

F. BENZÉCRI

**Introduction à l'analyse des correspondances
d'après l'analyse du commerce mondial
des phosphates**

Les cahiers de l'analyse des données, tome 10, n° 2 (1985),
p. 145-190

http://www.numdam.org/item?id=CAD_1985__10_2_145_0

© Les cahiers de l'analyse des données, Dunod, 1985, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

INTRODUCTION A L'ANALYSE DES CORRESPONDANCES D'APRÈS L'ANALYSE DU COMMERCE MONDIAL DES PHOSPHATES

[CORR. ECO]

par F. Benzécri *

Le present exposé s'appuie sur l'exemple d'un tableau (14×8) choisi dans la thèse de S. Maïza (cf. [PHOSPHATES] ou C.A.D. Vol IX, n° 1 pp 7-32). La petite taille du tableau (§ 1) restreint l'ampleur de l'interprétation (§§ 5 et 6) mais elle permet de suivre dans le détail la définition du nuage des profils (§ 2) et le calcul de la distance du χ^2 (§ 3) ; et d'exploiter les aides à l'interprétation (§ 6) sur un listage publié *in extenso*. De plus, avec l'introduction d'elements supplémentaires (§ 7), est posé un problème de classification déjà complexe.

1 Présentation et première lecture des données

1.0 Le commerce mondial des phosphates au cours de la période 1973-1980 : On dispose d'un tableau ternaire croisant :

I : ensemble de 14 pays importateurs :

Belgique	JBL	Japon	JJP	Brésil	JBR
Canada	JCA	Nederland	JNL	Pologne	JPL
France	JFR	Espagne	JSP	Roumanie	JRM
Deutschland	JDL	United Kingdom	JUK	autres pays d'Eu-	
Italie	JIT	Inde	JIN	rope de l'Est (*)	JEE

(la première lettre J des sigles de ces 14 pays rappelle que ce sont des pays importateurs) ;

J : ensemble des 8 pays exportateurs (J) :

Belgique	JBL	Sénégal	J SN
USA	JUS	Togo	JTG
Jordanie	JJR	Tunisie	JTN
Maroc	JMR	URSS (CCCP)	JCC

T : ensemble des 8 années : 1973 à 1980.

Ce tableau ternaire, désigné par $k_{I,J,T}$ est défini ainsi :

Pour tout triple (i,j,t) d'un pays importateur i de I , d'un pays exportateur j de J et d'une année t de T , le tableau donne le nombre $k(i,j,t)$ qui exprime en milliers de tonnes de P_2O_5 la quantité de phosphate exportée par le pays j vers le pays i , l'année t . Par exemple,

(*) Docteur-ès-Sciences.

(*) Bulgarie, Hongrie, RDA.

pour $i = \mathcal{J}FRance$, $j = \mathcal{Z}MaRoc$, $t = 1973$, le tableau donne $k(\mathcal{J}FR, \mathcal{Z}MR, 73) = 575$: en 1973 le Maroc a exporté 575 milliers de tonnes de P_2O_5 vers la France.

Remarque 1 : Que le produit exporté soit du minerai brut, de l'acide phosphorique, de l'engrais, ou toute autre préparation à base de phosphates, on s'intéresse à la masse de P_2O_5 qu'il contient. Ainsi, une même quantité de P_2O_5 peut être exportée une première fois l'année t_1 sous forme de minerai du pays j_1 vers le pays i_1 où elle est élaborée, puis exportée une seconde fois l'année t_2 sous forme de produit phosphaté du pays $j_2 = i_1$ vers un troisième pays i_2 etc. ; en sorte que cette même quantité de P_2O_5 est comptée dans le nombre $k(i_1, j_1, t_1)$, dans le nombre $k(i_2, j_2, t_2)$ etc. . C'est le rôle d'intermédiaire du pays $j_2 = i_1$ que joue la Belgique qui, de ce fait, figure à la fois dans les ensembles I et J .

Remarque 2 : Les pays constituant I et J sont les principaux importateurs et exportateurs de produits phosphatés ou de minerais, à l'exclusion du marché quasi fermé de l'Océanie (Australie, Nouvelle-Zélande et producteurs insulaires de l'Océan Indien et Pacifique : Christmas, Banaba et Nauru).

Remarque 3 : Tous les nombres $k(i, j, t)$ sont positifs ou nuls. L'analyse des correspondances s'applique à des tableaux rectangulaires de nombres dits encore tableaux binaires (i.e. croisant seulement deux ensembles). C'est pourquoi l'on construit à partir du tableau ternaire de données, les tableaux binaires décrits au § 1.2 ci-dessous.

1.1 Les tableaux binaires issus du tableau ternaire.

Les marges : On appelle tableau binaire de marge du tableau $k_{I,J,T}$ tout tableau obtenu par sommation des $k(i, j, t)$ sur l'un des trois ensembles. On construit ainsi les 3 tableaux binaires :

$k_{I,J}$: tableau donnant pour tout couple (i, j) d'un pays importateur i de I et d'un pays exportateur j de J , le nombre $k(i, j) = \sum \{k(i, j, t) \mid t \in T\}$, masse totale de P_2O_5 exportée par le pays j vers le pays i au cours de toute la période 1973-1980.

(C'est ce tableau $k_{I,J}$ que nous avons choisi pour présenter en détail les constructions géométriques de l'a. des c. et l'interprétation des résultats de calcul).

$k_{I,T}$: $k(i, t) = \sum \{k(i, j, t) \mid j \in J\}$, masse totale de P_2O_5 importée par le pays i l'année t en provenance de l'ensemble des 8 pays exportateurs de J .

$k_{J,T}$: $k(j, t) = \sum \{k(i, j, t) \mid i \in I\}$, masse totale de P_2O_5 exportée par le pays j l'année t vers l'ensemble des 14 pays importateurs de I .

En outre, le tableau ternaire des $k(i, j, t)$ peut être disposé en tableau binaire (qui n'est pas un tableau de marge) de trois façon différentes, en prenant pour première entrée le produit de deux des ensembles I , J , T et pour deuxième entrée, le troisième ensemble :

$$k_{I \times T, J} : k((i, t), j) = k(i, j, t) ;$$

$$k_{J \times T, I} : k((j, t), i) = k(i, j, t) ;$$

$$k_{I \times J, T} : k((i, j), t) = k(i, j, t).$$

Nous utiliserons le premier de ces trois tableaux : $k_{I \times T, J}$ au § 7 pour exposer la technique des éléments supplémentaires : chaque pays importateur y figure huit fois, une fois par année ; e.g. FR73 ≠ FR74 ≠ ... ≠ FR80.

On remarquera que chacun des tableaux de marge a un ensemble d'entrée en commun avec l'un des trois tableaux binaires du type produit défini ci-dessus :

$k_{I, J}$ avec $k_{I \times T, J}$; $k_{J, T}$ avec $k_{I \times J, T}$; $k_{I, T}$ avec $k_{J \times T, I}$.

Considérons par exemple $k_{I, J}$ et $k_{I \times T, J}$. Pour tout j de J et tout i de I , le tableau $k_{I, J}$ donne un nombre $k(i, j)$ et le tableau $k_{I \times T, J}$ donne huit nombres dont $k(i, j)$ est la somme :

$$k(i, j) = k(i73, j) + k(i74, j) + \dots + k(i80, j).$$

Si l'on dispose ces tableaux en rectangles en réservant :

pour chaque pays j de J , une colonne ;

pour chaque pays i de I , une ligne dans $k_{I, J}$,

huit lignes dans $k_{I \times T, J}$;

on peut les placer l'un sous l'autre.

j

	EBL	EUS	EJR	EMR	ESN	ETG	ETN	ECC	marge
JBL		1305		3573	25	500	110	293	5806
JCA	2	8335				8			8345
JFR	1311	2691	70	4891	1484	2526	1697	4	14674
JDL	1322	3808		1445	261	200	288	1442	8766
JIT	42	1883	194	2881	67	195	493	ε	5755
JJP	ε	4426	522	1540	239	93		ε	6820 ← k(i)
JNL	299	1484		1853	249	1584	59	84	5612
JSP	20	339		5073	2	85	11		5530
k(i, j) JUK	122	645	2	2852	971	36	197		4825
JIN		2559	996	1149	218		134		5056
JBR	29	4918		1398	15		241		6601
JPL	ε	1284	271	3311	33	540	548	1996	7983
k(j) JRM	ε	768	483	1541	6	138	22	1206	4164
JEE	2	201	136	1398			333	5533	7603
marge	3149	34646	2674	32905	3570	5905	4133	10558	97540 ← k
BL73		111		470	5	45	196	105	
BL74		75		482	5	112	186	60	
BL75		75		400		36	63	33	
....		...		...	...	...	...	...	
BL80		270		500		77	126	283	
CA73		1004							
CA74		1020				8			
....		...		...	...	...	...	...	

Les nombres représentent des milliers de tonnes de P_2O_5 ; les cases laissées blanches correspondent à des masses nulles ; les ε à des masses inférieures à 1000 tonnes ; les marges donnent les totaux. Jusqu'au § 7, nous ne nous occuperons que du tableau $k_{I, J}$.

Le tableau de nombres positifs ou nuls $k_{I,J}$ est appelé tableau de correspondance entre les ensembles I (lignes) et J (colonnes). Dans les calculs expliqués ci-dessous qui constituent l'analyse de ce tableau, les deux ensembles I et J jouent des rôles symétriques, en sorte qu'on ne changerait rien aux résultats en mettant en lignes les pays exportateurs (J) et en colonnes les pays importateurs (I). Toutefois, les calculs pouvant être conduits suivant deux voies différentes qui aboutissent aux mêmes résultats en des temps inégaux, certains programmes requièrent que le tableau soit présenté avec un nombre minimum de colonnes et introduit ligne par ligne pour que les calculs soient le plus rapides possible ; il va sans dire que le choix de la voie la plus rapide peut être incorporé dans le programme, auquel, dès lors, la présentation du tableau est indifférente.

1.2 Calcul des marges et de leurs profils : La première opération mathématique effectuée pour prendre connaissance du tableau de correspondance consiste à calculer les totaux des lignes et des colonnes. En inscrivant au bout de chaque ligne i son total $k(i)$ on obtient la *colonne de marge* du tableau $k_{I,J}$; en inscrivant au bas de chaque colonne j son total $k(j)$, on obtient la *ligne de marge*.

La colonne de marge donne pour chaque pays importateur i de I la masse totale $k(i)$ de P_2O_5 qu'il a importée durant la période 1973-80, sans distinction d'origine ; de même, la ligne de marge donne pour chaque pays exportateur j de J la masse totale $k(j)$ de P_2O_5 qu'il a exportée durant la période 1973-80, sans distinction de destination.

Ligne et colonne de marge ont même total, noté k , encore égal au total général du tableau.

$$k = \sum \{k(i,j) \mid i \in I ; j \in J\} = \sum \{k(i) \mid i \in I\} = \sum \{k(j) \mid j \in J\}.$$

En rapportant les 14 nombres $k(i)$ à leur total k , on obtient le *profil de la colonne de marge* :

$$\{(k(i)/k) \mid i \in I\} ; \text{ (profil sur I) ;}$$

pour chaque i , le rapport $k(i)/k$ mesure l'importance relative de i au sein de I ; plus précisément, la part afférente à i dans le total des importations recensées des phosphates. Par exemple la France représente 150/1000 du total des importations ; l'Allemagne, 90/1000 ; etc.. Dans la suite, $k(i)/k$ sera appelé *poïds de i* et noté f_i :

$$f_i = k(i)/k.$$

On définit de la même façon le *profil de la ligne de marge* en divisant chaque total de colonne $k(j)$ par le total général k :

$$\{k(j)/k \mid j \in J\}, \text{ (profil sur J),}$$

ce qui donne pour chaque j de J l'importance relative du pays j au sein de J : part afférente à j dans le total des exportations recensées des phosphates. Dans la suite, le rapport $k(j)/k$ sera appelé *poïds de j* et noté f_j :

$$f_j = k(j)/k$$

Ligne et colonne de marge du t. $k_{I,J}$, i.e. $\{k(j) \mid j \in J\}$ et $\{k(i) \mid i \in I\}$ peuvent être considérées (nous le verrons au § 2.3) comme définissant par leurs profils des pays fictifs moyens : pays importateur moyen pour la ligne, pays exportateur moyen pour la colonne. C'est pourquoi au § 1.3 le profil de la ligne de marge i.e. $\{f_j \mid j \in J\}$

figure à la dernière ligne du tableau des profils des lignes de $k_{I,J}$ et le profil de la colonne de marge i.e. $\{f_i | i \in I\}$ figure à la dernière colonne du tableau des profils des colonnes de $k_{I,J}$; tandis que comme poids, $\{f_i | i \in I\}$ constitue la colonne POIDS du tableau des profils des lignes, et $\{f_j | j \in J\}$ la ligne POIDS du t. des profils des colonnes (cf. § 1.3).

Remarque : $\Sigma\{f_i | i \in I\} = 1$ et $\Sigma\{f_j | j \in J\} = 1$.

1.3 Profils des lignes et des colonnes du tableau $k_{I,J}$: Considérons le tableau $k_{I,J}$ et ses marges et cherchons à comprendre en quoi les importations d'un pays i de I en provenance des divers pays de J diffèrent de celles d'un autre pays i' , ou de celles de la marge (pays importateur moyen). (Ensuite, nous comparerons de même les exportations de deux pays de J).

Par exemple, comparons les importations de JFRance et de JNederland. Pour tout j différent de $\mathcal{Z}_{CC}$, $k(\text{JFR}, j)$ est supérieur à $k(\text{JNL}, j)$; e.g. :

$$\begin{aligned} k(\text{JFR}, \mathcal{Z}_{BL}) &= 1311 ; k(\text{JFR}, \mathcal{Z}_{MR}) = 4891 \\ k(\text{JNL}, \mathcal{Z}_{BL}) &= 299 ; k(\text{JNL}, \mathcal{Z}_{MR}) = 1853. \end{aligned}$$

Mais que représentent ces nombres pour chacun des deux pays ? En provenance de la Belgique,

1311 représente $1311/14674 = 8\%$ des importations totales de JFR ;
299 représente $299/5612 = 5\%$ des importations totales de JNL ;
en provenance du Maroc,
4891 représente $4891/14674 = 33\%$ des importations totales de JFR ;
1853 représente $1853/5612 = 33\%$ des importations totales de JNL.

Ainsi, la Belgique, tout en exportant 4,4 fois plus vers la France que vers les Pays-Bas n'a que 1,6 fois plus d'importance au sein de J pour la France que pour les Pays-Bas. Le Maroc, bien qu'exportant 2,6 fois plus vers la France que vers les Pays-Bas, a une part égale dans les importations de l'un et de l'autre pays.

Afin de faire apparaître, comme dans ces exemples, la part des divers pays de J dans les importations de chacun des pays de I , on construit le tableau des *profils des lignes* en divisant chaque ligne i par son total $k(i)$:

profil de la ligne i : $f_j^i = \{k(i,j)/k(i) | j \in J\}$, (profil sur J).

Le rapport $k(i,j)/k(i) = f_j^i$, composante sur j du profil de i est la part relative du pays j dans le total des importations du pays i .

Remarquons que l'on a procédé de même au § précédent pour obtenir le profil de la ligne de marge : on a divisé la ligne par son total. On a adjoint au tableau des profils des lignes le profil de la ligne de marge ; ce qui permet de comparer chaque pays i de I au pays fictif moyen représenté par la ligne de marge. Ainsi apparaît le peu d'importance des USA dans les importations des pays de l'Est, de l'Espagne, de la France, de l'Angleterre ; leur importance dans les importations du Canada, du Brésil, du Japon, de l'Inde ; l'importance du Maroc dans les importations de l'Espagne, de la Belgique, de l'Angleterre, de l'Italie etc.

Au tableau des profils des lignes nous adjoignons également la colonne POIDS donnant pour chaque pays importateur i le rapport $k(i)/k = f_i$ mesurant l'importance de i au sein de I (cf. § 1.2).

		j								
I \ J	JBL	JUS	JJR	JMR	JSN	JTG	JTN	JCC	POIDS	
i	JBL	225		615	4	86	19	51	60	
	JCA	ε	999			1			86	
	JFR	89	183	5	333	101	172	116	150	
	JDL	151	434		165	30	23	33	90	
	JIT	7	327	34	501	12	34	86	59	
	JJP	ε	649	77	226	35	14		70	
	JNL	53	264		330	44	282	11	15	
	JSP	4	61		917	ε	15	2	58	
	JUK	25	134	8	591	201	8	41	57	
	JIN		506	197	227	43		26	49	
	JBR	4	745		212	2		36	52	
	JPL	ε	161	34	415	4	68	69	250	
	JRM	ε	184	116	370	1	33	5	290	
	JEE	ε	26	18	184			44	728	
marge		32	355	27	337	37	61	42	108	
									1000	

$f_j^i = \frac{k(i,j)}{k(i)}$

$f_j = \frac{k(j)}{k}$

$\frac{k(i)}{k} = f_i$

$\frac{k}{k}$

Tableau f_{IJ}^I : Profils des lignes du tableau k_{IJ} et profil de la ligne de marge de ce même t . ; la colonne POIDS indique pour chaque i son importance relative au sein de I (cf. § 1.2). Les nombres sont des millièmes.

		j								
	i \ j	BL	US	JR	MR	SN	TG	TN	CC	marge
i	BL		38		109	7	85	27	28	60
	CA	ε	241				1			86
	FR	416	78	26	149	416	428	411	ε	150
	DL	420	110		44	73	34	70	137	90
	IT	13	54	72	88	19	33	119	ε	59
	JP	ε	128	195	47	67	16		ε	70
	NL	95	43		55	70	268	14	8	58
	SP	6	10		154	1	14	3		57
	UK	39	19	1	87	272	6	48		49
	IN		74	373	35	61		32		52
	BR	9	142		43	4				68
	PL	ε	37	102	101	9	91	133	189	82
	RM	ε	22	181	47	2	23	5	114	43
	JEE	1	6	51	42			81	524	78
POIDS		32	355	27	337	37	61	42	108	1000

$f_j^i = \frac{k(i,j)}{k(j)}$

$f_j = \frac{k(j)}{k}$

$\frac{k(i)}{k} = f_i$

$\frac{k}{k}$

Tableau f_{IJ}^J : Profils des colonnes du tableau k_{IJ} et profil de la col. de marge de ce même t . ; la ligne POIDS indique pour chaque j son importance relative au sein de J (cf. § 1.2). Les nombres sont des millièmes.

