalités i de I dans la modalité j de J.

La notation $f_{i,j}^j$ est un peu difficile à saisir dans la mesure où l'on confond les rôles des j indices placés en haut et en bas de f. Pour y remédier on utilisera le procédé mnémotechnique suivant: quand on lit ou écrit $f_{i,j}^j$, se dire mentalement "f de j si i", ce qui signifie "fréquence conditionnelle de la modalité j de J si l'on se trouve dans la modalité i de I"; quand on écrit, écrire d'abord l'indice bas, puis l'indice haut.

Pour les ensembles aurignacoïdes et les catégories ordinales de Gatzarria on obtient le tableau des fréquences conditionnelles des différentes catégories dans chacun des ensembles:

	S	A	SE	E	B	P	
Cjn2	.489	.344	.070096	1
Cjn1	.560	.200	.168	.024	.048	1
Cbf	.548	.107	.193	.132	.020	1
	.530	.203	.147	.070	.051	1

et le tableau des fréquences conditionnelles des différents ensembles dans chacune des catégories:

	S	A	SE	E	B	P	
Cjn2	.316	.581	.164650342
Cjn1	.167	.156	.181	.055	.150158
Cbf	.517	.263	.655	.945	.200499
	1	1	1	1	1		1

2. LA METRIQUE DU KHI 2.

2.1 Le test du Khi 2.

On se propose de comparer un ensemble empirique ou observé 0,

distribué selon les modalités d'un caractère I, à une distribution théorique ou calculée C des mêmes modalités de ce caractère I.

Pour répondre à cette question, on calcule la somme des écarts quadratiques réduits, ou indicateur d'écart de Pearson, soit la quantité:

$$X^2 = \sum \frac{(\text{effectif observé} - \text{effectif théorique})^2}{\text{effectif théorique}} = \sum \frac{(O-C)^2}{C}$$

soit

$$X^2 = \sum_i \frac{(n_{ij} - \frac{n_i n_j}{N})^2}{\frac{n_i n_j}{N}}$$

et on la compare aux valeurs du Khi 2 à I-1 degrés de liberté.

Si cette quantité présente une valeur supérieure à celle du seuil de signification P = .050 du Khi 2, on rejette l'hypothèse nulle H₀ d'homogénéité ou d'indépendance des deux distributions au seuil de signification .050. Ce test est connu sous le nom de "test du Khi 2".

Notons que l'on peut encore calculer la quantité X² comme suit:

$$X^2 = \sum \frac{O^2}{C} - N \quad \text{soit} \quad X^2 = \sum \frac{n_{ij}^2}{n_i n_j} - N$$

2.2 La distance du Khi 2.

La construction de l'indicateur d'écart X² conduit à définir une distance entre distributions à l'aide d'une métrique analogue à celle que l'on a utilisée pour le test du Khi 2. En effet, si l'on considère trois distributions p, q et r, le carré de la distance entre p et q, calculée avec la "métrique du Khi 2 centrée sur r", est donnée par l'expression:

$$\left\| p - q \right\|_r^2 = \sum_i \frac{(p_i - q_i)^2}{r_i}$$

Avec cette notation, l'indicateur d'écart X² de Pearson, f étant la distribution observée et q étant la distribution théorique ou calculée, s'écrit:

$$N \left\| f - p \right\|_p^2 = \sum_i \frac{(f_i - p_i)^2}{p_i}$$

En conséquence, on voit qu'une métrique du Khi 2 dépend de la distribution sur laquelle elle est centrée: à chaque point du "simplexe", c'est à dire de la structure organisée par l'inclusion, des distributions peut-être associée une métrique différente.