De même, pour les colonnes du tableau k_{IJ} : si l'on cherche à comprendre en quoi les exportations d'un pays j de J vers les divers pays de I diffèrent de celles d'un autre pays j' , ou de celles du pays exportateur moyen (colonne de marge du tableau k_{IJ}), on est amené à rapporter tous les nombres d'une colonne à leur total, i.e. à constituer le tableau des profils des colonnes du t . k_{IJ} , (profils sur I):

profil de la colonne j : $f_I^j = \{k(i,j)/k(j) \mid i \in I\} = \{f_i^j \mid i \in I\}$;

profil de la colonne de marge : $f_I = \{k(i)/k \mid i \in I\} = \{f_i \mid i \in I\}$.

La composante sur i du profil de j : f_i^j mesure la part relative du pays i dans les exportations du pays j ; on peut la comparer à la composante sur i d'un autre pays j' , ou à celle du pays moyen qui figure en colonne de marge, profil de la col. de marge du t. $k_{I,J}$. A ce tableau on adjoint la ligne : POIDS, profil de la ligne de marge du tableau $k_{I,J}$, qui donne pour chaque j le rapport $f_j = k(j)/k$, mesure de l'importance de j au sein de J .

Dans l'analyse des correspondances que nous allons effectuer, un élément i de I sera représenté par son profil f_I^i ; un élément j de J , par son profil f_I^j . Il convient de noter ici la

Propriété des profils : La somme des composantes d'un profil est égale à 1.

En effet, considérons par exemple un profil sur J :

$$f_J^i = \{k(i,j)/k(i) \mid j \in J\} \text{ avec } k(i) = \sum \{k(i,j) \mid j \in J\} ;$$

On a :

$$\sum \{k(i,j)/k(i) \mid j \in J\} = (\sum \{k(i,j) \mid j \in J\})/k(i) = 1.$$

Et de même pour un profil sur I .

2 Les nuages de profils

2.1 Le nuage $N(I)$ dans le simplexe P_J des profils sur J .

2.1.1 L'espace vectoriel R_J des mesures sur J : En analyse des correspondances on utilise des lettres affectées d'indices haut ou bas, comme nous l'avons fait ici dès le § 1. Par exemple, dans f_j ou f_j^i , j est un indice bas ; dans f_i^j , i un indice haut. Nous insistons sur l'importance de cette distinction, manifestement essentielle, puisque $f_j^i = k(i,j)/k(i)$, tandis que $f_i^j = k(i,j)/k(j)$. Sans expliquer à fond ces notations issues du calcul tensoriel, nous dirons qu'un système μ_j de nombres de signes quelconques affectés d'un indice bas j parcourant l'ensemble J est appelé *mesure* (ou système de masses) sur J , chacun de ces nombres μ_j étant une composante de μ_J :

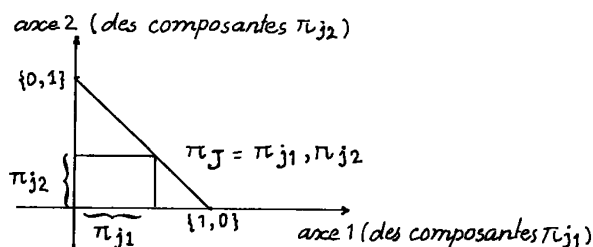
$$\mu_J = \{\mu_j \mid j \in J\}.$$

En particulier, $f_J^i = \{f_j^i \mid j \in J\}$ (profil de i) et $f_J = \{f_j \mid j \in J\}$ (profil de marge) sont des mesures sur J . L'ensemble de toutes les mesures sur J est noté R_J ; on écrit donc :

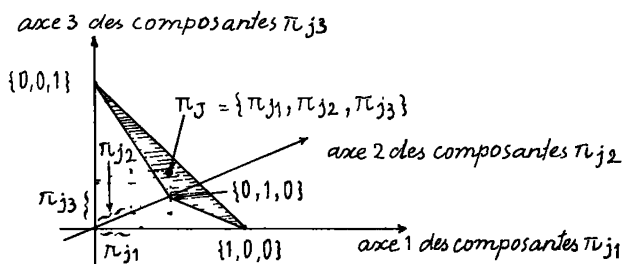
$$\mu_J \in R_J ; f_J^i \in R_J ; f_J \in R_J.$$

L'ensemble R_J est un espace vectoriel ayant autant de dimensions que J a d'éléments. Par exemple, si J n'a que deux éléments j_1 et j_2 : $J = \{j_1, j_2\}$, R_J est un plan ; si $J = \{j_1, j_2, j_3\}$ R_J est un espace usuel de dimension 3.

2.1.2 Le simplexe P_J des profils sur J : Plus précisément, ainsi qu'on l'a noté à la fin du § 1.3, un profil tel que f_J^i ou f_J est une mesure dont les composantes sont toutes positives ou nulles et ont pour somme 1. Reprenons le cas $J = \{j_1, j_2\}$; un profil π_J est un système de deux composantes $\{\pi_{j_1}, \pi_{j_2}\}$ positives ou nulles dont la somme est 1 ; le point π_J est donc situé, dans le plan, sur le segment qui joint les points $\{1, 0\}$ et $\{0, 1\}$ situés respectivement sur l'axe 1 et l'axe 2.



Dans le cas où $J = \{j_1, j_2, j_3\}$, un profil est de même caractérisé par la propriété d'être un point de la portion de plan triangulaire dont les trois sommets sont les extrémités des vecteurs unitaires des axes : $\{1, 0, 0\}$, $\{0, 1, 0\}$, $\{0, 0, 1\}$.



Dans le cas général où J a un nombre quelconque n d'éléments, les profils sur J sont des systèmes de n composantes positives ou nulles de somme 1 :

$$\pi_J = \{\pi_j | j \in J\} ; \forall j \in J: \pi_j \geq 0 ; \sum \{\pi_j | j \in J\} = 1.$$

L'ensemble de tous les profils sur J est, dans l'espace R_J (de dimension n) des mesures sur J , le "simplexe" (généralisation à la dim. n du triangle en dim. 3) de sommets les n points : $\{1, 0, \dots, 0\}$, $\{0, 1, 0, \dots, 0\}$ etc. extrémités des vecteurs unitaires des axes. On note P_J ce simplexe. Pour $n = 2$, P_J est un segment de droite (de dim. 1) ; pour $n = 3$, P_J est une portion triangulaire de plan (de dim. 2) ; pour n quelconque, P_J est une portion d'espace de dimension $n - 1$, espace qu'on appelle *support* du simplexe P_J .

Dans notre exemple où J est l'ensemble de 8 pays exportateurs, R_J a 8 dimensions et dans cet espace à 8 dimensions le simplexe des

profils P_J est une portion d'espace à 7 dimensions. On écrira :

$$\pi_J \in P_J \subset R_J \quad ; \quad f_J^i \in P_J \subset R_J$$

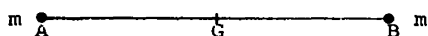
(les profils π_J, f_J^i sont des points du simplexe P_J , lequel est une partie de l'espace R_J des mesures sur J).

2.1.3 Le nuage $N(I)$: Les 14 points représentant les 14 profils de pays importateurs sont dispersés dans le simplexe P_J des profils sur J . En outre, chaque point f_J^i représentant le profil d'un pays importateur i est muni du poids de ce pays : $f_i = k(i)/k$, poids qui mesure l'importance relative de i au sein de I . Cet ensemble de 14 points géométriques pesants dans l'espace à 8 dimensions constitue la représentation géométrique de l'ensemble I des 14 pays importateurs et est appelé *nuage $N(I)$*

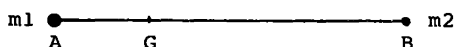
$$N(I) = \{(f_J^i ; f_i | i \in I) \subset P_J$$

2.2 Centre de gravité du nuage

Etant donné, dans l'espace usuel de la géométrie élémentaire, deux points A et B de même masse m , le centre de gravité de ces deux points n'est autre que le milieu du segment les joignant.



Si les deux points A et B sont doués de masses inégales, m_1 pour A , m_2 pour B , avec $m_1 \neq m_2$, le centre de gravité est comme attiré par le point le plus lourd ; de façon précise, c'est le point G du segment AB tel que : $|GA|/|GB| = m_2/m_1$.



Si les points A et B sont situés sur un axe et repérés par leurs abscisses $x(A)$ et $x(B)$ sur cet axe, l'abscisse de G sur ce même axe s'obtient par la formule :

$$x(G) = (1/(m_1 + m_2)) (m_1 x(A) + m_2 x(B))$$

$x(G)$ est la moyenne, pondérée par m_1 et m_2 , de $x(A)$ et de $x(B)$. Cette formule se déduit aisément de la proportion indiquée ci-dessus entre les longueurs de segments GA et GB .

Si les points A et B sont situés dans un plan rapporté à deux axes de coordonnées et repérés par leurs abscisses et ordonnées sur ces deux axes : $x(A)$, $y(A)$ et $x(B)$, $y(B)$, les coordonnées de G sur ces mêmes axes s'obtiennent par des formules semblables :

$$x(G) = (1/(m_1 + m_2)) (m_1 x(A) + m_2 x(B))$$

$$y(G) = (1/(m_1 + m_2)) (m_1 y(A) + m_2 y(B))$$

Ces formules se généralisent à un nombre quelconque de points pesants et à un nombre quelconque de dimensions en ce qui concerne l'espace où sont les points.

Pour le nuage $N(I)$ le nombre de points est : 14, et le nombre de dimensions de l'espace ambiant : 8. En outre, les masses attachées à chaque point sont telles que leur somme étendue à tous les points vaut 1 ;

en effet, on a (cf. § 1.2) : masse du point $i = k(i)/k$, où k est la somme des $k(i)$ étendue aux 14 pays importateurs ; quand on fait la somme des 14 rapports $(k(i)/k)$ on obtient 1. Les formules donnant les coordonnées du centre de gravité se simplifient (puisque le coefficient $(1/(m_1 + m_2 + \dots))$ vaut 1), et il vient :

coordonnée sur $\mathcal{ZBL}$ du centre de gravité de $N(I)$:

$$(k(\mathcal{JBL})/k) (k(\mathcal{JBL}, \mathcal{ZBL})/k(\mathcal{JBL})) +$$

$$(k(\mathcal{JCA})/k) (k(\mathcal{JCA}, \mathcal{ZBL})/k(\mathcal{JCA})) +$$

....

$$(k(\mathcal{JEE})/k) (k(\mathcal{JEE}, \mathcal{ZBL})/k(\mathcal{JEE})) .$$

↑
masses

↑
coordonnées sur $\mathcal{ZBL}$ des profils des pays de I .

Dans chaque terme de la somme relatif à un pays importateur i une simplification par $k(i)$ s'opère et il reste pour la coordonnée sur $\mathcal{ZBL}$ du centre de gravité de $N(I)$:

somme des $k(i,j)$ relatifs à la colonne $\mathcal{ZBL}$, divisée par k , total de tableau initial.

Cette coordonnée n'est autre que le nombre 32 figurant dans le tableau des profils des lignes (§ 1.3) à la croisée de la ligne de marge et de la colonne $\mathcal{ZBL}$.

De même, les autres coordonnées du centre de gravité se lisent à la suite dans la ligne de marge du tableau des profils des lignes. Ainsi, cette ligne de marge n'est autre que le profil, centre de gravité du nuage $N(I)$.

$$f_j = \{f_j | j \in J\} = \{k(j)/k | j \in J\} \in P_J.$$

↑
total de la colonne j

C'est pourquoi nous annonçons au § 1.2 que la ligne de marge du tableau $k_{I,J}$ définissait un pays fictif importateur moyen : nous voyons ici plus précisément que, ayant pris son profil comme pour les 14 pays réels, on a obtenu le centre de gravité des 14 profils de pays importateurs, chacun muni de son poids.

Remarque : Pour un sous-nuage de $N(I)$, i.e. pour une partie de $N(I)$, le centre de gravité se détermine de façon analogue à partir du tableau $k_{I,J}$.

Par exemple, considérons les trois points de $N(I)$ représentant les pays importateurs Allemagne, Pologne et Roumanie : $\mathcal{JDL}$, $\mathcal{JPL}$, $\mathcal{JRM}$; (nous verrons que dans la classification ascendante hiérarchique ces trois pays forment une classe caractérisée par de fortes importations en provenance de l'URSS). Dans l'espace à 8 dimensions, les profils de ces trois pays munis chacun de leur poids : 90 pour $\mathcal{JDL}$, 82 pour $\mathcal{JPL}$, 43 pour $\mathcal{JRM}$ (cf. § 1.3 tableau f_J^I) constituent un sous-nuage de $N(I)$. Le centre de gravité de ce nuage s'obtient en prenant le profil de la ligne cumulant les trois lignes $\mathcal{JDL}$, $\mathcal{JPL}$ et $\mathcal{JRM}$ du tableau $k_{I,J}$. Le t. $k_{I,J}$ donne :

	$\mathcal{ZBL}$	$\mathcal{ZUS}$	$\mathcal{ZJR}$	$\mathcal{ZMR}$	$\mathcal{ZSN}$	$\mathcal{ZTG}$	$\mathcal{ZTN}$	$\mathcal{ZCC}$	marge
$\mathcal{JDL}$	1322	3808		1445	261	200	288	1442	8766
$\mathcal{JPL}$	ε	1284	271	3311	33	540	548	1996	7983
$\mathcal{JRM}$	ε	768	483	1541	6	138	22	1206	4164

d'où la ligne (que nous appellerons DPR) obtenue en faisant la somme des trois lignes JDL, JPL et JRM, colonne par colonne.

	BL	EU	JR	MR	SN	TG	TN	CC	marge
J DPR	1322	5860	754	6297	300	878	858	4644	20913

le dernier nombre de la ligne étant son total : 20913.

Le profil sur J de la ligne J DPR (i.e. le centre de gravité des profils des trois pays JDL, JPL et JRM) se calcule comme pour toute ligne du tableau $k_{I,J}$, en divisant les nombres qui y figurent par le total de la ligne : 20913. On obtient :

	BL	US	JR	MR	SN	TG	TN	CC	POIDS
J DPR	.063	.280	.036	.301	.014	.042	.041	.222	.214

le poids : .214 est le rapport du total 20913 de la ligne J DPR au total général du tableau $k_{I,J}$: 97540.

2.3 Le nuage $N(J)$: Comme nous l'avons annoncé § 1.1 *in fine*, les ensembles I et J jouent des rôles symétriques : tout ce que nous avons dit de I relativement à J peut se dire de J relativement à I : chaque pays exportateur j de J est décrit, dans le tableau des profils sur I par 14 rapports positifs de somme 1 : $f_I^j = \{f_{I,i}^j \mid i \in I\}$ et son poids f_j .

Un système de 14 nombres étant considéré comme un point dans un espace de dimension 14, tout profil f_I^j est un point de R_I , espace de dim. 14 des mesures sur I. Mais comme f_I^j est un *profil*, il est situé dans le simplexe P_I des profils sur I, lui-même entièrement compris dans un sous-espace de dimension 14 : $14 - 1 = 13$.

Le nuage $N(J)$ est l'ensemble des 8 points pesants représentant chacun un pays exportateur de J :

$$N(J) = \{(f_I^j, f_j) \mid j \in J\} \subset P_I$$

Le centre de gravité de $N(J)$ est le profil de la colonne de marge :

$$f_I = \{f_{I,i} \mid i \in I\} = \{(k_{I,i}^{\text{total}})/k \mid i \in I\} \in P_I.$$

↑
total de la ligne i

Pour un sous-nuage $N(J')$ de $N(J)$: $J' \subset J$, le centre de gravité de $N(J')$ s'obtient en prenant le profil de la colonne cumulant les colonnes de J' dans le tableau $k_{I,J}$.

2.4 Le modèle d'indépendance : Considérons un tableau $k_{I,J}$ dont toutes les lignes aient même profil :

$$(1) \quad \forall i, i' \in I, f_I^i = f_I^{i'}$$

c'est-à-dire (en lisant "quel que soit" ou "pour tout" le symbole $\forall$) :

$$\forall i, i' \in I \text{ et } j \in J : k(i,j)/k(i) = k(i',j)/k(i').$$

On en déduit :

$$\begin{aligned} \forall i \in I, \forall j \in J : k(i,j)/k(i) &= \sum \{k(i,j) \mid i \in I\} / \sum \{k(i) \mid i \in I\} \\ &= k(j)/k ; \end{aligned}$$

d'où

$$(2) \quad \forall i \in I, \forall j \in J : k(i, j) = (k(i) \times k(j))/k$$

ou encore,

$$(2') \quad \forall i \in I, \forall j \in J : k(i, j) = k(k(i)/k)(k(j)/k) = k f_i f_j$$

De la relation (2) il résulte :

$$(3) \quad \forall j, j' \in J : f_I^j = f_I^{j'} = f_I.$$

En effet, d'après (2), $\forall i \in I, \forall j \in J, \forall j' \in J :$

$$k(i, j)/k(j) = k(i, j')/k(j') = k(i)/k.$$

Ainsi, non seulement toutes les lignes, y compris la ligne de marge, ont même profil, mais encore toutes les colonnes, y compris la colonne de marge, ont même profil. De même, si toutes les colonnes d'un tableau de correspondance ont même profil, la colonne de marge a même profil qu'elles et toutes les lignes, y compris la ligne de marge, ont même profil.

Nous dirons qu'un tel tableau, dont toutes les lignes ont même profil, ainsi que toutes les colonnes, satisfait au modèle d'indépendance.

Dans l'espace de dimension 8 des mesures sur J, les 14 profils constituant le nuage N(I) coïncident avec le centre de gravité du nuage ; et de même, dans l'espace de dimension 14 des profils sur I, le nuage N(J) est réduit à son centre de gravité.

S'il en était ainsi dans notre exemple, tout pays j de J aurait même part dans les importations des divers pays de I : cette part serait indépendante du pays i considéré dans I ; et de même tout pays i de I aurait même part dans les exportations des divers pays de J : cette part serait indépendante du pays j considéré dans J. Ce qui s'accorde avec le nom du modèle : modèle d'indépendance.

De plus, la formule (2') montre que le nombre $k(i, j)$ résulte exclusivement du poids de i et du poids de j, sans qu'intervienne aucune affinité entre i et j. L'analyse d'un tel tableau est sans objet, puisque l'analyse des correspondances du $t_{I, J}$ vise essentiellement à mettre en évidence les affinités qui existent entre certains éléments de I et certains éléments de J. Mais le modèle d'indépendance sert de référence : on parlera d'interaction attractive ou répulsive entre i et j selon que $k(i, j)$ est supérieur ou inférieur à $k f_i f_j = k(i) k(j)/k$.

Considérons par exemple les interactions entre d'une part, JFRance et d'autre part ZMaRoc, ZUSa et ZToGo ; les deux dernières colonnes du tableau ci-dessous donnant les valeurs calculées selon l'indépendance, puis les valeurs réelles.

i	j	k(i)	k(j)	$k(i)k(j)/k$	$k(i, j)$
JFR	ZMR	14674	32905	4950	4891
JFR	ZUS	14674	34646	5212	2691
JFR	ZTG	14674	5905	888	2526

$$k = 97540$$

entre JFR et ZUS, il y a répulsion : $2691 < 5212$
entre JFR et ZTG, il y a attraction : $2526 > 888$
entre JFR et ZMR, il n'y a pas d'interaction : $4891 \approx 4950$

Au § 1.3, pour interpréter les nombres $k(i,j)$ du tableau $k_{I,J}$ nous les avons rapportés d'une part au total de la ligne i ($f_j^i = k(i,j)/k(i)$, composante sur j du profil de i), d'autre part au total de la colonne j ($f_i^j = k(i,j)/k(j)$, composante sur i du profil de j). On dispose ainsi de trois moyens équivalents d'apprécier les interactions entre les i et les j . Par exemple une attraction entre i et j se manifestera des trois façons suivantes :

1°) f_j^i grand vis-à-vis de $f_j = k(j)/k$: la composante sur j du profil de i grande vis-à-vis de celle du profil moyen sur J . La part de j dans les importations de i est supérieure à la part de j dans les importations mondiales (représentées par la ligne de marge).

2°) f_i^j grand vis-à-vis de $f_i = k(i)/k$: la part de j dans les exportations de i est supérieure à la part de j dans les exportations mondiales (représentées par la colonne de marge).

3°) $k(i,j) > k(i)k(j)/k$, ce qui s'écrit encore, en posant $f_{ij} = k(i,j)/k$; $f_i = k(i)/k$; $f_j = k(j)/k$:

$$f_{ij} > f_i f_j$$

Dans 1°) on compare f_j^i et f_j ; dans 2°) f_i^j et f_i ; dans 3°) f_{ij} et $f_i f_j$. Or on a l'égalité des rapports :

$$f_j^i / f_j = f_i^j / f_i = f_{ij} / f_i f_j = (k(i,j) \times k) / (k(i) \times k(j)) ;$$

d'où l'équivalence de 1°), 2°) et 3°).

3 Dispersion et distance

Pour un nuage de points alignés sur une droite la notion de dispersion est clairement visible. Au § 3.0 nous définissons dans ce cas les notions classiques : moyenne des distances des points au centre de gravité, écart-type, variance, inertie.

Pour un nuage multidimensionnel, la notion même de distance entre deux points ne s'impose pas *a priori* : on se heurte à nouveau aux difficultés rencontrées au § 1.3 pour comparer deux lignes du tableau $k_{I,J}$. Il faut donc définir dans l'espace des profils où est situé le nuage une formule de distance mettant en jeu toutes les dimensions de l'espace : c'est l'objet des §§ 3.1 et 3.2.

Les diverses moyennes définies au § 3.0 dans le cas de la droite se généralisent alors à un nuage multidimensionnel (§ 3.3).

Pour rendre tangibles les relations de proximité que tisse l'espace au moyen de sa métrique entre les divers points du nuage on a recours à la projection sur des droites ou des plans (§ 3.4). Ainsi fait-on pour représenter un objet tridimensionnel ordinaire par diverses vues ou projections horizontales, frontales, etc. bien placées par rapport à l'objet. Le choix des directions privilégiées où un nuage multidimensionnel se projette le plus fidèlement possible sera traité au § 4.

3.0 Dispersion sur une droite : Un ensemble de nombres x^i munis respectivement de poids f_i : $\{(x^i, f_i) | i \in I\}$ peut être considéré comme un nuage de points dont les abscisses sont les x^i et les poids les f_i .

On écrira : $N(I) = \{(x^i, f_i) | i \in I\}$. Pour ce nuage, on définit classiquement :

la masse totale : $m_{\text{tot}} = \sum \{f_i | i \in I\}$

la moyenne, ou centre de gravité du nuage :

$$\bar{x} = \sum \{f_i x^i | i \in I\} / m_{\text{tot}} ;$$

le rayon moyen, ou moyenne des $|x^i - \bar{x}|$:

$$\text{ray} = \sum \{f_i |x^i - \bar{x}| | i \in I\} / m_{\text{tot}} ;$$

l'inertie par rapport à un point d'abscisse c de la droite :

$$\text{inert}_c = \sum \{f_i (x^i - c)^2 | i \in I\} ;$$

en particulier, l'inertie par rapport au centre $\bar{x}$:

$$\text{inert}_{\bar{x}} = \sum \{f_i (x^i - \bar{x})^2 | i \in I\} ;$$

la variance, ou moyenne des $(x^i - \bar{x})^2$:

$$\text{var} = \sum \{f_i (x^i - \bar{x})^2 | i \in I\} / m_{\text{tot}} ;$$

et l'écart-type, ou racine carrée de la variance :

$$\sigma = (\text{var})^{1/2} = \text{écart-type}.$$

Remarque : Dans le cas où $m_{\text{tot}} = 1$ (il en est ainsi en particulier quand $N(I)$ est défini par un tableau de correspondances : cf. § 1.2 *in fine*, remarque) inertie par rapport au centre de gravité et variance coïncident.

Pour mesurer la dispersion des x^i autour du centre $\bar{x}$, la moyenne des $|x^i - \bar{x}|$: *ray* s'impose d'abord à l'esprit. Pourtant, c'est l'*inertie* que l'on utilise, à cause de sa souplesse d'emploi dans les calculs. En effet, c'est le carré de la distance, non la distance, qui se décompose sur des axes orthogonaux, dans un espace euclidien, par le théorème de Pythagore. Et c'est à l'inertie que s'applique le théorème de Huygens : quel que soit le nombre réel c :

$$\text{inert}_c = \text{inert}_{\bar{x}} + m_{\text{tot}} (c - \bar{x})^2 ;$$

(l'inertie par rapport à un point c quelconque de la droite est égal à l'inertie par rapport au centre de gravité $\bar{x}$ augmentée de l'inertie par rapport à c de la masse totale concentrée au centre de gravité $\bar{x}$).

Ce choix de l'inertie comme mesure de la dispersion rend les résultats très sensibles aux effets de grandes distances (celles-ci intervenant élevées au carré) : deux masses égales situées à des distances au centre de rapport 10 fournissent des inerties de rapport 100.

Signalons enfin des inégalités que nous ne démontrons pas ici (cf. Pratique de l'A. des D. t. 1 §§ 3.2.3 et 10.3.1 Note) mais que nous illustrerons par un exemple. Notons *Max* et *min* les valeurs maxima et minima de $|x^i - \bar{x}|$ pour $i \in I$; on a :

$$\text{min} \leq \text{ray} \leq \text{var}^{1/2} \leq \text{Max}.$$

Ces inégalités expriment que, en bref, *ray* et $\text{var}^{1/2}$ donnent l'ordre de grandeur et la distance au centre de gravité ; mais $\text{var}^{1/2}$ est

une estimation plus sensible à la valeur de Max comme l'atteste l'exemple ci-dessous :

Exemple : Soit sur un axe ox les points d'abscisses -10 ; 0 ; $+10$, avec pour masses $0,005$; $0,99$; $0,005$ respectivement.

$$\bar{x} = (0,005 \times (-10)) + (0,99 \times 0) + (0,005 \times 10) = 0 ;$$

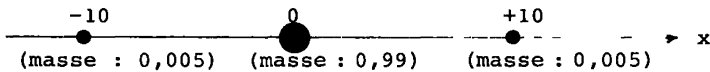
$$\min = 0 ; \quad \text{Max} = 10 ;$$

$$\text{ray} = (0,005 \times 10) + (0,99 \times 0) + (0,005 \times 10) = 0,1 ;$$

$$\text{var} = (0,005 \times 100) + (0,99 \times 0) + (0,005 \times 100) = 1 ;$$

$$\text{var}^{1/2} = 1 = \text{écart-type } (\sigma) ;$$

$$(\min = 0) < (\text{ray} = 0,1) < (\text{var}^{1/2} = 1) < (\text{Max} = 10) .$$

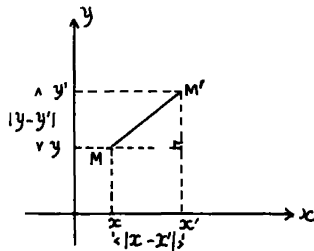


3.1 Formule de distance euclidienne dans un espace multidimensionnel

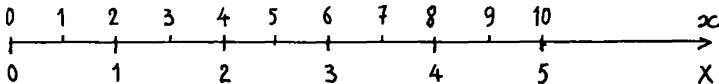
Comme nous l'avons annoncé au § 3.0, nous nous placerons dans le cadre de la géométrie euclidienne qui offre de remarquables facilités dans les calculs d'inertie.

Dans le plan de la géométrie euclidienne rapporté à deux axes rectangulaires gradués selon l'unité de distance du plan, le carré de la distance de deux points M , M' de coordonnées (x,y) et (x',y') s'exprime, en vertu du théorème de Pythagore, par la formule :

$$(1) \quad d^2(M,M') = (x' - x)^2 + (y' - y)^2$$



Changeons l'unité de graduation de l'un des axes, par exemple celui des x . Prenons comme nouvelle unité le double de l'ancienne : un point d'ancienne abscisse x a pour nouvelle abscisse $X = (1/2)x$; d'où $x = 2X$.



Les points M et M' d'anciennes coordonnées (x,y) et (x',y') ont pour nouvelles coordonnées (X,y) et (X',y') avec : $X = (1/2)x$ et $X' = (1/2)x'$; et la formule de distance s'écrit dans les nouvelles coordonnées :

$$(2) \quad d^2(M,M') = 4(X' - X)^2 + (y' - y)^2 ;$$

un coefficient 4 s'introduit dans le terme en $(X' - X)^2$ du fait de la graduation particulière choisie sur l'axe des X. Dans l'espace à 8 dimensions rapporté aux axes $\mathcal{E}_{BL}$, $\mathcal{E}_{US}$, $\mathcal{E}_{JR}$ etc., nous posons *a priori* la formule de distance : Pour tout couple (i, i') de pays importateurs, de profils sur J : f_J^i et $f_J^{i'}$,

$$d^2(f_J^i, f_J^{i'}) = \alpha_1 (f_{\mathcal{E}_{BL}}^{i'} - f_{\mathcal{E}_{BL}}^i)^2 + \alpha_2 (f_{\mathcal{E}_{US}}^{i'} - f_{\mathcal{E}_{US}}^i)^2 + \text{etc.}$$

$\uparrow$ coordonnée sur $\mathcal{E}_{BL}$ du profil de i
 $\uparrow$ coordonnée sur $\mathcal{E}_{BL}$ du profil de i'
 $\uparrow$ coefficient à préciser

L'absence de terme produits tels que $\beta_{12} (f_{\mathcal{E}_{BL}}^{i'} - f_{\mathcal{E}_{BL}}^i) (f_{\mathcal{E}_{US}}^{i'} - f_{\mathcal{E}_{US}}^i)$ implique l'orthogonalité des axes deux à deux. quant aux coefficients α_1, α_2 etc. α_8 il sont déterminés à partir du tableau $I \times J$ de façon à ce que soit satisfait le principe d'équivalence distributionnelle objet du § 3.2 suivant.

3.2 Distance distributionnelle

Nous avons envisagé au § 2.4 le cas où toutes les lignes du tableau $k_{I,J}$ étaient proportionnelles entre elles : le nuage $N(I)$ était alors concentré en un seul point coïncidant avec son centre de gravité.

Supposons ici que deux lignes seulement i_1 et i_2 soient proportionnelles entre elles ; les autres lignes étant quelconques :

$$\forall j \in J, k(i_1, j) = C k(i_2, j)$$

où C est une constante indépendante de j.

On a, pour les totaux $k(i_1)$ et $k(i_2)$ des lignes i_1 et i_2 du tableau $k_{I,J}$:

$$k(i_1) = C k(i_2) ;$$

d'où l'identité des profils de i_1 et i_2 :

$$\forall j \in J, k(i_1, j) / k(i_1) = k(i_2, j) / k(i_2).$$

Dans l'espace R_J , les deux points représentant les profils de i_1 et i_2 coïncident, chacun ayant son poids propre :

$$f_{i_1} = k(i_1)/k = C k(i_2)/k \text{ et } f_{i_2} = k(i_2)/k.$$

Constituons une ligne i_3 cumul des lignes i_1 et i_2 : son profil :

$$\{((k(i_2, j))(1+C))/((k(i_2))(1+C)) | j \in J\} = \{k(i_2, j)/k(i_2) | j \in J\}$$

coïncide avec celui de i_1 et i_2 ; et son poids :

$$f_{i_3} = k(i_2)(1+C)/k = f_{i_2} + f_{i_1}$$

est la somme des poids de i_1 et de i_2 .

En sorte que si l'on remplace les lignes i_1 et i_2 par l'unique ligne i_3 (cumul de i_1 et de i_2), on remplace dans le nuage $N(I)$ deux points confondus par un point ayant même position qu'eux et muni de la somme de leurs poids.

Mais ce changement remplace l'ensemble I par un ensemble

$I' = I - i_1 - i_2 + i_3$ et l'espace où est représenté $N(J)$ n'est plus R_I mais $R_{I'}$.

Le principe d'équivalence distributionnelle assure que pour un choix convenable de la formule de distance dans les espaces R_I et $R_{I'}$, c'est-à-dire en se restreignant aux formules du type :

$$(d(j, j'))^2 = \sum \{\alpha_i (f_{i1}^j - f_{i1}^{j'})^2 \mid i \in I\}$$

$$(d'(j, j'))^2 = \sum \{\alpha'_i (f_{i1}^j - f_{i1}^{j'})^2 \mid i \in I'\}$$

semblables à celles considérées au § 3.1 dans R_J) pour un choix convenable des α_i et α'_i , les distances distributionnelles des f_{i1}^j , au sein de $R_{I'}$, sont les mêmes que celles des f_{i1}^j au sein de R_I . Ce principe peut être énoncé en termes moins techniques : On ne change pas les résultats de l'analyse en cumulant dans le tableau de correspondance $k_{I,J}$ deux lignes (ou deux colonnes) proportionnelles entre elles. Nous avons ajouté dans le 2-ème énoncé : (ou deux colonnes) car les ensembles I et J sont traités de façon symétrique, comme nous l'avons signalé dès le § 1.1.

Il se trouve qu'en prenant dans R_I : $\alpha_i = 1/f_{i1}$ (inverse de la composante sur 1 du profil de la colonne de marge (cf. § 1.3) et dans R_J : $\alpha_j = 1/f_{j1}$ (inverse de la composante sur j du profil de la ligne de marge cf. § 1.3), on obtient des formules compatibles avec le principe d'équivalence distributionnelle. La distance ainsi définie dans R_I et dans R_J s'appelle *distance distributionnelle* ou *distance du chi2*: $\forall i, i'$ éléments de I , de profils sur J : f_{i1}^j et $f_{i1}^{j'}$,

$$d^2(f_{i1}^j, f_{i1}^{j'}) = \sum \{(1/f_{j1}) (f_{i1}^j - f_{i1}^{j'})^2 \mid j \in J\};$$

et $\forall j, j'$ éléments de J , de profils sur I : f_{i1}^j et $f_{i1}^{j'}$,

$$d^2(f_{i1}^j, f_{i1}^{j'}) = \sum \{(1/f_{i1}) (f_{i1}^j - f_{i1}^{j'})^2 \mid i \in I\}.$$

Toute autre formule euclidienne sans termes rectangles (i.e. du type considéré au § 3.1), compatible avec le principe d'équivalence distributionnelle se déduit de la formule du chi2 écrite ci-dessus par multiplication par une constante $\mathcal{L}$: $d' = \mathcal{L}d$.

L'analyse des correspondances utilise la formule du chi2 pour définir la distance de deux points dans R_I ou dans R_J . Ainsi l'on a, dans l'espace à 8 dimensions où est représenté $N(I)$ (cf. ligne de marge du tableau f_{I1}^J § 1.3) :

$$\alpha_1 = 1/.032; \alpha_2 = 1/.355; \alpha_3 = 1/.027; \alpha_4 = 1/.337$$

$$\alpha_5 = 1/.037; \alpha_6 = 1/.061; \alpha_7 = 1/.042; \alpha_8 = 1/.108$$

et dans l'espace à 14 dimensions où est représenté $N(J)$ (cf. col. de marge du tableau f_{I1}^J § 1.3) :

$$\alpha'_1 = 1/.060; \alpha'_2 = 1/.086 \text{ etc. } \alpha'_{14} = 1/.078.$$

Remarque : Le coefficient $\alpha_j = 1/f_j$ est d'autant plus fort que l'exportateur j a moins d'importance : ceci n'est pas absurde, car le terme de rang j : $(1/f_j)(f_j^i - f_j^{i'})^2$ est finalement de l'ordre de f_j comme l'est, généralement, la différence élevée au carré ; la forte valeur de α_j ne compense ainsi qu'en partie la petitesse des parts f_j^i , $f_j^{i'}$, $(f_j^i - f_j^{i'})$ afférentes à l'exportateur j . (En bref on peut écrire : $(1/\epsilon)\epsilon^2 \approx \epsilon$).

A titre d'exemple, calculons la distance entre le centre de gravité des trois lignes $\mathcal{JDL}$, $\mathcal{JPL}$, $\mathcal{JRM}$ et le centre de gravité du nuage $N(I)$ (cf. § 2.2).