En conclusion, nous référant aux tableaux des fréquences conditionnelles, on utilisera, pour calculer la distance entre les distributions Xⁱ et X^{i'}, la métrique du Khi 2 centrée sur f_J:

$$d^2(i, i') = \left\| f_J^i - f_J^{i'} \right\|_{f_J}^2$$

$$\text{soit } d^2(i, i') = \sum \frac{1}{f_j} (f_j^i - f_j^{i'})^2 = \sum \frac{1}{f_j} \left(\frac{f_{ij}}{f_i} - \frac{f_{i'j}}{f_i} \right)^2$$

et, symétriquement, pour calculer la distance entre les distributions X^j et $X^{j'}$, la métrique du Khi 2 centrée sur f_I :

$$d^2(j, j') = \left\| \left\| f_I^j - f_I^{j'} \right\|_{f_I} \right\|^2$$

$$\text{soit } d^2(j, j') = \sum \frac{1}{f_i} (f_i^j - f_i^{j'})^2 = \sum \frac{1}{f_i} \left(\frac{f_{ij}}{f_i} - \frac{f_{ij'}}{f_i} \right)^2$$

2.3 L'inertie d'un nuage de points.

Nous rapportant à notre exemple concret, si l'on considère la fréquence conditionnelle f_j^i , à savoir la fréquence conditionnelle de la modalité j du caractère "catégorie typologique" dans la modalité i du caractère "ensemble industriel", le vecteur f_j^i représente le profil typologique, au niveau des ordres, de l'ensemble industriel i . On note X^i le point de l'espace à $k = 6$ dimensions de coordonnées $x_{ij} = f_j^i$, la somme des f_j^i étant égale à l'unité. On associe à X^i la masse f_i , qui représente l'importance relative de l'ensemble industriel i dans le tableau des fréquences, la somme f_i étant égale à l'unité.

Si l'on note $N(I)$ le "nuage" des points X^i munis des masses f_i , le centre de gravité G de $N(I)$ est donné par l'expression:

$$G = \sum_I f_i X^i$$

et sa coordonnée est
$$g_j = \sum_I f_i f_j^i = f_j$$

Le centre de gravité G de $N(I)$ est donc le point représentant le profil des "catégories typologiques", au niveau des ordres, dans la totalité des "ensembles industriels" considérés.

Parvenus à ce point, il faut souligner qu'un espace à $n > 3$ dimensions de l'espace usuel est inaccessible à notre imagination, le résultat des calculs n'étant qu'un ensemble de nombres.

L'inertie du nuage $N(I)$ formé par les points X^i munis des masses f_i est un paramètre de dispersion. On le calcule avec une expression, identique à celle de la variance, où G est le centre de gravité G du nuage $N(I)$:

$$\sum_I f_i \left\| X^i - G \right\|^2$$

Comme il est nécessaire de disposer d'une métrique sur $N(I)$, c'est-à-dire d'une métrique permettant de calculer les distances entre les points X^i , on choisit la métrique du Khi 2 centrée sur f_J . En conséquence, l'inertie du nuage $N(I)$ devient:

$$\sum_I f_i \left\| f_j^i - f_j \right\|_{f_J}^2$$

quantité qui s'écrit:
$$\sum_{ij} \frac{(f_{ij} - f_i f_j)^2}{f_i f_j}$$

expression remarquable parce que:

- a - elle fait jouer le même rôle aux lettres i et j;
- b - elle est nulle quand $f_{ij} = f_i f_j$, à savoir que lorsque le tableau est égal au produit des marges le nuage est concentré en un seul point;
- c - elle donne une mesure approchée de l'information apportée par un tableau de contingence, comme nous le verrons ultérieurement.

Dans cet espace à k dimensions R^k où se situe le nuage N(I) on dispose désormais d'un "triplet" constitué par:

- a - les points des coordonnées x_{ij} ;
- b - la répartition de masse f_i entre ces points, donnée par les f_i ;
- c - la distance du Khi 2 entre ces points.

Pour visualiser ce nuage, on le projette sur un espace à deux dimensions, c'est-à-dire un plan, de manière à obtenir le maximum d'allongement selon les deux axes d'inertie déterminés par l'analyse factorielle à partir du "triplet". C'est ainsi que cette projection optimale révèle les structures masquées. On procède de la même manière pour le nuage N(J) dans l'espace à r dimensions R^r . Or, on peut superposer les nuages N(I) et N(J) car il existe entr'eux des "formules de transition" qui font que l'on peut se servir de l'un pour interpréter l'autre. Telles sont les bases de l'analyse factorielle des correspondances.