	ΣBL	ΣUS	ΣJR	ΣMR	ΣSN	ΣTC	ΣTU	ΣCC
$f_J^{\mathcal{J}DPR}$	.063	.280	.036	.301	.014	.042	.041	.222
f_J	.032	.355	.027	.337	.037	.061	.042	.108
Δ	.031	.075	.009	.036	.023	.019	.001	.114
$10^3 \Delta^2$	.961	5.625	.081	1.296	.529	.361	.001	12.996
α	1/.032	1/.355	1/.027	1/.337	1/.037	1/.061	1/.042	1/.108
$\alpha \Delta^2$	.030	.016	.003	.004	.014	.006	.	.120

$f_J^{\mathcal{J}DPR}$: profil sur J du centre de gravité de la classe $\{\mathcal{JDL}, \mathcal{JPL}, \mathcal{JRM}\}$;

f_J : profil sur J du centre de gravité du nuage $N(I)$;

Δ : valeur absolue de la différence des coordonnées de f_J et $f_J^{\mathcal{J}DPR}$;

$10^3 \Delta^2$: élevé au carré et multiplié par 1000, pour format.

α : coefficients de la formule de distance distributionnelle.

Le carré de la distance entre f_J et $f_J^{\mathcal{J}DPR}$ est la somme des termes $\alpha \Delta^2$:

$$.030 + .016 + .003 + .004 + .014 + .006 + . + .120 = .193$$

3.3 Dispersion d'un nuage multidimensionnel

Les diverses caractéristiques définies au § 3.0 pour un nuage de points sur la droite peuvent être définies pour $N(I)$ au sein de R_J où la distance du χ^2 entre deux points joue le rôle de la valeur absolue de la différence des abscisses sur la droite :

masse totale du nuage :

$$m_{\text{tot}} = \Sigma \{f_i \mid i \in I\}$$

(pour $N(I)$ défini à partir du tableau $k_{I,J}$, $m_{\text{tot}} = 1$; mais si l'on considère un sous-nuage $N(I')$ avec $I' \subset I$, alors $m_{\text{tot}} = \Sigma \{f_i \mid i \in I'\}$ n'est pas égal à 1).

rayon moyen, ou moyenne des $d(f_J^i, f_J)$, distances des divers points f_J^i du nuage au centre de gravité f_J :

$$\text{ray} = \Sigma \{f_i d(f_J^i, f_J) \mid i \in I\} / m_{\text{tot}}$$

inertie par rapport à un point quelconque π_J de R_J :

$$\text{inert } \pi_J = \sum \{f_i d^2(f_J^i, \pi_J) | i \in I\}$$

inertie par rapport au centre de gravité (ou in. totale) :

$$\text{inert } f_J = \sum \{f_i d^2(f_J^i, f_J) | i \in I\}$$

variance, ou moyenne des $d^2(f_J^i, f_J)$:

$$\text{var} = \sum \{f_i d^2(f_J^i, f_J) | i \in I\} / m_{\text{tot}}$$

écart-type ou racine carrée de la variance :

$$\sigma = (\text{var})^{1/2}.$$

La remarque du § 3.0 peut être reprise également : le théorème de Pythagore sera précisé au § 3.4 à propos des projections ; quant au théorème de Huygens il s'énonce : pour tout point c_J de R_J ,

$$\text{inert } c_J = \text{inert } f_J + m_{\text{tot}} d^2(f_J, c_J) ;$$

l'inertie du nuage par rapport à un point quelconque c_J de R_J est égale à l'inertie du nuage par rapport à son centre de gravité augmentée de l'inertie par rapport à c_J de la masse totale du nuage concentrée en son centre de gravité).

Une inertie étant une quantité positive, il découle de la formule ci-dessus que l'inertie par rapport à un point c_J quelconque est supérieure à l'inertie par rapport au c. de g. du nuage, de la quantité $\text{inert } c_J(f_J, m_{\text{tot}})$ qui est positive, nulle seulement si c_J coïncide avec f_J . Autrement dit, c'est par rapport à son centre de gravité que le nuage $N(I)$ a l'inertie la plus faible.

Enfin, si l'on note min et Max les valeurs minima et maxima de $d(f_J^i, f_J)$ sur l'ensemble du nuage, on a les inégalités :

$$\min \leq \text{ray} \leq \text{var}^{1/2} \leq \text{Max}$$

(ray et $\text{var}^{1/2}$ donnent un ordre de grandeur de la distance au centre de gravité).

Remarque Les définitions données ci-dessus pour le nuage $N(I)$ de l'espace R_J se transposent pour le nuage $N(J)$ de l'espace R_I . Un fait très remarquable (qu'on ne démontrera pas) est que les deux nuages $N(I)$ et $N(J)$ ont même inertie totale. Ce fait confirme la symétrie existant entre I et J en analyse des correspondances.

3.4 Projection d'un nuage multidimensionnel sur un axe

Grâce à la distance définie au § 3.2, la dispersion du nuage au sein de l'espace multidimensionnel R_J prend un sens. Mais nous ne la voyons que par le truchement de projections sur des axes ajustés au nuage : les axes principaux d'inertie, qui seront définis au § 4. Ces axes, nous le verrons, passent par le centre de gravité du nuage. L'opération de base à considérer est donc la projection du nuage $N(I)$ sur une droite de R_J passant par f_J .

Quest-ce qu'une droite dans l'espace à 8 dimensions ? En géométrie élémentaire une droite est définie par deux points. De même, en dimension 8, une droite passant par f_J est définie par le point f_J et un second point a_J :

$$f_J = (f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8)$$

(nous désignons par 1, 2... les coordonnées $\geq BL$, $\leq US$ etc.) ;

$$a_J = (a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8) ;$$

un point quelconque c_J de R_J de coordonnées :

$$c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8$$

appartient à la droite (f_J, a_J) si et seulement si :

$$(c_1 - f_1)/(a_1 - f_1) = (c_2 - f_2)/(a_2 - f_2) = \dots = (c_8 - f_8)/(a_8 - f_8).$$

La projection orthogonale d'un point f_J^i du nuage $N(I)$ sur la droite (f_J, a_J) est le point de cette droite le plus proche de f_J^i : notons-le : $\text{pr} f_J^i$. En outre, le point f_J^i étant muni de son poids f_i , sa projection sur la droite (f_J, a_J) est munie du même poids f_i . On définit alors le nuage projeté sur la droite (f_J, a_J) comme l'ensemble des projections de ses divers points pesants :

$$\text{pr} N(I) = \{(\text{pr} f_J^i, f_i) | i \in I\}.$$

Le centre de gravité du nuage projeté n'est autre que f_J , et les caractéristiques de dispersion du nuage projeté autour de son centre f_J se ramènent à celles définies au § 3.0 en choisissant sur la droite une origine 0 et un sens positif, et en posant pour l'abscisse x d'un point c de la droite :

$$|x| = d(0, c) ;$$

(le choix de l'orientation de la droite, qui définit le signe de x , étant indifférent).

4 Axes principaux d'inertie ; facteurs

Ainsi qu'on l'a annoncé au § 3, l'objet de l'analyse factorielle est de déterminer dans l'espace ambiant à un nuage un système d'axes relativement auquel la forme de celui-ci apparaît au mieux.

Au § 4.0 on rappelle les propriétés d'un système d'axes orthonormé. Aux §§ 4.1 et 4.2 on définit de façon précise le système des axes factoriels pour le nuage $N(I)$. Une définition identique vaut pour le nuage $N(J)$ (§ 4.3). L'analyse factorielle des correspondances permet d'identifier les axes factoriels des deux nuages (§ 4.4). Réciproquement, la connaissance des résultats d'analyse factorielle pour les deux nuages permet de reconstituer le tableau $k_{I,J}$ (§ 4.5).

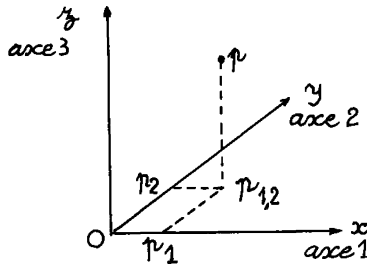
4.0 Système d'axes orthonormé : décomposition de l'inertie

Dans R_J comme dans l'espace usuel à 3 dimensions, un système d'axes orthonormés est un système d'axes deux à deux perpendiculaires et munis chacun, comme unité de graduation, de l'unité de distance dans l'espace. Cette définition est équivalente à la suivante : un système

d'axes est orthonormé si et seulement si l'espace étant rapporté à ce système, l'expression de la distance de deux points en fonction des coordonnées est une formule euclidienne sans coefficients. Par exemple, en dimension 3, soit $p = (x(p), y(p), z(p))$ et $p' = (x(p'), y(p'), z(p'))$; le repère est orthonormé si et seulement si l'on a :

$$d^2(p, p') = (x(p) - x(p'))^2 + (y(p) - y(p'))^2 + (z(p) - z(p'))^2.$$

Aux §§ 3.1 et 3.2 nous avons considéré des repères orthogonaux (ortho = droit) mais non "normés" : les formules de distance comportaient des coefficients.



Numérotons : 1, 2, 3 les axes des coordonnées x, y, z dans l'espace usuel à 3 dimensions. L'axe 3 est perpendiculaire aux axes 1 et 2; donc au plan des axes 1, 2. Et de même, par permutation circulaire, l'axe 1 est perpendiculaire aux axes 2 et 3 donc au plan des axes 2, 3 etc. .

La perpendiculaire abaissée de p sur le plan (1,2) est parallèle à l'axe 3 et coupe le plan (1,2) au point noté $p_{1,2}$, projection orthogonale de p sur le plan (1,2).

Les perpendiculaires abaissées de $p_{1,2}$ dans le plan (1,2) sur les axes 1 et 2 sont respectivement parallèles aux axes 2 et 1 et coupent : l'une l'axe 1 en p_1 et l'autre l'axe 2 en p_2 .

Le plan $p, p_{1,2}, p_1$ est perpendiculaire à l'axe 1 (comme contenant deux droites $(p, p_{1,2})$ et $(p_{1,2}, p_1)$ perpendiculaires à cet axe) et coupe l'axe 1 en p_1 , projection orthogonale de p et de $p_{1,2}$ sur l'axe 1.

Les projections de p sur les autres axes et plans de coordonnées s'obtiennent de la même façon : $O, p_1, p_2, p_{1,2}, p_3, p_{1,3}, p, p_{2,3}$ sont les sommets d'un parallélépipède rectangle dont les arêtes sont parallèles aux axes de coordonnées.

On a sur l'axe 1 : $\overline{Op_1} = x(p)$;

sur l'axe 2 : $\overline{Op_2} = y(p)$;

sur l'axe 3 : $\overline{Op_3} = z(p)$.

La projection sur l'axe 1 du point (x, y, z) est le point $(x, 0, 0)$; sur l'axe 2 : $(0, y, 0)$; sur l'axe 3 : $(0, 0, z)$; sur le plan (1,2) : $(x, y, 0)$; sur le plan 2,3 : $(0, y, z)$; sur le plan 1,3 $(x, 0, z)$.

Propriété caractéristique de la projection : p_1 est le point de l'axe 1 le plus proche de p ; $p_{1,2}$ est le point du plan 1,2 le plus proche de p :

$$d^2(p, p_1) = y^2 + z^2$$

et pour un point p' de l'axe 1 : $p' = (x', 0, 0)$

$$d^2(p, p') = (x - x')^2 + y^2 + z^2$$

$(x - x')^2$ est une quantité positive, nulle seulement si $x = x'$.

On voit de la même façon que $d^2(p, p_{1,2})$ est inférieur à $d^2(p, p')$ pour tout point p' du plan 1,2.

La généralisation à un nombre quelconque de dimensions est immédiate :

soit dans l'espace à 4 dimensions $p = (x, y, z, t)$;
la projection sur l'axe 1 est $p_1 = (x, 0, 0, 0)$;
etc.

et le repère est orthonormé si et seulement si :

$$d^2(p, p') = (x(p) - x(p'))^2 + (y(p) - y(p'))^2 + (z(p) - z(p'))^2 + (t(p) - t(p'))^2 ;$$

etc.

La propriété caractéristique de la projection d'un point sur un axe ou sur un plan d'être le point de l'axe ou du plan le plus proche du point que l'on projette reste vraie en dimension quelconque : nous l'avons utilisée au § 3.4 pour définir la projection d'un point de $N(I)$ sur une droite de R_J .

Considérons, dans le sous-espace de R_J support du simplexe des profils sur J P_J , un système d'axes orthonormé centré en f_J centre de gravité du nuage $N(I)$. Notons 1, 2, 3, ..., 7 les axes, (en nombre 7 car le support de P_J est de dimension 7), de ce système et $x_1, x_2, \dots, x_7$ les coordonnées sur ces axes. Un point f_J^i du nuage a , dans ce système, 7 coordonnées que nous notons, $x_1(i), x_2(i), \dots, x_7(i)$. On a (en notant pr_1 la projection sur l'axe 1)

$$|x_1(i)| = d(f_J, pr_1(f_J^i)) ; |x_2(i)| = d(f_J, pr_2(f_J^i)) \text{ etc. ;}$$

$$d^2(f_J, f_J^i) = (x_1(i))^2 + (x_2(i))^2 + \dots + (x_7(i))^2$$

D'où la décomposition de l'inertie de (f_J^i, f_i) par rapport au centre de gravité f_J sur le système d'axes :

$$f_i d^2(f_J, f_J^i) = f_i (x_1(i))^2 + f_i (x_2(i))^2 + \dots + f_i (x_7(i))^2 ;$$

↑
inert. f_J de $pr_1(f_J^i)$

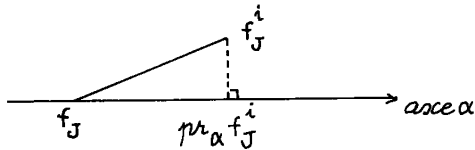
et en faisant la somme des $f_i d^2(f_J, f_J^i)$ pour tous les points du nuage, on obtient la décomposition de l'inertie totale du nuage par rapport à son centre f_J en somme des inerties du nuage projeté sur chacun des 7 axes.

$$\text{inert}_{f_J}(N(I)) = \text{inert}_{f_J}(\text{pr}_1(N(I))) + \dots \\ + \text{inert}_{f_J}(\text{pr}_7(N(I))) .$$

Remarque : De la décomposition ci-dessus il découle l'inégalité valant pour tout axe α du système :

$$\text{inert}_{f_J}(\text{pr}_\alpha(N(I))) \leq \text{inert}_{f_J}(N(I)) ;$$

inégalité qui découle également de la situation géométrique des segments (f_J, f_J^i) et $(f_J, \text{pr}_\alpha(f_J^i))$ par rapport à la droite $(f_J^i, \text{pr}_\alpha(f_J^i))$: le premier est oblique, le second perpendiculaire.



La décomposition de l'inertie sur des axes orthogonaux est celle que nous avons en vue au § 3 quand nous choisissons dans l'espace une métrique euclidienne et que nous prenons pour mesure de la dispersion autour du centre f_J l'inertie du nuage par rapport à ce centre.

Pareille décomposition s'applique e.g. à la projection de f_J^i sur le plan des axes 1,2 : $\text{pr}_{1,2}(f_J^i)$, et aux axes 1 et 2. On a :

$$\text{pr}_1(f_J^i) = \text{pr}_1(\text{pr}_{1,2}(f_J^i))$$

$$\text{pr}_2(f_J^i) = \text{pr}_2(\text{pr}_{1,2}(f_J^i)) ;$$

d'où l'expression de la distance à f_J dans le plan (1,2) :

$$d^2(f_J, \text{pr}_{1,2}(f_J^i)) = (x_1(i))^2 + (x_2(i))^2$$

ce qui entraîne l'inertie du nuage projeté sur le plan (1,2) :

$$\text{inert}_{f_J} \text{pr}_{1,2}(N(I)) = \text{inert}_{f_J} \text{pr}_1(N(I)) \\ + \text{inert}_{f_J} \text{pr}_2(N(I)) .$$

4.1 Premier axe du nuage $N(I)$; λ_1 , τ_1 , $F_1(i)$

Dans l'espace R_J , parmi toutes les droites passant par le centre de gravité f_J de $N(I)$, le premier axe principal d'inertie est celui sur lequel $N(I)$ se projette avec le maximum de dispersion. (Si, par exception ce maximum est réalisé pour plusieurs droites, on choisit l'une de celles-ci). On appelle première valeur propre, et on note λ_1 , l'inertie par rapport à f_J , du nuage projeté sur le 1-er axe principal d'inertie. (Rappelons, cf. § 3.4, que f_J , qui coïncide avec sa projection sur le 1-er axe puisque ce dernier passe par f_J , est centre de gravité du nuage projeté).

En analyse des correspondances, cette première valeur propre est toujours comprise entre 0 et 1.

On appelle *taux d'inertie* relatif à la première valeur propre le rapport $\tau_1 = \lambda_1 / \text{inertie totale de } N(I)$, exprimé en % ; c'est-à-dire l'inertie par rapport à f_J du nuage projeté divisée par l'inertie par rapport à f_J de $N(I)$. Ce rapport τ_1 est toujours inférieur ou égal à 1. (cf. § 4.0 Remarque).

L'origine de l'axe est fixée en f_J ; le sens positif est choisi arbitrairement ; l'unité de graduation est l'unité de distance dans l'espace R_J muni de la distance distributionnelle. Tout point de l'axe peut alors être repéré par son *abscisse* ; en particulier, pour un point f_J^1 du nuage $N(I)$, la projection $pr_1(f_J^1)$ sur le premier axe principal a une abscisse notée $F1(i)$ et appelée *premier facteur* de i . Le point f_J , centre de gravité du nuage et origine de l'axe a pour abscisse : 0.

4.2 Définition des axes successifs

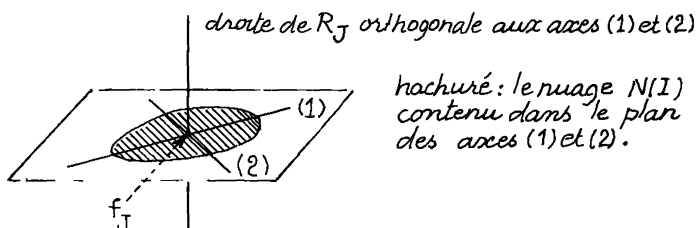
Parmi toutes les droites passant par f_J et orthogonales au 1-er axe principal d'inertie, le *deuxième axe* principal d'inertie est celui sur lequel $N(I)$ se projette avec le maximum de dispersion. On appelle *deuxième valeur propre*, et on note λ_2 , l'inertie par rapport à f_J du nuage projeté sur ce 2-ème axe.

$$0 \leq \lambda_2 \leq \lambda_1 \leq 1.$$

Parmi toutes les droites passant par f_J et orthogonales aux deux premiers axes principaux d'inertie, le 3-ème axe principal d'inertie est celui sur lequel $N(I)$ se projette avec le maximum de dispersion. On appelle 3-ème valeur propre, et on note λ_3 , l'inertie du nuage projeté sur le 3-ème axe, par rapport à f_J :

$$0 \leq \lambda_3 \leq \lambda_2 \leq \lambda_1 \leq 1$$

etc. jusqu'à ce qu'on ait épuisé l'épaisseur du nuage. Comme $N(I)$ est compris dans un espace de dimension 7 (cf. § 2.1.2), il existe au plus 7 axes principaux d'inertie. Nous verrons au § 5 en étudiant les listages de sortie de l'ordinateur qu'il existe effectivement 7 axes principaux d'inertie correspondant chacun à une valeur propre non nulle. Si l'on cherche à projeter $N(I)$ sur une droite passant par f_J et orthogonale dans R_J aux 7 premiers axes principaux d'inertie, le nuage projeté se réduira à f_J (muni de m_{tot}) et la dispersion sera nulle. C'est ce qu'illustre le dessin que nous proposons en dimension 3.



Il se peut, pour un tableau $k_{I,J}$ présentant certaines particularités, que le nuage $N(I)$ soit compris dans un espace de dimension n strictement inférieure à $(Card J - 1)$. Dans ce cas, les valeurs propres de rang supérieur à n sont nulles : le nuage projeté sur une droite quelconque passant par f_J et orthogonale aux n premiers axes principaux d'inertie se réduit à f_J (muni de m_{tot}).

Chacun des axes est centré en f_J ; muni d'un sens positif arbitraire et d'une unité de graduation coïncidant avec l'unité de distance dans l'espace R_J , comme il a été dit pour l'axe 1.

Ainsi l'espace support du nuage, sous-espace de R_J , est muni d'un système d'axes orthonormés (cf. § 4.0), centré au centre de gravité du nuage.

Relativement à ce système, un point f_J^i du nuage $N(I)$ a 7 coordonnées, appelées facteurs de i :

$$F_1(i), F_2(i), \dots, F_7(i) ;$$

(le centre de gravité F_J a ses 7 coordonnées nulles).

Ce que nous avons dit au § 4.0 s'applique au système des axes principaux d'inertie : en projection sur un axe :

$$d(f_J, pr_\alpha(f_J^i)) = |F_\alpha(i)| ;$$

$$\lambda_\alpha = \sum \{f_i(F_\alpha(i))^2 | i \in I\}$$

(λ_α est la variance de F_α ; $\lambda_\alpha^{1/2}$ donne l'ordre de grandeur de F_α cf. § 3.0) ; l'inertie par rapport à f_J du nuage $N(I)$ s'écrit :

$$inert_{f_J}(N(I)) = \lambda_1 + \lambda_2 + \dots + \lambda_7 ;$$

cette somme est encore appelée *trace* (ce qui rappelle son calcul algébrique). Le *taux d'inertie* τ_α apparaît maintenant comme la part relative de la projection sur l'axe α dans l'inertie totale du nuage $N(I)$:

$$\tau_\alpha = \lambda_\alpha / inert_{f_J}(N(I)) = \lambda_\alpha / trace$$

$$= \lambda_\alpha / (\lambda_1 + \lambda_2 + \dots + \lambda_7).$$

Comme nous l'avons fait au § 4.0, on décompose sur deux axes e_1 et e_2 , l'inertie de la projection du nuage sur le plan de ces axes :

$$inert_{f_J} pr_{1,2} N(I) = \lambda_1 + \lambda_2 ,$$

ce qui représente une part relative de l'inertie totale égale à :

$$(\lambda_1 + \lambda_2) / trace = \tau_1 + \tau_2.$$

N. B. Nous ne dirons rien du procédé de calcul des directions des axes principaux d'inertie et des valeurs propres qui leur sont relatives.

4.3 Axes et facteurs pour le nuage $N(J)$: Rappelons (cf. § 2.3) que $N(J)$ est l'ensemble des 8 profils sur I des 8 pays exportateurs constituant J . Dans l'espace à 14 dimensions : $JBL, JCA, etc., JEE$, chaque profil f_I^j est muni de son poids : importance relative du pays j

dans l'ensemble des 8 pays exportateurs. On note f_I le centre de gravité du nuage $N(J)$.

On peut définir dans l'espace R_I où est situé $N(J)$, un système d'axes principaux d'inertie centré en f_I , centre de gravité du nuage, comme nous l'avons fait pour $N(I)$ aux §§ 4.1 et 4.2 ci-dessus, *mutatis mutandis*.

A priori, on s'attend à déterminer 13 axes, puisque le support du simplexe des profils sur I est de dimension 13. Or il se trouve que, nécessairement, les valeurs propres sont les mêmes pour $N(J)$ et pour $N(I)$: il y a donc, pour $N(J)$ comme pour $N(I)$, 7 valeurs propres non nulles en sorte que $N(J)$ admet, comme $N(I)$, un système de 7 axes principaux d'inertie ; on voit ainsi que $N(J)$ est contenu dans un sous-espace de R_I de dimension 7 (et non 13). En outre, on retrouve le fait que $N(J)$ et $N(I)$ ont même inertie totale : $\lambda_1 + \lambda_2 + \dots + \lambda_7$ (cf. *supra* § 3.3 Remarque).

On note $G_1, G_2, \dots, G_7$ les coordonnées sur le système orthonormé des axes principaux d'inertie de $N(J)$. On a :

$$d(f_I, \text{pr}_\alpha f_I^J) = |G_\alpha(j)|$$

(la distance au centre f_I de la projection sur l'axe α d'un point f_I^J du nuage $N(J)$ est égale à la valeur absolue de la coordonnée de ce point sur l'axe) ;

$$\lambda_\alpha = \sum \{f_j G_\alpha^2(j) \mid j \in J\}.$$

4.4 Identification des supports des deux nuages : formule de transition ; principe barycentrique : Comme nous l'avons vu au § 4.3 précédent, les axes principaux d'inertie des deux nuages se correspondent biunivoquement :

le 1-er axe de $N(I)$ correspond au 1-er axe de $N(J)$ avec la même valeur propre : λ_1 ;

le 2-ème axe de $N(I)$ au 2-ème axe de $N(J)$ avec la même v. pr. λ_2 ;
etc.

les 7-èmes axes se correspondent avec la v. pr. λ_7 .

D'où l'idée de superposer les deux sous-espaces supports des nuages en faisant coïncider f_J avec f_I ; le 1-er axe de $N(I)$ avec le 1-er axe de $N(J)$; le 2-ème axe de $N(I)$ avec le 2-ème axe de $N(J)$; ainsi de suite jusqu'aux 7-èmes axes.

Mais comme les axes sont orientés chacun arbitrairement, il y a deux façons de faire coïncider les axes de même rang. Au total, il y a 2^7 façons de faire coïncider les deux systèmes de 7 axes. Mais pour l'une des 2^7 superpositions possibles, les deux nuages satisfont au *principe barycentrique* que l'on peut énoncer ainsi : à une correction près (cf. *infra*) tout point f_I^i de $N(I)$ est situé au centre de gravité des points f_I^j de $N(J)$ affectés, comme coefficients de pondération, des composantes sur eux du profil de i , et de même, tout point f_I^j de $N(J)$ est situé au centre de gravité des points f_I^i de $N(I)$ affectés,

comme coefficients de pondération, des composantes sur eux du profil de j . De façon précise, on a, pour une orientation cohérente des axes des deux nuages les formules suivantes, dites de *transition* :

$$\forall \epsilon [1, 2, \dots, 7] :$$

$$F_{\alpha}(i) = (1/(\lambda\alpha))^{1/2} \sum \{f_j^1 G_{\alpha}(j) | i \in J\} ;$$

$$G_{\alpha}(j) = (1/(\lambda\alpha))^{1/2} \sum \{f_i^j F_{\alpha}(i) | i \in I\} .$$

(on notera la présence du coefficient $(1/\lambda\alpha)^{1/2}$: c'est pourquoi on a dit : "à une correction près").

Ici apparaît la parfaite symétrie qui existe entre les ensembles I et J dans l'analyse du tableau $k_{I,J}$. La conséquence pratique de ces formules est qu'il suffit de déterminer les axes principaux d'inertie et les facteurs relatifs à l'un seulement des nuages $N(I)$ ou $N(J)$; les facteurs relatifs à l'autre nuage se calculent par les formules de transition. Quant au principe barycentrique, il permet d'interpréter les groupes de points qui apparaissent sur les graphiques plans : la présence d'un pays exportateur j au sein d'un groupe c de pays importateurs est le signe que le profil de j a de fortes composantes sur les pays de c i.e. que les pays importateurs de c absorbent une part importante des exportations de j . Inversement, la présence d'un pays importateur i au sein d'un groupe s de pays exportateurs est le signe que les pays exportateurs de s fournissent une part importante des importations de i .

4.5 Formule de reconstitution et modèle d'indépendance

A partir d'un tableau $k_{I,J}$ on a obtenu une description des éléments i de I et j de J par des points de R_J et R_I rapportés respectivement aux axes principaux d'inertie de $N(I)$ et de $N(J)$, la dispersion de $N(I)$ et de $N(J)$ sur les axes successifs étant mesurée par la même suite de valeurs propres : $\lambda_1, \lambda_2, \dots, \lambda_n$ (dans l'exemple traité, $n=7$).

Inversement, connaissant les facteurs sur I et sur J , la suite des valeurs propres et, en outre, les marges du tableau $k_{I,J}$ i.e.

$\{k(i) | i \in I\}$ totaux des lignes et $\{k(j) | j \in J\}$ totaux des colonnes, on reconstitue le tableau $k_{I,J}$ grâce à la *formule de reconstitution* qui s'écrit :

$$k(i,j) = k f_i f_j (1 + (\lambda_1)^{-1/2} F_1(i) G_1(j) + (\lambda_2)^{-1/2} F_2(i) G_2(j) + \dots + (\lambda_n)^{-1/2} F_n(i) G_n(j)) ;$$

$$\text{avec: } k = \sum \{k(i) | i \in I\} = \sum \{k(j) | j \in J\} ;$$

$$f_i = k(i)/k ; \quad f_j = k(j)/k .$$

L'ordre de grandeur des termes successifs de cette formule diminue comme $(\lambda\alpha)^{1/2}$ (en effet, on a vu au § 4.2 que $(\lambda\alpha)^{1/2}$ donnait l'ordre de grandeur de $F_{\alpha}(i)$ et de $G_{\alpha}(j)$) ; ce qui permet de donner une reconstitution approchée du tableau $k_{I,J}$ en ne gardant que les premiers termes de la formule :

$$\text{approx. d'ordre } 0 : k_0(i,j) = k f_i f_j$$

$$\text{approx. d'ordre } 1 : k_1(i,j) = k f_i f_j (1 + (\lambda_1)^{-1/2} F_1(i) G_1(j))$$

etc.

L'approximation d'ordre 0 satisfait au modèle d'indépendance : on reconnaît la formule (2') du § 2.4. C'est la simplification extrême du tableau $k_{I,J}$ dont on ne conserve que les marges. Au fur et à mesure que l'on ajoute des termes, on obtient des approximations du tableau $k_{I,J}$ de plus en plus riches correspondant à des nuages présentant de plus en plus de dimensions. Il arrive qu'au-delà d'un certain rang p , les facteurs issus du tableau $k_{I,J}$ ne rendent compte que de fluctuations aléatoires. Dans ce cas, l'approximation d'ordre p du tableau $k_{I,J}$ est plus fidèle à la réalité que le tableau $k_{I,J}$ lui-même.

Au § 2.4 nous avons dit que l'interaction entre i et j était attractive ou répulsive selon que $k(i,j)$ était supérieur ou inférieur à $k f_i f_j$. La formule de reconstitution, en donnant du rapport $k(i,j)/(k f_i f_j)$ une expression analytique, permet de nuancer cette mesure globale de l'interaction : i et j s'attirent de par le facteur α si le terme $F_\alpha(i) G_\alpha(j)$ est positif (et tend par conséquent à rendre $k(i,j)$ supérieur à $k f_i f_j$), c'est-à-dire si les projections sur l'axe α des points représentant i et j sont situées du même côté de l'origine ; au contraire, i et j se repoussent de par l'axe α si $F_\alpha(i) G_\alpha(j)$ est négatif et tend à rendre $k(i,j)$ inférieur à $k f_i f_j$. Eventuellement, i et j s'attirent par certains axes, se repoussent par d'autres : il n'est pas possible d'avoir des attractions et des répulsions un schéma simpliste ; l'analyse factorielle donne justement de la complexité une image ordonnée.

5 Interprétation des résultats d'analyse : valeurs propres et graphiques.

Les résultats sont imprimés sur le listage qui comprend : l'histogramme des valeurs propres ; des tableaux donnant, avec les valeurs des facteurs, des informations complémentaires qui seront expliquées dans la suite ; des graphiques donnant chacun les projections des points sur un plan engendré par deux axes principaux d'inertie (appelés encore : axes factoriels). Dans le présent § 5 nous présenterons l'histogramme des valeurs propres et interpréterons l'analyse d'après les graphiques plans sans recourir aux tableaux numériques. Au § 6, nous expliquerons le contenu de ces tableaux qui aident à confirmer et préciser l'interprétation.

5.1 Histogramme des valeurs propres : Valeurs propres et taux d'inertie sont communs aux deux nuages $N(I)$ et $N(J)$ qui ont (cf. 4.3) même inertie totale et mêmes valeurs propres (ce qui permet d'identifier les axes principaux d'inertie de même rang des deux nuages).

Chaque ligne de ce tableau correspond à un axe principal d'inertie et on lit de haut en bas :

- dans la colonne VAL PROPRE : $\lambda_1 = .45$; $\lambda_2 = .32$; $\lambda_3 = .15$ etc. ; ces valeurs sont données avec un nombre de décimales qui dépasse la précision permise dans l'interprétation. Le total de la colonne n'est pas imprimé : il est égal à la trace, i.e. à l'inertie totale des nuages $N(I)$ ou $N(J)$.
- dans la colonne POURCENT : $\tau_1 = 40,5\%$; $\tau_2 = 28,4\%$; $\tau_3 = 13,7\%$ etc.. Rappelons que $\tau_\alpha = \lambda_\alpha / \text{trace}$ et mesure la part afférente à l'axe α dans l'inertie totale du nuage.

■ dans la colonne CUMUL : $\tau_1 = 40,5\%$; $\tau_1 + \tau_2 = 68,9\%$; $\tau_1 + \tau_2 + \tau_3 = 82,6\%$ etc. . La dernière ligne de cette colonne indique toujours 100%. Dans le cas présent il y a 7 axes et l'on a : $\tau_1 + \tau_2 + \dots + \tau_7 = 100\%$ car cette somme n'est autre que trace/trace = 1.

■ L'histogramme représente chaque valeur propre λ_α par un segment de longueur proportionnelle à λ_α ; on peut ainsi apprécier d'un regard la décroissance des valeurs propres quand leur rang augmente. Sur notre exemple, on voit que chacune des 3 premières valeurs propres est nettement séparée de celle qui la suit : λ_1 vaut près de 1,5 fois λ_2 ; $\lambda_2 \approx 2 \times \lambda_3$; $\lambda_3 \approx 2 \times \lambda_4$.

VALEUR PROPRE 1 : 1

NUM	ITER	VALEUR PROPRE	POURCENT	CUMUL	HISTOGRAMME DES VALEURS PROPRES DE LA MATRICE
1	1	.45050079	40,5	40,5	*****
2	1	.31613224	28,42	68,92	*****
3	1	.15236717	13,7	82,61	*****
4	4	.081875021	7,36	89,97	*****
5	2	.057501991	5,17	95,14	*****
6	1	.033471503	3,01	98,15	*****
7	3	.02056181	1,85	100	*****

Placé en tête du listage, ce tableau guide l'utilisateur dans le choix des plans à examiner et du nombre des facteurs susceptibles d'être clairement interprétés. Par exemple, $CUMUL3 = 1 + 2 + 3 = 82,6\%$ suggère que la représentation dans l'espace engendré par les axes 1, 2, 3 rend compte assez fidèlement de la dispersion totale des nuages. Comme d'autre part les valeurs propres sont bien séparées, les axes sont susceptibles d'être interprétés individuellement et peuvent être associés librement pour construire des graphiques plans. Dans la suite, nous considérerons les plans 1-2 et 2-3.

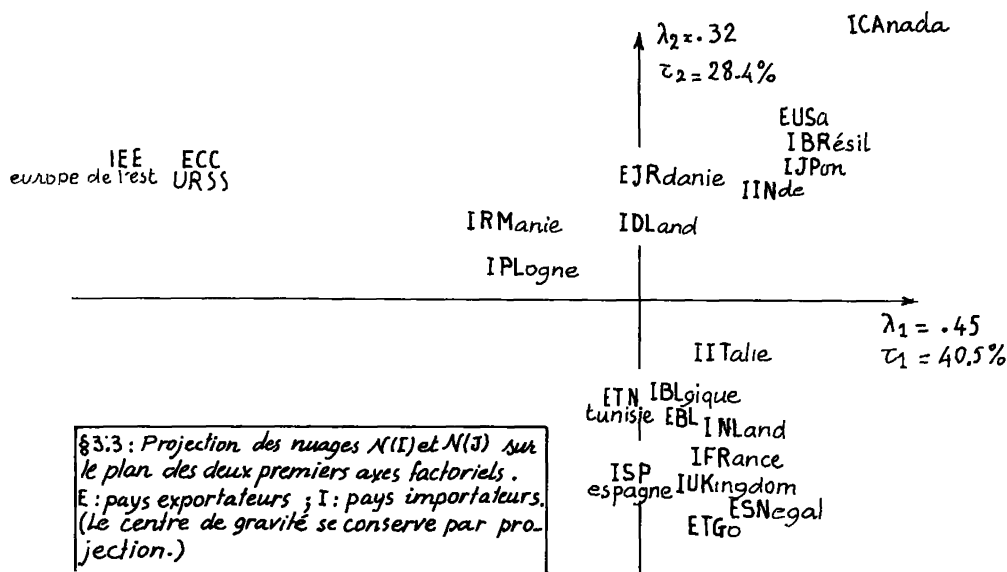
Parfois deux valeurs propres consécutives sont voisines (étant bien séparées de celle qui les précède immédiatement et de celle qui les suit) : c'est le cas dans notre exemple, de λ_4 et λ_5 . On peut craindre alors que la direction des deux axes correspondants dans le plan qu'ils définissent soit instable, c'est-à-dire dépende d'une modification contingente dans le choix des données ; par exemple de l'introduction d'un nouveau pays. L'interprétation devra donc être prudente.