3. LE "LIEN" ENTRE DEUX CARACTERES.

3.1 Définition du "Lien".

On appelle "Lien des caractères I et J mesuré sur une série ou ensemble statistique E" la quantité Lien (I,J) telle que:

$$\text{Lien (I,J)} = \sum_{ij} \frac{(f_{ij} - f_i f_j)^2}{f_i f_j}$$

On note immédiatement que Lien (I,J) est égal au carré de la distance entre les distributions f_{IJ} et $f_I f_J$ mesurée selon la métrique du Khi 2 centrée sur $f_I f_J$.

Une constatation tout empirique autorise à utiliser le Lien pour mesurer l'information apportée par un tableau de contingence (VOLLE 1981). Ainsi, le Lien peut être interprété comme une mesure du gain d'information apporté par la connaissance du contenu d'un tableau, c'est-à-dire la distribution f_{IJ} , à qui connaît au préalable les distributions marginales f_I et f_J . On notera que l'expression du Lien étant identique à celle de l'inertie des nuages N(I) et N(J) construits pour effectuer une analyse des correspondances, c'est par son intermédiaire que l'analyse des correspondances se raccorde à la théorie de l'information et que se justifie, dans cette analyse, l'usage de la métrique du Khi 2.

Le Lien présente des propriétés remarquables:

a - Lien (I,J) = Lien (J,L) en raison de la symétrie du lien en i et en j;

b - Lien (I,J) ≥ 0;

c - Lien (I,J) = 0 lorsque $f_{IJ} = f_i f_j$;

d - Lien (I,J) = $-\frac{X^2}{N}$ d'où $X^2 = N$ Lien (I,J)

expression qui suit, dans l'hypothèse nulle H_0 de l'homogénéité ou de l'indépendance de I et de J, une loi du Khi 2 à $(r-1)(k-1)$ degrés de liberté. Pour établir cette relation entre le Lien (I,J) et l'indicateur d'écart X^2 on part du tableau de contingence avec:

$$X^2 = \sum \frac{(O - C)^2}{C} \quad \text{soit} \quad X^2 = \sum_{ij} \frac{(n_{ij} - \frac{n_i n_j}{N})^2}{\frac{n_i n_j}{N}}$$

en multipliant le numérateur et le dénominateur par N, on obtient successivement:

$$X^2 = \sum_{ij} \frac{(Nn_{ij} - \frac{Nn_i n_j}{N})^2}{\frac{Nn_i n_j}{N^2}} = \sum_{ij} \frac{(Nf_{ij} - Nf_i f_j)^2}{Nf_i f_j} = \sum_{ij} \frac{N^2 (f_{ij} - f_i f_j)^2}{Nf_i f_j}$$

et, en définitive

$$X^2 = \sum_{ij} \frac{N(f_{ij} - f_i f_j)^2}{f_i f_j}$$

e - Lien (I,J) = $\sum_{ij} \frac{f_{ij}^2}{f_i f_j} - 1$

3.2 Du tableau de contingence au tableau du Lien.

Si l'on considère la distribution croisée de deux caractères I et J sur une série statistique E, on obtient l'expression du Lien (I,J) qui peut s'interpréter comme le gain d'information réalisé lorsque l'on passe de la connaissance des distributions marginales f_I et f_J à la connaissance du contenu du tableau f_{IJ} . Ainsi, le Lien nous fournit-il l'expression d'une "distance", calculée selon la métrique du Khi 2 centrée sur la distribution $f_i f_j$, entre le tableau f_{IJ} et le tableau "produit des marges" $f_i f_j$. Cette "distance" nous permet d'apprécier ce que la connaissance de f_{IJ} nous apporte:

a - si le tableau f_{IJ} s'identifie au "produit des marges" $f_i f_j$, le Lien est nul et la connaissance du tableau f_{IJ} ne présente aucun intérêt;

b - plus le tableau f_{IJ} est éloigné du "produit des marges" $f_i f_j$, plus le Lien est élevé et plus le tableau f_{IJ} est instructif. Ainsi existe-t-il entre les caractères I et J une correspondance que le Lien (I,J) nous permet de connaître de façon détaillée.

On peut poser Lien (I,J) = $\sum_{ij} c_{ij}$ avec $c_{ij} = \frac{(f_{ij} - f_i f_j)^2}{f_i f_j}$

l'expression c_{ij} étant dénommée "contribution de la case (i,j) à l'information apportée f_{IJ} par le tableau f_{IJ} ". Cette décomposition de l'expression du Lien permet non seulement de clarifier rapidement et simplement un tableau de contingence, mais encore de caractériser par un graphique l'information apportée par chaque ligne ou chaque colonne, savoir $\sum_j c_{ij}$ pour la ligne et $\sum_i c_{ij}$ pour la colonne. Rappelons que nous avons (LAPLACE 1974) la même démarche méthodologique pour l'analyse d'un tableau des X^2 relatif à un tableau de contingence, démarche procédant par comparaison des sommations catégorielles des X^2 .

Comme dans le cas de la corrélation quadruple \emptyset , mesurant l'intensité de l'association de deux caractères dans un tableau de contingence 2 par 2, on peut affecter la valeur de chaque contribution de case c_{ij} du signe + ou du signe - selon que la quantité $(f_{ij} - f_i f_j)$ est positive ou négative, les sommations des c_{ij} ne portant, bien entendu, que sur les valeurs absolues des c_{ij} .

Par conséquent, nous dressons le tableau du Lien correspondant au tableau de contingence:

		Caractère J (k modalités)		
		j		
	i	$\pm c_{ij}$		$\sum_{j=1}^k c_{ij}$
		$\sum_{i=1}^r c_{ij}$		Lien (I,J)

Calculons le tableau du Lien pour le tableau de contingence des catégories ordinales dans les ensembles aurignacoïdes de Gatzarria:

	S	A	SE	E	B	
Cjn2	- 0,00108	+ 0,03386	- 0,01368	- 0,02385	+ 0,01404	0,08651
Cjn1	+ 0,00027	- 0,00001	+ 0,00047	- 0,00475	- 0,00002	0,00552
Cbf	+ 0,00032	- 0,02278	+ 0,00715	+ 0,02778	- 0,00910	0,06713
	0,00167	0,05665	0,0213	0,05638	0,02136	0,15916

que nous pouvons traduire en tableau des fréquences en faisant le quotient de chaque valeur de case et de chaque valeur marginale par la valeur du Lien:

	S	A	SE	E	B	
Cjn2	- .007	+ .213	- .086	- .150	+ .088	.544
Cjn1	+ .002	- .000	+ .003	- .030	- .000	.035
Cbf	+ .002	- .143	+ .045	+ .174	- .057	.421
	.010	.356	.134	.354	.146	1

3.3 Les cases importantes d'un tableau de contingence.

La case la plus importante étant, du point de vue de l'information, celle dont le c_{ij} est le plus grand, on classe les cases selon les c_{ij} décroissants en c_{ij} prenant soin de noter la fréquence de l'information contenue dans chaque case ainsi que la fréquence cumulée de l'information pour chaque case, c'est-à-dire la somme des fréquences de l'information de toutes les cases antérieures, jusqu'à la case considérée comprise:

Cjn2/A	.213		Cjn2/SE	.086	.854	Cjn1/SE	.003	.996
Cbf/E	.175	.387	Cbf/B	.057	.912	Cbf/S	.002	.998
Cjn2/E	.150	.537	Cbf/SE	.045	.957	Cjn1/S	.002	1
Cbf/A	.143	.680	Cjn1/E	.030	.986	Cjn1/B	.000	1
Cjn2/B	.088	.768	Cjn2/S	.007	.993	Cjn1/A	.000	1

Il est évident qu'une case apparaît d'autant plus aberrante dans le tableau des c_{ij} qu'elle occupe un rang plus élevé dans la séquence, toute anomalie pouvant être l'indice d'une différence significative d'un phénomène réel. Ainsi, si l'on se rapporte à la séquence des c_{ij} de Gatzarria on constate:

a - que la sommation des fréquences, à savoir la valeur des fréquences cumulées, franchit le seuil des 90% de l'information entre Cjn2/SE et Cbf/B;

b - que ces fréquences cumulées se rapportent exclusivement aux ensembles Cjn2 et Cbf et, principalement, aux catégories A et E, le rôle des catégories B et SE étant mineur.