Valeurs propres et pourcentages varient grandement selon les types de données et les cas traités. Des valeurs propres élevées signalent des oppositions tranchées dont l'interprétation est souvent à la fois évidente et attendue. Des valeurs propres faibles peuvent correspondre à des corrélations plus discrètes que seule l'analyse aura révélées. Il est manifeste que les pourcentages sont *a priori* d'autant plus élevés qu'il y a moins d'axes entre lesquels se répartisse l'inertie, c'est-à-dire moins de dimensions dans l'espace ; autrement dit que le plus petit des deux nombres (Card I-1) et (Card J-1) est plus faible (cf. § 2.1).

Nous répéterons donc que l'histogramme des valeurs propres propose d'abord d'utiles suggestions, mais que seul l'examen approfondi des graphiques et des tableaux numériques peut asseoir l'interprétation. Aucune règle arithmétique ne peut dispenser le savant de la réflexion critique seul garant de la validité de son oeuvre.

5.2 Examen des graphiques plans

5.2.1 Le plan 1.2



Dans ce plan apparaissent comme trois pôles bien distincts les trois pays exportateurs les plus "lourds": ΣCC : URSS (POIDS = 102, cf. § 1.3), ΣUSA (POIDS = 355) et $\Sigma MaRoc$ (POIDS = 337).

ΣCC : URSS et ΣEE : autres pays d'Europe de l'Est (qui se projette à proximité de ΣCC) forment un groupe très isolé du reste des points à l'extrémité de l'axe 1.

ΣUSA et $\Sigma MaRoc$, tous deux opposés à ΣCC : URSS sur l'axe 1, s'opposent entre eux sur l'axe 2.

Dans l'orbite de ΣUSA se projettent $\Sigma Canada$ (qui importe la quasi-totalité de ses phosphates des USA: cf. tableau des profils de lignes § 1.3), $\Sigma BRésil$, $\Sigma Japon$, $\Sigma Inde$ qui ont ΣUSA pour principal fournisseur de phosphates.

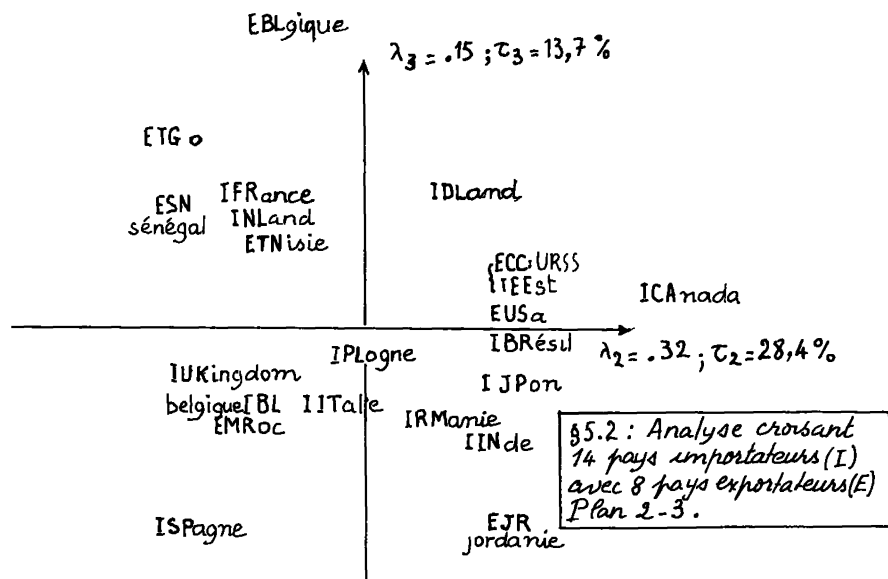
La situation de $\Sigma MaRoc$ est moins claire du fait de la présence à son voisinage des autres exportateurs africains et de $\Sigma Belgique$ qui, rappelons-le, joue le rôle d'intermédiaire. Dans ce groupe figurent la plupart des pays importateurs européens occidentaux.

La position de $\Sigma Jordanie$, $\Sigma Roumanie$, $\Sigma Pologne$ et $\Sigma Deutschland$, (centrale par rapport aux trois pôles ΣCC : URSS, ΣUSA , $\Sigma MaRoc$) reste à expliquer.

En conclusion, ce plan 1-2 fait apparaître l'isolement de ΣCC : URSS dans $N(J)$, et dans $N(I)$ l'isolement de $\Sigma Europe$ de l'Est suivi de loin par $\Sigma Pologne$ et $\Sigma Roumanie$, faits qui ont déterminé le premier axe principal d'inertie dans chacun des deux nuages. En sorte que la dispersion du reste du nuage $N(I)$ - ΣEE ou $N(J)$ - ΣCC se mesure sur les axes suivants: 2, 3 etc. L'axe 2 apparaît dans ce plan 1-2 comme opposant ΣUSA aux exportateurs africains (principalement $\Sigma MaRoc$ qui

est le plus important). Nous le retrouverons dans le plan 2-3 objet au § suivant.

5.2.2 Le plan 2.3



On distingue quatre groupes :

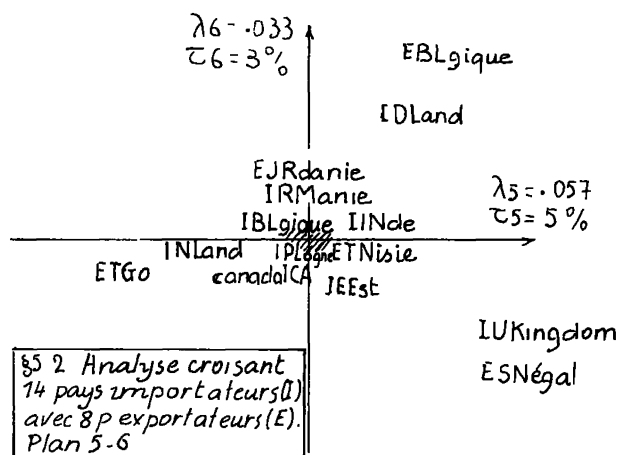
USA sur l'axe 2, entouré de CANada, BRésil, JJaPon ; (la position de l'Europe de l'Est et ECC : URSS dans ce secteur doit être nuancée par le fait qu'en relief par rapport au plan 2-3 ces points s'écartent sur l'axe 1, fait sur lequel nous reviendrons au § 6.3 en lisant les listages de facteurs).

MaRoc est isolé cette fois des autres exportateurs africains et entouré de l'Espagne, Belgique, United Kingdom, l'Italie. Le pays Pologne est situé dans le même quart de plan mais très près de l'origine des axes. Les principaux clients de MaRoc sont l'Espagne, la France et les Pays-Bas, mais ces deux derniers pays : FR et NL importent davantage des autres pays africains).

Sur le 3-ème axe les exportateurs africains (autres que MaRoc) ainsi que Belgique, celui-ci très éloigné sur le 3-ème axe positif, associés à l'France et l'NederLand s'opposent à MaRoc.

Enfin, la Jordanie : JR est isolée en tant qu'exportateur dans un quadrant où se projettent l'Inde, Roumanie et JJaPon ses trois premiers clients (cf. profils des colonnes § 1.3), JR s'oppose à MaRoc sur le 2-ème axe et aux autres pays africains sur le 3-ème axe.

La position intermédiaire de l'DeutschLand reste une question posée qui sera reprise à l'aide des listages de facteurs.

5.2.3 Le plan 5-6

Nous représentons le plan 5-6 afin de montrer comment se dispersent les nuages dans un plan perpendiculaire aux premiers axes factoriels (1, 2, 3, 4) : la plupart des points sont massés au centre et quelques uns s'écartent de l'origine. Ce sont :

ETGo et JNederLand s'opposant sur l'axe 5 à ESÉNégal et JUnited King. d'une part, à EBLgique et JDeutschLand d'autre part, ces deux derniers groupes s'opposant entre-eux sur l'axe 6. Du côté de EBLgique sur l'axe 6, EJordanie s'écarte du centre avec JRouManie.

Les faits les plus massifs où entrent en jeu les trois plus gros exportateurs : EUSA, EMaroc et ECC : URSS, ayant déterminé les premiers axes, le plan 5-6 se trouve étranger à ces faits : les exportateurs USA, Maroc et URSS se projettent au centre de ce plan. Tandis qu'apparaissent des faits n'intéressant qu'un petit nombre de pays :

ETGo 2-ème fournisseur de JNederLand après EUSA ;

EBLgique, 4-ème fournisseur de JDeutschLand après EUSA ,
ECC : URSS et EMaroc ;

ESÉNégal, 2-ème fournisseur de JUnited King. après EUSA ; et enfin, quoique plus central :

EJordanie, 4-ème fournisseur de JRouManie, après EMaroc, ECC et EUSA.

Ainsi le plan 5-6 rend compte de faits réels bien que ne représentant que 8% de l'inertie du nuage. Mais ces faits, d'autant plus intéressants qu'ils sont moins évidents *a priori*, se verront avec plus de précision sur le tableau des facteurs (§ 6).

Au cours de l'examen des graphiques plans deux questions se posent constamment : quels éléments ont déterminé les axes ? Où est situé dans le nuage multidimensionnel un point dont on observe la projection sur tel ou tel plan ? La réponse à ces questions se fait à l'aide des coefficients CTR et COR qui sont imprimés avec les facteurs : cf. § 6.

6 Les tableaux des facteurs sur I et sur J : aides à l'interprétation

Les facteurs sur I et sur J sont présentés dans deux tableaux distincts mais semblables : en lignes, les points représentant les éléments de I (ou de J) dont les facteurs sont les coordonnées sur les axes principaux d'inertie. Trois premières colonnes Qualité, Poids, INertie en concernent pas un facteur particulier (cf. § 6.1). Puis pour chaque facteur, un groupe de trois colonnes : le facteur lui-même, F_α ; COR ou CO2 (cf. § 6.3) ou part de l'axe α dans l'explication de l'écart du point considéré au centre de gravité du nuage ; CTR (cf. § 6.2) ou part du pays considéré dans la dispersion du nuage le long de l'axe α . En général, on n'imprime pas tous les facteurs : la largeur du papier d'imprimante permet le listage sur une ligne de 7 facteurs. C'est ce qui justifie la présence de la colonne OLT donnant la qualité de la représentation de chaque point par ses 7 premiers facteurs.

N.B.1 Sauf avis contraire, toutes les valeurs imprimées sur les tableaux sont en millièmes ; e.g. 73 est à lire : 0,073.

N.B.2 Les sigles désignant les diverses colonnes des tableaux peuvent différer d'un programme à l'autre. Par exemple le poids est tantôt désigné par POID, tantôt par PDS etc. .

6.1 Les colonnes : qualité, poids, inertie

■ PDS : $f_i = k(i)/k$; $f_j = k(j)/k$; f_i : rapport du total $k(i)$ de la ligne i du tableau $k_{I,J}$ au total k du tableau ; f_j : rapport du total $k(j)$ de la colonne j du tableau $k_{I,J}$ au total général k . Les quantités déjà présentées au § 1.2 sont les poids attribués aux divers éléments de I et de J dans l'analyse. Les f_i constituent la colonne POIDS du tableau des profils des lignes ; les f_j la ligne POIDS du tableau des profils des colonnes. Le total de la colonne poids vaut pour chacun des tableaux (de I ou de J) : $1000/1000 = 1$ car $\sum\{k(i) | i \in I\} = \sum\{k(j) | j \in J\} = k$.

■ INR : $INR(i) = \text{inert}_{fJ}(f_J^i) / \text{inertie totale de } N(I)$;

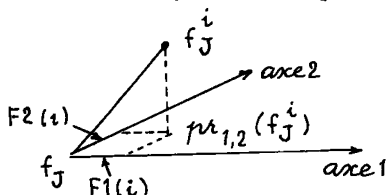
$$INR(j) = \text{inert}_{fI}(f_I^j) / \text{inertie totale de } N(J) .$$

Pour chaque point f_J^i de $N(I)$ (profil de la ligne afférente au pays importateur i) ou f_I^j de $N(J)$ (profil de la colonne du pays exportateur j), INR donne en millièmes la valeur de l'inertie de ce point par rapport au centre de gravité du nuage (i.e. du produit de la masse par le carré de la distance au centre), rapportée à l'inertie totale du nuage.

Remarque la colonne POIDS donne pour chaque pays i la part qu'il a dans le total du tableau ; et la colonne INR la part qu'il a dans l'inertie totale du nuage. Si pour ce pays, INR est nettement supérieur à POIDS, cela ne peut être dû qu'à sa distance élevée à l'origine. On a ainsi le moyen de repérer les points excentriques de même que les points centraux, pour lesquels POIDS est nettement supérieur à INR. Par exemple, JEE est excentrique, JPL relativement proche du centre (en effet, l'Europe de l'Est a l'URSS pour fournisseur quasi exclusif, tandis que la PoLogne (cf. *infra*) a recours à plusieurs exportateurs (en particulier au Maroc et aux USA).

$$\blacksquare \text{ QLT} : \text{QLT}(i) = d^2(f_J, \text{pr}_{1, \dots, n}(f_J^i)) / d^2(f_J, f_J^i) \leq 1.$$

Soit i un élément de I , de profil f_J^i dans l'espace R_J . Le tableau comporte, ligne i , n facteurs imprimés (en général $n = 7$ ou $n = 14$ cf. *supra*). Ces facteurs sont les coordonnées du profil f_J^i sur les n premiers axes principaux d'inertie et décrivent la projection de f_J^i sur l'espace engendré par ces axes ; projection que nous notons : $\text{pr}_{1, \dots, n}(f_J^i)$ (cf. § 4.0). La figure correspond au cas $n = 2$:



(f_J est le centre de gravité du nuage $N(I)$, origine des axes).

La projection orthogonale diminue les distances, en sorte que l'on a toujours :

$$d(f_J, \text{pr}_{1, \dots, n}(f_J^i)) \leq d(f_J, f_J^i),$$

l'égalité n'ayant lieu que si f_J^i coïncide avec sa projection, i.e. est située dans l'espace engendré par les n premiers axes. Dans ce cas, $\text{QLT} = 1$ et la représentation de f_J^i par ses n premiers facteurs est parfaite. S'il n'en est pas ainsi, la représentation dans l'espace engendré par les n premiers axes sera d'autant meilleure que f_J^i sera plus proche de sa projection et QLT plus voisine de 1.

Dans notre exemple, il y a exactement 7 axes factoriels, en sorte que tous les facteurs sont imprimés. Quel que soit i , $\text{QLT} = 1 = 1000/1000$. Nous reviendrons sur QLT à propos de COR (§6.3.4).

Ce que nous venons de dire se transpose au tableau des facteurs sur J : pour tout élément j de J de profil f_I^j dans R_I ,

$$\text{QLT}(j) = d^2(f_I, \text{pr}_{1, \dots, n}(f_I^j)) / d^2(f_I, f_I^j) \leq 1.$$

6.2 Contribution relative d'un point à un axe : CTR

Nous considérons successivement le calcul de cette contribution CTR pour un point i du nuage $N(I)$ (§ 6.2.1) ; le calcul, tout analogue, fait pour un point j de $N(J)$ (§ 6.2.2) ; l'utilisation simultanée des CTR afférents à I et à J (§ 6.2.3).

6.2.1 Calcul de CTR pour un point i du nuage $N(I)$

$$\blacksquare \text{ CTR}_\alpha(i) = f_i(F_\alpha(i))^2 / \lambda_\alpha.$$

On sait (cf. § 4.2) que : $\lambda_\alpha = \sum \{f_i(F_\alpha(i))^2 \mid i \in I\}$: $\text{CTR}_\alpha(i)$ donne donc pour tout pays i de I la part de ce pays i dans l'inertie λ_α du nuage $N(I)$ projeté sur l'axe α .

La somme $\sum \{\text{CTR}_\alpha(i) \mid i \in I\}$ vaut $1000/1000 = 1$: le total 1000 est souvent écrit au bas de la colonne CTR.

6.2.2 Calcul de CTR pour un point j du nuage N(J) : La symétrie avec N(I) est totale :

$$\blacksquare \text{CTR}_\alpha(j) = f_j(G_\alpha(j))^2 / \lambda_\alpha.$$

On a : $\lambda_\alpha = \sum \{f_j(G_\alpha(j))^2 \mid j \in J\}$: $\text{CTR}_\alpha(j)$ donne donc pour tout pays j de J la part de ce pays j dans l'inertie λ_α du nuage N(J) projeté sur l'axe .

La somme $\sum \{\text{CTR}_\alpha(j) \mid j \in J\}$ vaut 1000/1000 = 1.

6.2.3 Utilisation simultanée des CTR afférents à I et à J

En parcourant la colonne CTR pour un nuage et un axe donnés, on repère les éléments qui ont le plus de part à l'inertie du nuage projeté sur l'axe. On dit que ces éléments font l'axe. Voici ce qu'on entend par là. D'une part, si on refait l'analyse sur un tableau de données modifié en supprimant certaines lignes ou colonnes dont aucune ne correspond à l'un de ces éléments, le facteur se retrouve généralement peu modifié. Si, d'autre part, on supprime des éléments apportant une contribution majeure à un facteur, celui-ci disparaît. Quand un facteur est créé par un ou deux éléments comme c'est le cas ici pour le facteur 1 créé par JEE et ECC (cf. *infra*), la suppression de ceux-ci fait disparaître ce facteur, le reste de l'analyse subsistant peu modifié mais avec un décalage des axes. Dans le cas présent, le plan 2-3 se retrouverait comme plan 1-2 ; etc. . On peut ainsi répondre à la question formulée à la fin du § 5 : quels éléments ont déterminé les axes?