En conclusion, si le calcul des c_{ij} , en nous donnant une mesure de la part d'information relative apportée par le croisement des caractères, met en lumière les associations séquentielles anormales, il ne saurait représenter que l'ébauche de l'analyse d'un tableau de contingence.

3.4 Le "spectre" d'une ligne ou d'une colonne.

Si l'on considère tous les c_{ij} des cases constitutives d'une même ligne ou d'une même colonne, on peut en donner une représentation graphique par un bloc-fréquences dénommé "spectre" de la ligne i ou de la colonne j .

Pour ce faire, on fait correspondre à chaque modalité de I ou de J un rectangle de surface proportionnelle à chaque c_{ij} , rectangle orienté vers le haut ou vers le bas selon que le c_{ij} est positif ou négatif. Ce "spectre", en visualisant la séquence de i ou de j , nous en permet une meilleure connaissance. C'est ainsi que la surface définie de part et d'autre de l'axe de la modalité i de I ou de la modalité j de J de chaque bloc-fréquences étant respectivement égale à $\sum_j c_{ij}$ ou à $\sum_i c_{ij}$, c'est-à-dire à l'information totale apportée par i ou par j , donne la mesure de leur anomalie. D'autre part, si l'on considère l'expression de la fré-

quence conditionnelle $f_j^i = \frac{f_{ij}}{f_i}$, on en tire la valeur $f_{ij} = f_j^i f_i$ que l'on porte dans l'expression:

$$c_{ij} = \frac{(f_{ij} - f_i f_j)^2}{f_i f_j} \quad \text{pour obtenir} \quad c_{ij} = \frac{(f_j^i f_i - f_i f_j)^2}{f_i f_j} = \frac{f_i^2 (f_j^i - f_j)^2}{f_i f_j}$$

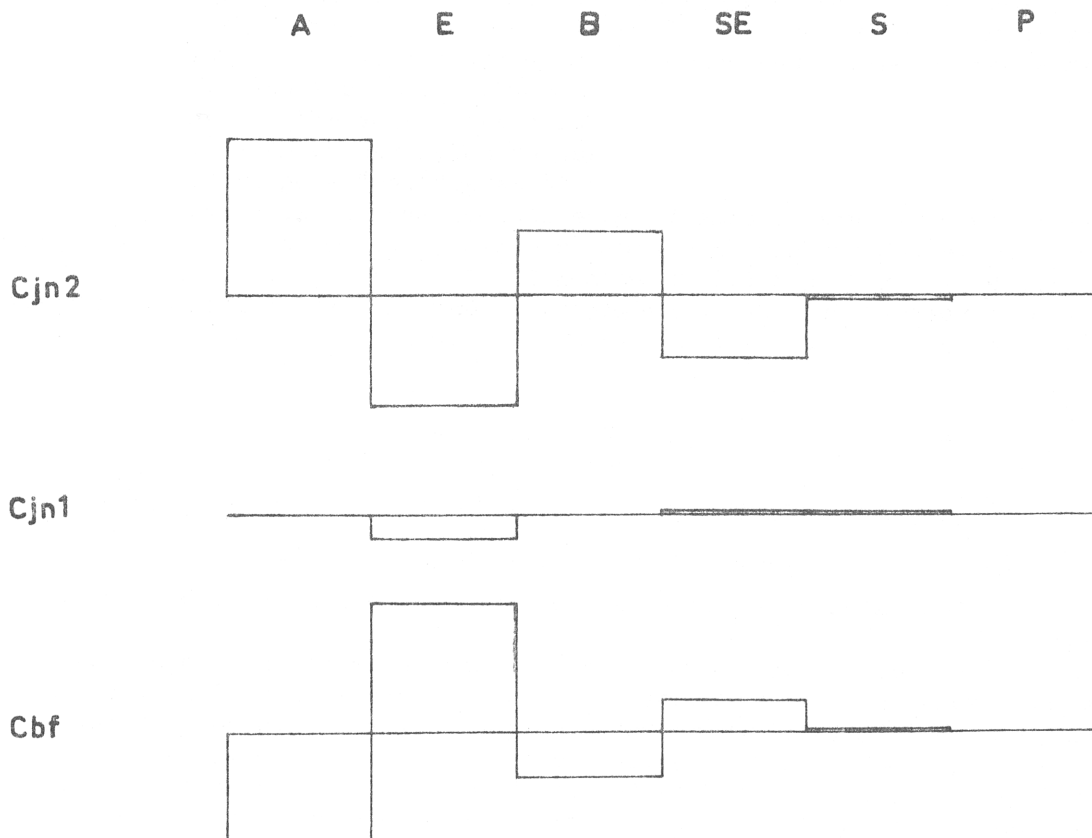
et, par conséquent, l'expression $c_{ij} = f_i \frac{(f_j^i - f_j)^2}{f_i}$

où, si l'on se réfère au tableau de contingence des catégories ordinales dans les ensembles aurignacoïdes de Gatzarria:

- a - f_i est la fréquence de l'ensemble i;
- b - f_j est la fréquence de la catégorie j;
- c - f_j^i est la fréquence conditionnelle de la catégorie j dans l'ensemble i.

Pour dresser le tableau des "spectres" des trois ensembles aurignacoïdes de Gatzarria, évidemment arrangés selon leur séquence diachronique, soit Cjn2, Cjn1 et Cbf, on ordonne les catégories typologiques selon la séquence dégressive des sommations marginales de leurs fréquences, soit A, E, B, SE, S et P:

	A	E	B	SE	S	P	
Cjn2	+ .213	- .150	+ .088	- .086	- .007544
Cjn1	- .000	- .030	- .000	+ .003	+ .002035
Cbf	+ .143	+ .174	- .057	+ .045	+ .002421
	.356	.354	.146	.134	.010	1



3.5 Du Lien à l'analyse structurale.

La contribution c_{ij} de la case (i,j) à l'information apportée par le tableau f_{IJ} n'est autre que l'écart quadratique réduit de sa fréquence dans le tableau de contingence. La sommation des c_{ij} par ligne ou par colonne, visualisée graphiquement par le "spectre" de ligne ou de colonne, exprime l'information apportée par chaque ligne ou par chaque colonne, soit, respectivement, $\sum_{j=1}^k c_{ij}$ ou $\sum_{i=1}^r c_{ij}$. Le Lien (I,J) est obtenu par la sommation soit de tous les c_{ij} , soit de tous les $\sum_{j=1}^k c_{ij}$, soit de tous les $\sum_{i=1}^r c_{ij}$. On conçoit que chaque c_{ij} constitutif d'une ligne ou d'une colonne traduise l'instabilité relative de la modalité j dans la ligne i ou de la modalité i dans la colonne j . De même, les quantités $\sum_{j=1}^k c_{ij}$ ou $\sum_{i=1}^r c_{ij}$ traduisent respectivement l'instabilité respective de la ligne i ou de la colonne j dans l'ensemble des lignes ou des colonnes. On notera que pour faciliter la lecture du tableau du Lien il est avantageux de le transformer en tableau des fréquences en faisant le quotient des valeurs des cases et des valeurs marginales de lignes et de colonnes par la valeur du Lien.

Examinons donc le tableau des fréquences du Lien dressé à partir du tableau de contingence des catégories ordinales dans les ensembles aurignacoïdes de Gatzarria. Interrogeons-nous sur les phénomènes d'instabilité mis en lumière et, éventuellement, visualisés par le diagramme des "spectres". Cette instabilité est-elle due à des fluctuations aléatoires ou est-elle significative d'une hétérogénéité réelle ? Pour répondre à cette question, on doit avoir recours à une épreuve diacritique: celle de l'application du test du Khi 2 à des tableaux de contingence 2 par k ou 2 par 2, dérivés du tableau de contingence r par k, comportant dans la première ligne les effectifs complémentaires cumulés des k-1 autres modalités du même caractère. Ainsi, pour nous limiter aux modalités du caractère "catégorie typologique", on testera successivement, dans l'ordre décroissant de leurs fréquences marginales dans le tableau des fréquences du Lien: A et non A (A / \bar{A}), E et non E (E / \bar{E}), B et non B (B / \bar{B}), SE et non SE (SE / \bar{SE}), S et non S (S / \bar{S}), P étant omis puisque nul. Par conséquent, on obtient les indicateurs d'écart χ^2 et les probabilités P, ordonnés selon la séquence dégressive des