I PR	QALT	PDS	INR	F#1	CO2 CTR	F#2	CO2 CTR	F#3	CO2 CTR	F#4	CO2 CTR	F#5	CO2 CTR	F#6	CO2 CTR	F#7	CO2 CTR
IBL	1000	60	22	1	0	-404	390	31	358	386	50	-293	206	63	-194	90	39
ICA	1000	86	139	787	342	117	1038	595	291	130	9	-265	39	74	45	1	3
IFR	1000	150	182	208	57	14	-648	557	200	498	328	245	153	31	43	16	0
IDL	1000	90	50	-59	6	1	242	94	17	516	427	157	107	19	13	306	149
IIT	1000	59	15	226	180	7	-208	155	8	317	356	39	51	9	2	6	0
IJP	1000	70	37	514	450	41	473	381	50	222	84	23	190	62	31	0	0
INL	1000	58	51	242	59	7	-522	279	50	444	281	74	5	0	0	599	367
ISP	1000	57	77	8	0	0	-766	389	105	822	449	252	435	126	132	37	1
IUK	1000	49	56	156	19	3	715	410	80	191	29	12	47	2	1	681	370
IIN	1000	52	63	398	117	18	411	124	28	-471	164	76	891	585	502	2	0
IBR	1000	68	44	569	444	49	554	422	66	-52	4	1	-269	100	60	18	0
IPL	1000	82	29	-558	883	57	-93	23	2	-145	55	11	3	0	0	-150	59
IRM	1000	43	38	-642	525	39	178	40	4	379	184	41	381	184	76	147	28
IEE	1000	78	284	1932	922	647	523	67	67	129	4	0	-55	1	3	27	0

J PR	QALT	PDS	INR	F#1	CO2 CTR	F#2	CO2 CTR	F#3	CO2 CTR	F#4	CO2 CTR	F#5	CO2 CTR	F#6	CO2 CTR	F#7	CO2 CTR
EBJ	1000	32	67	145	9	2	439	84	20	1150	574	280	47	1	1	436	83
EUS	1000	355	201	528	443	220	585	543	385	50	4	6	-75	9	25	-9	0
EJR	1000	27	80	0	0	0	466	67	19	-784	190	111	1519	713	773	-164	8
ENR	1000	337	122	-44	5	2	-491	682	259	-370	342	384	-133	45	74	20	1
ESN	1000	37	64	299	46	7	-767	386	68	432	97	45	460	109	94	707	258
ET6	1000	61	95	156	14	3	-821	387	129	693	275	191	136	11	14	-721	299
ETN	1000	42	32	-100	12	1	-449	239	27	407	195	46	184	40	17	118	17
ECC	1000	108	339	-1784	913	765	523	78	94	155	7	17	33	0	2	1	0

Sur l'axe 1, on repère, dans I :

colonne CTR1 : JEE : 647 avec F1 = -1932

JCA : 117 avec F1 = 787 ;

on en déduit, quant à l'ensemble des importateurs N(I), que l'axe 1

est fait principalement par JEE, secondairement par JCA qui s'oppose sur cet axe à JEE.

Dans J, on repère :

colonne CTR1 : $\Sigma CC = 765$ avec $F1 = -1784$;
 $\Sigma US = 220$ avec $F1 = 528$;

l'axe 1 est fait par ΣCC ; secondairement par ΣUSA qui s'oppose sur l'axe à $\Sigma CC(URSS)$.

Dans la représentation simultanée des deux nuages, en projection sur le plan 1-2, on voit en effet aux deux extrémités de l'axe 1 : Europe de l'Est associé à ΣCC d'une part ; JCanada, ΣUSA d'autre part.

Sur l'axe 2, on repère :

pour I, JCA : CTR = 291 ; F2 = 1038
 JFR : CTR = 200 ; F2 = -648
 JSP : CTR = 105 ; F2 = -766
 pour J, ΣUS : CTR = 385 ; F2 = 585
 ΣMR : CTR = 259 ; F2 = -491

L'axe 2 est donc fait par l'opposition entre les exportateurs USA($F2 > 0$) et Maroc($F2 < 0$) au sein de N(J) ; et, corrélativement, au sein de N(I), par l'opposition entre Canada d'une part ($F2 > 0$) ; et France et Espagne d'autre part ($F2 < 0$), oppositions qui apparaissent dans la représentation simultanée des deux nuages sur les plans 1-2 et 2-3 : cf. §§ 5.2.1 et 5.2.2. La part prépondérante des USA dans ΣJ , alors que son poids (355) est voisin de celui du Maroc (337) est due à l'attraction de JCanada dont la totalité des importations de phosphates vient de USA($f_{\Sigma JCA}^{\Sigma US} = 99,9\%$).

L'axe 3 est déterminé par les oppositions suivantes :

dans N(I) : Espagne : CTR = 252 ; F3 = -822
 France : CTR = 245 ; F3 = 498
 dans N(J) : MaRoc : CTR = 305 ; F3 = -370
 Belgique : CTR = 280 ; F3 = 1150

En poids la Belgique représente le dixième du poids du Maroc. La France et l'Allemagne absorbent 83% de ses exportations de phosphates (cf. profil de ΣBL § 1.3). Nous trouvons ici (axe 3) la Belgique associée à la France ; sur l'axe 6 nous la trouverons associée à l'Allemagne.

Poursuivons l'examen des axes.

Axe 4 :

dans N(I) : JIN : CTR = 503 ; F4 = 891
 dans N(J) : ΣJR : CTR = 773 ; F4 = 1520

L'axe 4 est donc l'axe de la Jordanie, associée à son principal client : l'Inde.

L'axe 5 oppose $\Sigma SÉNÉgal$, associé à JUnited Kingdom, à $\Sigma ToGo$ associé à JNederLand.

L'axe 6 oppose $\mathcal{E}$ Belgique associé à $\mathcal{J}$ DeutschLand, à $\mathcal{E}$ Sénégal associé à $\mathcal{J}$ UnitedKingdom.

L'axe 7 est l'axe de la Tunisie : $\mathcal{E}$ TN : COR = 836 ; F7 = 637. C'est le seul pays exportateur se projetant sur la partie positive de l'axe, associé à $\mathcal{J}$ France, $\mathcal{J}$ Italie, et $\mathcal{J}$ Pologne, qui importent respectivement 12%, 9% et 7% de leurs phosphates de Tunisie et qui, à eux trois, absorbent 76% des exportations de ce pays.

Ainsi, comme nous l'annoncions au § 5.1, tous les axes de cette analyse rendent compte de réalités clairement interprétables.

6.3 Contribution relative d'un axe à l'écart d'un point au centre

Nous touchons ici au but essentiel de l'analyse qui est de déterminer en quoi un pays s'écarte de la moyenne, laquelle est représentée par le centre de gravité du nuage, origine des axes factoriels. Après les formules de définition de $COR\alpha(i)$ et de $COR\alpha(j)$ (§ 6.3.1) et (6.3.2), nous considérons successivement la définition usuelle de cette contribution COR comme un $\cos^2$ (§ 6.3.3) ; le lien entre COR et QLT (§ 6.3.4) ; l'interprétation de COR comme corrélation (§ 6.3.5) ; et enfin nous utilisons les COR pour compléter l'interprétation des graphiques commencée au § 5.

6.3.1 Calcul de COR pour un point i du nuage $N(I)$

$$\blacksquare \quad COR\alpha(i) = (F\alpha(i))^2 / d^2(f_J, f_J^i).$$

On se rappelle que f_J est le centre de gravité du nuage $N(I)$ (§ 2.2) ; f_J^i , profil du pays i , le point qui représente i dans le nuage $N(I)$ (§ 2.1.2) ; et que le carré de la distance de f_J à f_J^i a pour expression en fonction des facteurs (§§ 4.0 et 4.2) :

$d^2(f_J, f_J^i) = \Sigma \{ (F\alpha(i))^2 \mid \alpha = 1, \dots, 7 \}$; (7 est le nombre de dimensions du nuage dans notre exemple, i.e. le nombre de valeurs propres non nulles). $COR\alpha(i)$ est donc la part du facteur α dans le carré de la distance du point représentant le pays i au centre du nuage, f_J , qui est l'origine des axes factoriels.

6.3.2 Calcul de COR pour un point j du nuage $N(J)$

$$\blacksquare \quad COR\alpha(j) = (G\alpha(j))^2 / d^2(f_I, f_I^j).$$

f_I est le centre de gravité du nuage $N(J)$; f_I^j le point représentant le pays j dans le nuage $N(J)$ et l'on a :

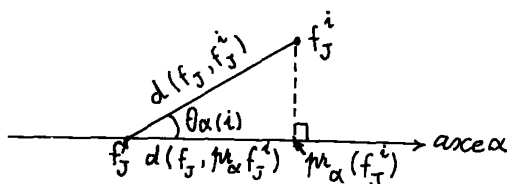
$$d^2(f_I, f_I^j) = \Sigma \{ (G\alpha(j))^2 \mid \alpha = 1, \dots, 7 \} ;$$

$COR\alpha(j)$ est donc la part du facteur α dans le carré de la distance du point représentant le pays j au centre f_I du nuage, origine des axes factoriels.

Dans la suite, ce que nous dirons d'un $COR\alpha(i)$ ou $COR\beta(i)$ vaut pour un j .

6.3.3 COR est un cosinus au carré : Reprenons la formule de définition : $COR\alpha(i) = (F\alpha(i))^2 / d^2(f_J, f_J^i)$. Le numérateur $(F\alpha(i))^2$ n'est autre que le carré de la longueur de la projection du rayon (f_J, f_J^i) joignant l'origine : f_J au point représentant i : f_J^i , projection que l'on

note : $(f_J, pr_\alpha(f_J^i))$.



$$COR\alpha(i) = d^2(f_J, pr_\alpha(f_J^i)) / d^2(f_J, f_J^i).$$

Ce rapport caractérise l'angle $\theta_\alpha(i)$ que fait avec l'axe α la direction dans laquelle le pays i s'écarte du centre :

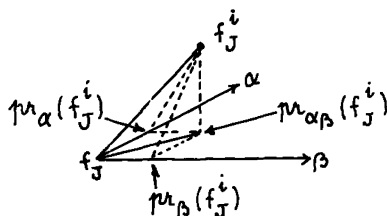
$$COR\alpha(i) = \cos^2 \theta_\alpha(i).$$

Si cet angle est nul, i.e. si le point f_J^i est sur l'axe, $COR\alpha(i) = 1$. On dit alors que l'axe α explique à lui seul l'écart de i au centre. C'est à peu près le cas de l'axe 1 pour JEE : $COR1(JEE) = 922/1000 \approx 1$. Au contraire, si le point i s'écarte du centre dans une direction perpendiculaire à l'axe considéré l'angle θ vaut 90° et $COR\alpha(i) = 0$. On dit que l'axe est étranger à l'écart du point i au centre. C'est le cas de l'axe 5 pour JFR : $COR5(JFR) = 0$.

Ainsi, en parcourant dans la ligne d'un pays les divers COR afférents aux divers axes, on repère les axes qui expliquent l'écart des pays au centre : ce sont ceux pour lesquels COR a une forte valeur ; dans le nuage, le pays est situé à proximité de ces axes. On a ainsi le moyen de répondre à la question : où se trouve, dans le nuage multidimensionnel, un point dont on observe la projection (cf. § 5 *in fine*).

Considérons deux axes α et β et le plan qu'ils engendrent ; notons $pr_{\alpha\beta}$ la projection sur ce plan ; pr_α la projection sur l'axe α ; pr_β celle sur l'axe β . Pour un point f_J^i du nuage (comme pour tout autre point) la projection de f_J^i sur l'axe α coïncide avec celle de la projection de f_J^i sur le plan α - β .

$$pr_\alpha(f_J^i) = pr_\alpha(pr_{\alpha\beta}(f_J^i)) ;$$



et de même :

$$pr_\beta(f_J^i) = pr_\beta(pr_{\alpha\beta}(f_J^i)).$$

Notons $\theta_\alpha(i)$, $\theta_\beta(i)$, $\theta_{\alpha\beta}(i)$ les angles que fait le rayon $f_J f_J^i$ avec l'axe α , l'axe β et sa projection sur le plan $\alpha\beta$: $(f_J, pr_{\alpha\beta}(f_J^i))$, on a :

$$d^2(f_J, pr_{\alpha\beta}(f_J^i)) = d^2(f_J, pr_\alpha(f_J^i)) + d^2(f_J, pr_\beta(f_J^i))$$

d'où :

$$\cos^2 \theta_{\alpha\beta}(i) = \cos^2 \theta_\alpha(i) + \cos^2 \theta_\beta(i).$$

En sorte que la somme : $(COR_\alpha(i) + COR_\beta(i))$ donne le $\cos^2$ de l'angle que fait le rayon (f_J, f_J^i) avec le plan $\alpha\beta$.

L'interprétation de COR comme un $\cos^2$ justifie la notation COR^2 parfois utilisée pour COR.

6.3.4 Calcul de la qualité QLT comme somme de COR : Au § 6.1 nous avons défini la qualité de représentation QLT d'un point i de I par les n premiers facteurs comme le rapport :

$$QLT = d^2(f_J, pr_{1, \dots, n}(f_J^i)) / d^2(f_J, f_J^i).$$

Or, $pr_{1, \dots, n}(f_J^i) = \{F_1(i), F_2(i), \dots, F_n(i)\}$, ensemble des coordonnées de f_J^i sur les n premiers facteurs (cf. § 4.0), et :

$$d^2(f_J, pr_{1, \dots, n}(f_J^i)) = F_1(i)^2 + F_2(i)^2 + \dots + F_n(i)^2 ;$$

en sorte que le rapport QLT n'est autre que la somme des n coefficients de corrélation COR afférents au point i :

$$QLT = COR_1 + COR_2 + \dots + COR_n.$$

Si, comme c'est le cas dans notre exemple, la totalité des facteurs se trouve imprimée, i.e. si n est égal au nombre de valeurs propres non nulles, on a, pour tout élément i :

$$d^2(f_J, f_J^i) = (F_1(i))^2 + (F_2(i))^2 + \dots + (F_n(i))^2 ;$$

et la somme des COR de la ligne vaut 1 :

$$COR_1(i) + \dots + COR_n(i) = ((F_1(i))^2 + \dots + (F_n(i))^2) / ((F_1(i))^2 + \dots + (F_n(i))^2)$$

(de même que la somme des CTR d'une colonne vaut 1 : cf. § 6.2.2). Nous retrouvons ici le fait que QLT vaut 1 quand la représentation est parfaite. Sion, la somme des COR afférents à un élément i est inférieure à 1 ; de façon précise, si, sur le nombre total p de facteurs on n'en imprime que n (les n premiers), la qualité est à diminuer des COR afférents aux facteurs non pris en compte :

$$QLT = 1 - (COR_{(n+1)} + COR_{(n+2)} + \dots + COR_p)$$

6.3.5 Interprétation de COR comme corrélation : On peut montrer que $COR_1(i)$ est un coefficient de corrélation entre fonctions sur J :

Revenons au tableau de données $\{k(i, j) | i \in I, j \in J\}$ et divisons tous les nombres $k(i, j)$ par le total k . On obtient le tableau $\{f_{ij} | i \in I, j \in J\}$ avec $f_{ij} = k(i, j)/k$. Posons encore $d^{ij} = f_i f_j$ où $f_i f_j$ sont les poids des éléments i et j définis au § 1.2. Le pays i étant fixé, $\{d^{ij} | j \in J\}$ est une fonction sur J : d^{iJ} . Et $COR_\alpha(i)$ est le coefficient de corrélation entre le facteur C_α (fonction sur J) et d^{iJ} .

(fonction sur J également). De même $COR\alpha(j)$ est un coefficient de corrélation entre $F\alpha$ et d^{1j} (fonctions sur I).

C'est pourquoi l'on dit d'un pays i ou j pour lequel COR est voisin de 1 qu'il est *bien corrélé à l'axe α* .

6.3.6 Usage de COR dans l'interprétation des graphiques : Comme nous l'avons dit plus haut § 6.3.3, en parcourant dans la ligne d'un pays les COR afférents aux divers axes, on repère les axes qui expliquent l'écart du pays au centre : ce sont ceux pour lesquels COR a une forte valeur; dans le nuage, le pays est situé à proximité de ces axes. On a ainsi le moyen de répondre à la question : où est situé, dans le nuage multidimensionnel le point dont on observe la projection sur un plan (cf. § 5 *in fine*).

Voici comment la lecture des COR aide à interpréter les groupements observés sur les graphiques plans.

Dans le plan 1-2 (§ 5.2.1) se trouvent groupés : TuNisie , MaRoc , Belgique , Sénégal et ToGo . L'examen des COR donne pour chacun de ces pays le sous-espace de R_J dont il est le plus proche.

TuNisie : axes 7, 2, 3 ; MaRoc : axes 2, 3 ; Belgique : axes 3, 6 ; Sénégal : axes 2, 5 ; ToGo : axes 2, 3, 5.

Il faut donc exclure de ce groupe la Belgique dont l'écart au centre n'est expliqué qu'à 8% par l'axe 2 ($COR = 84$) ; le point Belgique est proche du plan 3-6 : $COR3 + COR6 = 820$; nous avons vu en consultant les CTR que BL déterminait l'axe 3 en opposition à MaRoc et l'axe 6 en opposition à Sénégal . Nous voyons ici, en consultant les COR que l'axe 3 explique 57% de son écart au centre, et l'axe 6 : 25%. En consultant de même les COR sur l'ensemble I , on voit que France et DeutschLand sont bien corrélés tous deux à l'axe 3 : $COR = 328$ pour la France et 427 pour l'Allemagne ; ainsi la position de BL à l'extrémité positive de l'axe 3 dans le Plan 2-3 (§ 5.2.2) et celle, du même côté de l'axe, de chacun des deux pays France et DeutschLand traduit bien l'association de ceux-ci avec Belgique dont ils absorbent à eux deux, nous l'avons déjà dit, 84% des exportations.

Il reste à expliquer la position isolée de DeutschLand sur les plans présentés au § 5. La ligne des COR de JDL fait ressortir les axes 3, 5, 6. L'Allemagne a un profil de composantes moyennes sur les principaux exportateurs CC , USA , mais de composante inférieure à la moyenne sur MaRoc . Son originalité est son association avec la Belgique. Donc, sur le plan 1-2 JDL occupe une position intermédiaire entre CC et USA à cause de son déficit en MaRoc . Dans le plan 2-3, JDL se projette sur l'axe 2 avec USA ; sur l'axe 3 avec Belgique . Dans le plan 5-6 il se projette sur l'axe 5 du côté de Sénégal à cause de son déficit en ToGo (alors que son profil a une composante moyenne sur SN) et, sur l'axe 6, avec Belgique .

Roumanie , dans le plan 1-2 occupe comme JDL une place intermédiaire entre CC et USA . Sa ligne de COR indique une très bonne corrélation avec l'axe 1 (525) et une corrélation non négligeable avec les axes 3 et 4 (184 pour chacun d'eux, soit 368 pour le plan 3-4). Par rapport à la moyenne, le profil de la Roumanie est excédentaire en CC : 29% contre 10% ; déficitaire en USA : 18% contre 35% ; très légèrement excédentaire en MaRoc (37% contre 33%) ce qui explique sa position sur le plan 1-2 : dans le quadrant de CC mais toutefois assez centrale par rapport aux trois pôles qui caractérisent ce plan : CC , MR et US . Mais transversalement au plan 1-2, RM s'écarte sur l'axe 3 attiré par MaRoc qui caractérise ce même côté de l'axe ; RM s'écarte également dans la direction de l'axe 4, attiré par Jordanie d'où il importe 12% de ses phosphates (alors que la composante sur JR du profil moyen est 3%).