$f \sum_i c_{ij}$:	A / \bar{A}	E / \bar{E}	B / \bar{B}	SE / \bar{SE}	S / \bar{S}
(f)	(.356)	(.354)	(.146)	(.134)	(.010)
χ^2	56,07	47,82	19,25	19,70	2,82
P	.000	.000	.000	.000	.245

Les valeurs calculées de la probabilité P pour chaque catégorie montrent:

- a - que l'hypothèse nulle H_0 d'une répartition homogène de la catégorie S dans les trois ensembles aurignacoïdes peut être retenue; en conséquence, l'instabilité de cette catégorie est attribuée au hasard;
- b - l'hétérogénéité hautement significative de la répartition des catégories SE, B, E et A dans les trois ensembles aurignacoïdes; l'instabilité de ces catégories étant significative d'une dynamique structurale, ces catégories sont, de ce fait, dénommées catégories sensibles.

Si l'on considère les ensembles industriels Cjn2 et Cjn1, on obtient, successivement, le tableau du Lien:

	S	A	SE	B	E	
Cjn2	- 0,00068	+ 0,00478	- 0,00644	+ 0,00197	- 0,00519	0,01906
Cjn1	+ 0,00146	- 0,01033	+ 0,01392	- 0,00426	+ 0,01121	0,04118
	0,00214	0,01511	0,02036	0,00623	0,01640	0,06024

le tableau des fréquences du Lien:

	S	A	SE	B	E	
Cjn2	- .011	+ .079	- .107	+ .033	- .086	.316
Cjn1	+ .024	- .171	+ .231	- .071	+ .186	.684
	.036	.251	.338	.103	.272	1

les indicateurs d'écart X^2 et les probabilités P, ordonnés selon la séquence dégressive des $f \sum_i c_{ij}$:

	SE / \overline{SE}	A / \overline{A}	E / \overline{E}	B / \overline{B}	S / \overline{S}
(f)	(.338)	(.272)	(.251)	(.103)	(.036)
X^2	8,95	8,51	6,53	2,68	1,73
P	<u>.003</u>	<u>.004</u>	<u>.031</u>	.102	.189

Les valeurs calculées de la probabilité P mettent en lumière:

a - la répartition homogène des catégories S et B dont l'instabilité est attribuée au hasard.

b - l'instabilité significative de la catégorie sensible E et très significative des catégories sensibles A et SE, révélatrice d'une dynamique évolutive dont le sens, indiqué par les signes du tableau du Lien, est positif pour E et SE et négatif pour A.

De la même manière, on obtient successivement, pour Cjn1 et Cbf, le tableau du Lien:

	S	SE	A	E	B	
Cjn1	+ 0,00003	- 0,00046	+ 0,00938	- 0,01527	+ 0,00395	0,02909
Cbf	- 0,00001	+ 0,00015	- 0,00298	+ 0,00485	- 0,00125	0,00923
	0,00005	0,00061	0,01236	0,02012	0,00520	0,03832

le tableau des fréquences du Lien:

	S	SE	A	E	B	
Cjn1	+ .001	- .012	+ .245	- .398	+ .103	.759
Cbf	- .000	+ .004	- .078	+ .127	- .033	.241
	.001	.016	.323	.525	.136	1

les indicateurs d'écart X^2 et les probabilités P, ordonnés selon la séquence

dégressive des $f \sum_i c_{ij}$:

	E / \bar{E}	A / \bar{A}	B / \bar{B}	SE / \bar{SE}	S / \bar{S}
(f)	(.525)	(.323)	(.136)	(.016)	(.016)
X^2	11,68	7,63	2,77	0,39	0,05
P	<u>.0006</u>	<u>.007</u>	.093	.534	.818

Les valeurs calculées de la probabilité P mettent en évidence:

- a - la répartition homogène des catégories S, SE et B dont l'instabilité est attribuée au hasard;
- b - l'instabilité très significative de la catégorie sensible A et hautement significative de la catégorie sensible E, révélatrice d'une dynamique évolutive dont le sens, déduit des signes du tableau du Lien, est positif pour E et négatif pour A.

Enfin, on peut comparer de la même façon les ensembles extrêmes Cjn2 et Cbf pour lesquels on obtient, successivement, le tableau du Lien:

	S	A	SE	E	B	
Cjn2	- 0,00096	+ 0,03984	- 0,01502	- 0,03184	+ 0,01615	0,10381
Cbf	+ 0,00066	- 0,02730	+ 0,01029	+ 0,02182	- 0,01106	0,07114
	0,00162	0,06713	0,02532	0,05367	0,02721	0,17495

le tableau des fréquences du Lien:

	S	A	SE	E	B	
Cjn2	- .005	+ .228	- .086	- .182	+ .092	.593
Cbf	+ .004	- .156	+ .059	+ .125	- .063	.407
	.009	.384	.145	.307	.156	1

les indicateurs d'écart X^2 et les probabilités P, ordonnés selon la séquence

dégressive des $f \sum_i c_{ij}$:

	A / \bar{A}	E / \bar{E}	B / \bar{B}	SE / \bar{SE}	S / \bar{S}
(f)	(.384)	(.307)	(.156)	(.145)	(.009)
X^2	55,95	38,66	19,04	19,62	2,26
P	<u>.000</u>	<u>.000</u>	<u>.000</u>	<u>.000</u>	.133

Les valeurs calculées de la probabilité P mettent en relief:

- a - une répartition homogène de la catégorie S dont l'instabilité est attribuée au hasard;
- b - une instabilité hautement significative des catégories sensibles SE, B, E et A, révélatrice d'une forte dynamique évolutive dont le sens, tiré des signes c_{ij} dans le tableau du Lien, est positif pour E, SE et S, et négatif pour A.

APPENDICE.

Programme de calcul du tableau du Lien pour Calculateur HP 67.

1. Instructions du programme.

01. f LBL E	08. RCL 0	15. RCL 2	22. $g x^2$	29. RCL 5	36. R/S
02. DSP 5	09. $\cdot/.$	16. RCL 4	23. RCL 5	30. -	37. RCL 7
03. h Rv	10. STO 3	17. x	24. $\cdot/.$	31. R/S	38. h RTN
04. RCL 0	11. RCL 1	18. STO 5	25. STO + 6	32. RCL 6	
05. $\cdot/.$	12. RCL 0	19. RCL 3	26. STO + 7	33. R/S	
06. STO 2	13. $\cdot/.$	20. RCL 5	27. R/S	34. 0	
07. h Rv	14. STO 4	21. -	28. RCL 3	35. STO 6	

2. Introduction des données.

1. N en STO 0
2. n_i ou n_j en STO 1
3. n_{ij} en ENTER
4. n_j ou n_i en ENTER

3. Demande des résultats.

1. E pour c_{ij}
2. R/S pour $\pm (f_{ij} - f_i f_j)$
3. R/S pour $\sum_j c_{ij}$ ou $\sum_i c_{ij}$
4. R/S pour effacer $\sum_j c_{ij}$ ou $\sum_i c_{ij}$
5. R/S pour Lien (I,J)

BIBLIOGRAPHIE.

- LAPLACE G. (1966) - Les niveaux castelperronien, protoaurignaciens et auri-
gnaciens de la Grotte Gatzarria à Suhare en Pays Basque. Fouilles
1961-1963. Quartär, 17, 117-140, 4 fig. 5 tabl.
- LAPLACE G. (1974) - De la dynamique de l'analyse structurale ou la typologie
analytique. Rivista di Scienze Preistoriche, XXIX, 1-71, 2 fig.
3 tabl. hors texte.
- VOLLE M. (1981) - Analyse des données. Economica, Paris, 320 p.