Nous ne dirons rien ici de JPoLogne qui se projette sur le plan 1-2 au voisinage de la Roumanie et, sur le plan 2-3 près de l'origine, dans le quadrant du Maroc : JPoLogne servira d'exemple au § 7 où nous étudierons l'évolution temporelle à l'aide des éléments supplémentaires.

Il reste à commenter la position de $\mathcal{E}CC$ et $\mathcal{J}EE$ sur le plan 2-3 (§ 5.2.2). L'écart au centre du nuage de $\mathcal{E}CC$ est expliqué à 91% par l'axe 1. Transversalement au plan 2-3, $\mathcal{E}CC$ s'oppose fortement à $\mathcal{E}US$ dans la direction de l'axe 1 négatif ; mais dans le plan 2-3, en opposition à $\mathcal{E}MaRoc$, $\mathcal{E}CC$ est associé à $\mathcal{E}US$. Ceci s'explique parce que la pénétration de $\mathcal{E}CC$ en dehors de l'Europe de l'Est ($\mathcal{J}EE$) se fait en particulier en Deutschland, pays où, des deux fournisseurs mondiaux $\mathcal{E}MaRoc$, et $\mathcal{E}US$, celui-ci l'emporte nettement.

7 Les tableaux supplémentaires

7.1 Projection d'un profil quelconque sur les axes factoriels

Supposons donné un tableau $IS \times J$ ayant même ensemble J de colonnes que le tableau principal 14×8 objet des §§ précédents, mais dont l'ensemble IS des lignes est un ensemble *supplémentaire* ajouté à l'ensemble I . Par exemple, une ligne is de IS pourra représenter un pays importateur non compris dans I tel que le Portugal ou le Danemark : on aura, à la croisée de is et d'une colonne j , la masse $k(is, j)$ de P_{205} exportée par j vers is pendant la période 1973-1980 ; et le total $k(is)$ de cette ligne sera la masse totale de P_{205} importée par is durant cette période, sans distinction de provenance. Le *Profil* de is sur J s'obtient comme pour les éléments principaux (§ 1.3) en divisant la ligne is par son total $k(is)$:

$$\text{profil de } is \text{ sur } J = f_J^{is} = \{(k(is, j)/k(is)) \mid j \in J\}.$$

Ce profil est un point de R_J appartenant au simplexe des profils sur J : $P_J \subset R_J$; $f_J^{is} \in P_J$. Dans P_J (où est situé également le nuage $N(I)$), l'analyse du tableau $k_{I, J}$ à 14 lignes et 8 colonnes a mis en place un système de 7 axes factoriels centrés en f_J , centre de gravité de $N(I)$. on utilise ce système de coordonnées pour repérer n'importe quel profil sur J .

La calcul pratique des coordonnées d'un profil quelconque sur J se fait très simplement par la formule de transition (cf. § 4.4). Ainsi, pour le profil f_J^{is} de l'élément supplémentaire is , on a :

$$Fa(f_J^{is}) = (\lambda\alpha)^{-1/2} \sum \{f_J^{is} G\alpha(j) \mid j \in J\}.$$

Remarque 1 : Il est clair qu'on peut procéder de même pour adjoindre à l'analyse une colonne supplémentaire js en appliquant la formule de transition aux composantes de son profil sur I .

Remarque 2 : Supposons qu'un tableau principal $I \times J$ comporte une ligne $i \in I$ dont le total soit très faible relativement au total général k . Les résultats de l'analyse seront les mêmes que si cette ligne était retranchée du tableau principal, l'élément $i \in I$ recevant sur les axes factoriels les coordonnées que lui donne la formule des éléments supplémentaires. C'est pourquoi on dit que, tandis que les éléments principaux se placent sur des axes qu'ils ont créés en ce sens que ces axes sont ajustés au nuage $N(I)$, les éléments supplémentaires se projettent sur les axes mais ne participent pas à la création de ceux-ci.

7.2 Utilisation des éléments supplémentaires

Nous ne pourrions donner ici qu'un aperçu de l'utilisation des éléments supplémentaires. Quiconque aura étudié attentivement de nombreux exemples concrets (e.g. dans le domaine de l'économie) se convaincra que tout problème énoncé en termes généraux se résout mathématiquement par les trois étapes suivantes : collecte de données pertinentes, construction à partir de ces données d'un tableau approprié au problème, lecture de la solution, le plus souvent d'après la position d'éléments supplémentaires.

Citons trois cas d'utilisation d'éléments supplémentaires.

Après une première analyse, on peut être amené à retirer du tableau principal des éléments qui ont une inertie telle qu'ils déterminent à eux seuls un axe, tandis que se trouvent masqués des faits qui intéressent l'ensemble du nuage. On retire ces éléments perturbateurs du tableau principal, mais on les place en éléments supplémentaires, en sorte qu'ils se projettent sur les plans factoriels créés par les éléments laissés en principal. Ainsi dans notre exemple, on aurait pu placer Σ CC(URSS) en colonne supplémentaire et Σ EE (Europe de l'Est) en ligne supplémentaire. Comme nous l'avons signalé au § 5, on supprime par ce procédé le 1-er axe de la 1-ère analyse mais les axes suivants restent axes principaux d'inertie, avec un rang décalé d'une unité par rapport à celui qu'ils ont dans la 1-ère analyse : le 2-ème axe devient le 1-er, le 3-ème devient le 2-ème etc. (Mais il n'en est pas toujours ainsi : il se peut qu'une fois retirés du tableau principal les éléments qui créent à eux seuls un axe, les axes principaux d'inertie du nouveau nuage principal diffèrent des axes issus de la 1-ère analyse).

On peut être également amené à placer un élément nouveau en supplémentaire, qui n'a jamais figuré dans le tableau principal : pour reprendre l'exemple du § 7.1, ce pourrait être le Portugal ou le Danemark.

Nous nous apesantirons davantage sur l'emploi systématique qui a été fait des éléments supplémentaires pour suivre, année par année, les profils annuels des divers pays importateurs de I. Pour simplifier l'exposé nous ne suivrons ainsi, que les profils annuels de la Pologne (§ 7.3).

7.3 Mise en éléments supplémentaires des profils annuels de pays ; exemple de la Pologne

7.3.1 Le nuage des pays-années : Nous avons présenté au § 1.0 le tableau $k_{I,J}$: pour chaque pays i et chaque année t , ce tableau $k_{I \times T, J}$ donne, ligne it , les quantités de P205 importées par i l'année t en provenance des divers pays exportateurs de J : Σ Belgique, Σ USA etc..

Pour chaque pays it on définit son profil f_J^{it} ; son poids f_{it} : total de la ligne it divisé par le total k du tableau principal $k_{I,J}$; d'où, dans l'espace R_J des profils sur J le nuage des pays-années, ensemble de tous les points $\{f_J^{it} | i \in I ; t \in T\}$, chacun muni de son poids f_{it} . Nous ne considérons dans ce § qu'une partie de ce nuage des pays-années : le sous-nuage constitué par les huit points représentant les profils annuels de la Pologne : PL73, PL74, ..., PL80. Ce nuage de huit points $\{PL73, \dots, PL80\}$ est imbriqué dans le nuage $N(I)$ analysé aux §§ précédents et peut être projeté comme nous l'avons dit au § 7.1 sur les axes factoriels et les divers plans qu'ils définissent, issus de l'analyse de $N(I)$.

7.3.2 Calcul des facteurs des profils annuels : A chaque profil annuel de pays e.g. PL73 correspond une ligne du tableau $k_{IT,J}$:

	BL	US	JR	MR	SN	TG	TN	CC	total
PL73	20	3390		47130			5538	40800	96878

d'où un profil sur J dont les composantes s'obtiennent en divisant les nombres de cette ligne par le total de la ligne : 96878 (c'est ainsi qu'on a défini, pour le tableau principal, un profil par ligne au § 1.3).

	BL	US	JR	MR	SN	TG	TN	CC	total
f_J PL73		.035		.486			.057	.421	1

L'analyse du tableau principal a fourni les facteurs de US, MR, TN, CC ; ainsi que les valeurs propres :

	US	MR	TN	CC
F1	.528	-.044	-.100	-1.784
F2	.585	-.491	-.449	.523

$\lambda_1 = .45$; $\lambda_2 = .32$ etc. .

En appliquant la formule de transition, on obtient les facteurs ou coordonnées sur les axes factoriels de PL73 :

$$F1(PL73) = (.45)^{-1/2} ((.528 \times 0.35) - (.044 \times .486) - (.100 \times .057) - (1.784 \times .421)) = -1.132$$

$$F2(PL73) = (.32)^{-1/2} ((.585 \times 0.35) - (.491 \times .486) - (.449 \times .057) + (.523 \times .421)) = -.042$$

Le programme calcule tous les facteurs pour chacun des 8 points : PoLogne-années, et les résultats sont présentés dans le tableau des facteurs sur ISUPplémentaires.

7.3.3 Le tableau des facteurs sur ISUP

D'une façon générale, le tableau des facteurs sur ISUP est présenté comme celui sur IPRincipal avec trois premières colonnes : QLT, POIDS, INR puis des groupes de colonnes F, COR, CTR afférents à chaque axe. Mais les colonnes POIDS, INR, CTR n'ont plus ici total 1 comme pour les lignes (ou colonnes) principales, puisque le total du tableau, l'inertie du nuage et les inerties en projection sur les axes du nuage ont été calculés sur le seul tableau principal. Par exemple

ISUP	QALT	PDS	INR	F#1	C02	CTR	F#2	C02	CTR	F#3	C02	CTR	F#4	C02	CTR	F#5	C02	CTR	F#6	C02	CTR	F#7	C02	CTR
PL73	1000	10	13	1133	905	28	42	1	0	270	37	3	250	44	8	69	3	1	29	1	0	109	8	6
PL74	1000	9	9	853	660	14	308	86	3	391	139	9	281	72	8	88	7	1	27	1	0	194	34	16
PL75	1000	11	6	647	716	10	132	30	1	228	89	4	172	51	4	95	15	2	69	8	2	230	98	27
PL76	1000	11	5	529	519	7	392	284	5	242	109	4	19	1	0	129	31	3	70	2	0	172	55	16
PL77	1000	11	4	311	133	1	51	7	0	172	45	1	211	110	6	587	635	48	94	23	3	142	49	11
PL78	1000	11	2	250	269	1	31	4	0	9	0	0	138	81	2	352	535	23	53	13	1	151	98	12
PL79	1000	10	2	401	618	4	49	10	0	29	3	0	20	2	0	261	262	12	69	19	2	150	86	11
PL80	1000	10	4	587	523	6	247	123	2	220	99	3	312	197	12	152	48	4	52	5	1	50	5	1

POID (PL73) =(total de la ligne PL73 / total de $k_{I,J}$) mesure l'importance de PL73 par rapport à l'ensemble I, sans que PL73 fasse partie de I.

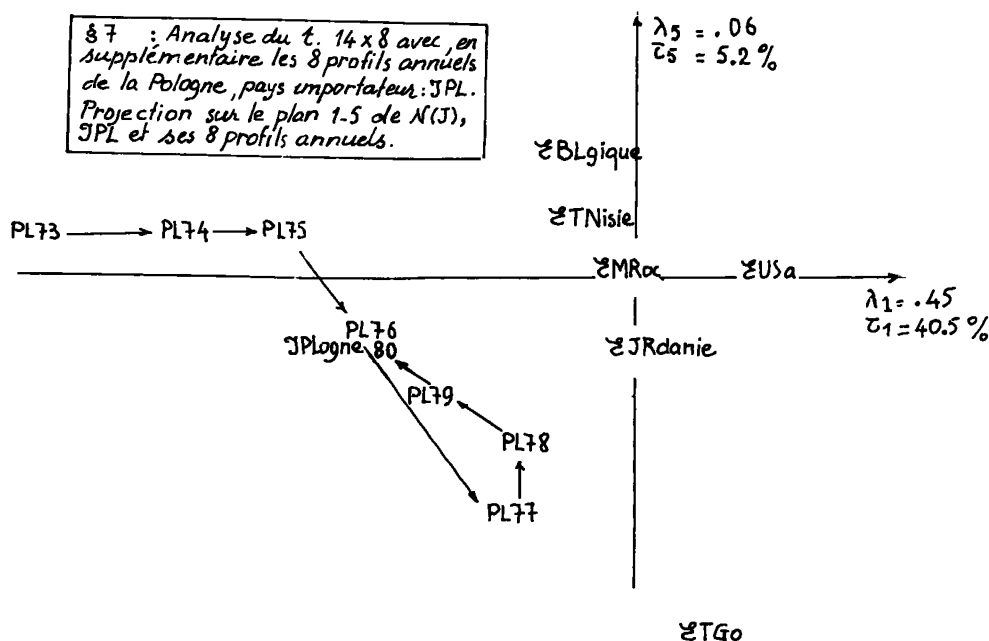
7.3.4 Interprétation des résultats

Les 8 lignes ont toutes approximativement même importance quant au poids. En revanche la colonne INR montre de grandes différences entre les années : PL73 est le plus excentrique des 8 points. Avec les années, le point PoLogne se rapproche de l'origine : PL78 et PL79 sont les plus proches de l'origine ; puis PL80 s'en éloigne.

L'examen des coefficients COR fait apparaître que les 8 points sont bien corrélés à l'axe 1 ; pour tous, sauf deux : PL77 et PL78, $COR > .500$; quant à ces deux derniers points, c'est à l'axe 5 qu'ils sont le mieux corrélés.

Nous donnons le graphique plan 1-5, plan qui se rapproche au mieux de l'ensemble des 8 points ; mais nous devons nuancer l'interprétation à l'aide des facteurs qui ont part à l'écart de tel ou tel point à l'origine.

Σ SNégal



Rappelons que l'axe 1 est l'axe de l'URSS : ΣCC , qui s'oppose à ΣUSA ; et que l'axe 5 est fait de l'opposition entre $\Sigma ToGo$ et $\Sigma S\acute{e}n\acute{e}gal$. Le Maroc ne joue aucun rôle dans ce plan : il se projette à l'origine d'où il s'écarte dans le plan 2-3.

Dans le plan 1-5 représenté ici, on peut suivre d'année en année le cheminement de la Pologne qui s'écarte d'abord de l'URSS parallèlement à l'axe 1 : PL73, 74, 75 ; puis incline sa route vers $\Sigma ToGo$; et

rebrousse chemin en 78, 79, 80, le dernier point PL 80 se trouvant à l'emplacement de PL76.

	BL	US	JR	MR	SN	TG	TN	CC	POIDS
PL73	ε	35	0	486			57	421	10
PL74		6		619			70	304	9
PL75		131	7	484	16	9	83	269	11
PL76		55	30	536	15	79	78	207	11
PL77		270	55	255		172	76	172	11
PL78		267	52	316		122	71	172	11
PL79		231	57	357		99	69	215	10
PL80		264	92	296		40	40	263	10
f_J	32	355	27	337	37	61	42	108	

Profils sur J des huit points Pologne-années.

La colonne $\mathcal{E}CC$ du tableau des profils sur J des huit points Pologne-années montre une décroissance de la composante sur $\mathcal{E}CC$ de PL73 à PL 77, 78 puis une croissance de la composante de PL78 à PL80 : ce qui explique le mouvement de la Pologne en projection sur l'axe 1. La colonne $\mathcal{E}ToGo$ explique le mouvement de la Pologne en projection sur l'axe 5 à partir de 1976, le point extrême, PL77, correspond au maximum de la composante sur $\mathcal{E}ToGo$.

Ce graphique plan donne une vue générale de la trajectoire de la Pologne au cours de la période 73-80 ; mais laisse dans l'ombre des faits ponctuels qui apparaissent sur le listage des 7 facteurs :

PL74, bien corrélé au 3-ème facteur, doit sa position particulière à la forte composante de son profil sur $\mathcal{E}MaRoc$;

PL76, bien corrélé au plan 2-3, est, de même, associé au Maroc.

PL77 et PL80 s'associent à la Jordanie sur l'axe 4.

PL76 et PL80 s'opposent entre eux sur l'axe 2 : PL76 est associé au Maroc ; quant à PL80, il ne s'associe pas à $\mathcal{E}USA$, puisque la composante de son profil sur $\mathcal{E}USA$ est inférieure à la moyenne, mais se projette au voisinage de $\mathcal{E}JR$ auquel il s'associe dans le nuage.

Dans le tableau principal $k_{I,J}$, la ligne JPologne est le cumul des huit lignes supplémentaires PoLogne-années. Dans l'espace R_J , le point JPL est donc le centre de gravité du nuage des huit profils annuels de la Pologne. Nous donnons ici le profil de JPL ainsi que le profil f_J de la ligne de marge de $k_{I,J}$ (cf. § 1.3).

	BL	US	JR	MR	SN	TG	TN	CC
JPL	ε	161	34	415	4	68	69	250
f_J	32	355	27	337	37	61	42	108

La comparaison de ces deux lignes montre que des affinités particulières des divers profils annuels avec les pays de J ne subsiste pour le cumul JPL que celle avec $\mathcal{E}MaRoc$. Quant à l'affinité avec $\mathcal{E}CC$, commune aux 8 profils annuels, elle est très forte pour JPL.

La position de JPL sur les plans 1-2 et 2-3 présentés au § 5 s'explique par l'association : JPologne, $\mathcal{E}MaRoc$.

7.4 L'étude des graphiques plans et la lecture des tableaux nous ont permis de suivre dans l'espace multidimensionnel des profils sur J la trajectoire de la Pologne année par année. Il resterait à faire de même pour les trajectoires des 13 autres importateurs principaux. Et, au-delà de cette étude individuelle des pays une question se pose : les 14 trajectoires occupent-elles des domaines séparés ou s'enchevêtrent-elles ? A combien de types principaux conviendrait-il de réduire la diversité des 112 (14×8) profils annuels des 14 pays importateurs ? Quels rapports ces types ont-ils entre eux ? Autant de questions auxquelles seule la classification ascendante hiérarchique fournira une réponse à la fois synthétique et précise.