

C. BOURGARIT

Analyse du tableau cumulatif du code génétique

Les cahiers de l'analyse des données, tome 4, n° 2 (1979),
p. 219-230

http://www.numdam.org/item?id=CAD_1979__4_2_219_0

© Les cahiers de l'analyse des données, Dunod, 1979, 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/>

ANALYSE DU TABLEAU CUMULATIF DU CODE GÉNÉTIQUE [CORR. GEN.]

par C. Bourgarit (1)

0 Origine du problème : On sait (cf [Code Gén.] Cah. Vol IV n° 2), que la synthèse des protéines chez le vivant, s'effectue par traduction d'un programme dont le support chimique est l'a.r.n. messager, molécule, qui en tant que message peut être assimilée à un mot écrit dans un alphabet à 4 lettres : {U,C,A,G} (symbolisant quatre bases uracyle ; adénine ; cytosine ; guanine). Dans cette traduction, à un triplet de bases (ou codon), correspond une molécule d'acide aminé : e.g. au triplet UGG, correspond une molécule de tryptophane ; etc. Comme il y a $64 = 4^3$ triplets servant de codons pour seulement 20 aminoacides et un signe de terminaison (noté TAU ou TRM) il y a entre les triplets de nombreuses synonymies. Pour en étudier le système, on a construit le tableau de correspondance k_{IJ} , point de départ du présent problème :

I = ensemble des 20 acides aminés et du signe de terminaison (chacun désigné par un sigle à trois lettres : e.g. TRP pour tryptophane etc.)

J = ensemble des bases {U,C,A,G} avec un chiffre 1, 2 ou 3 pour préciser sa place dans le codon.

$k(i,j)$ = nombre de fois que le caractère j se rencontre dans un codon signifiant i : e.g. $k(ALA,C2) = 4$ parce que quatre codons comportant un C (cytosine) en position 2 signifient alanine.

L'analyse de ce tableau a révélé les particularités suivantes :

1° Avec 12 colonnes et entre elles deux relations linéaires (les colonnes allant par bloc de somme constante) on attendait 9 facteurs non triviaux : il y en a 8, parce que les colonnes U3 et C3 sont identiques!

2° Dans le plan des axes 1-2 (cf § 2.6 *in fine*) l'ensemble I est concentré en 4 points ; et l'ensemble J en 5 points cette propriété résulte de ce qu'un codon comportant G1 ne peut être synonyme que d'un codon comportant G1 (et de même pour U2) : c'est la propriété P de la question 3°.

3° Plus précisément, comme on le démontre dans la 2° question pour le facteur lié à G1, les facteurs 1 et 2 sont relatifs à la v. p. (1/3) : pourquoi 1/3? et pourquoi cette v. p. est-elle la plus grande possible pour un tableau de la forme considérée ici : c'est l'objet des questions 1.4 et 1.5.

1 Problème proposé en Mars 1979 aux étudiants du D.E.A. de statistique

L'objet du présent problème est de découvrir dans la structure du tableau k_{IJ} (21×12) : Card I = 21 ; Card J = 12) les raisons de plusieurs particularités observées à l'analyse. Certaines de ces propriétés seront démontrées sur le tableau lui-même ; d'autres seront situées dans un cadre plus général.

(1) Assistante au laboratoire de statistique de l'université Pierre et Marie Curie

	U ₁	C ₁	A ₁	G ₁	U ₂	C ₂	A ₂	G ₂	U ₃	C ₃	A ₃	G ₃
Tryptophane : W : TRP	1							1				1
Isoleucine : I : ILE			3		3				1	1	1	
Tyrosine : Y : TYR	2						2		1	1		
Phenylalanine : F : PHE	2				2				1	1		
Proline : P : PRO		4				4			1	1	1	1
Leucine : L : LEU	2	4			6				1	1	2	2
Valine : V : VAL				4	4				1	1	1	1
Methionine : M : MET			1		1							1
Lysine : K : LYS			2				2				1	1
Cysteine : C : CYS	2							2	1	1		
Alanine : A : ALA				4		4			1	1	1	1
Arginine : R : ARG		4	2					6	1	1	2	2
Threonine : T : THR			4			4			1	1	1	1
Serine : S : SER	4		2			4		2	2	2	1	1
Glycine : G : GLY				4				4	1	1	1	1
Histidine : H : HIS		2					2		1	1		
Aspartic acid : D : ASP				2			2		1	1		
Asparagine : N : ASN			2				2		1	1		
Glutamic acid : E : GLU				2			2				1	1
Glutamine : Q : GLN		2					2				1	1
Terminaison : Z : TAU	3						2	1			2	1

Tableau k_{IJ}

1.1 On considère le tableau k_{IJ} :

$$I = \{TRP, ILE, TYR, \dots, GLN, TAU\}$$

$$J = \{U_1, C_1, A_1, G_1, U_2, C_2, A_2, G_2, U_3, C_3, A_3, G_3\} ;$$

chaque élément i de I est désigné par un sigle de trois lettres - chaque élément j de J est désigné par l'une des quatre lettres U, C, A, G suivies d'un chiffre, 1, 2 ou 3.

a) écrire le tableau de fréquence f_{IJ} ($f_{ij} = k(i,j)/k$) associé au tableau donné k_{IJ} ; en plaçant en marge les lois f_I et f_J

b) écrire le tableau f_I^J de la transition de J vers I ($f_i^j = f_{ij}/f_j$).

c) écrire le tableau f_J^I de la transition de I vers J .

1.2 On considère une fonction φ^J sur J définie par les conditions suivantes :

$$\varphi(U_1) = \varphi(C_1) = \varphi(A_1) = a ; \quad \varphi(G_1) = b ;$$

$\varphi(s_2) = \varphi(s_3) = 0$; où s désigne l'une quelconque des quatre lettres $\{U, C, A, G\}$; autrement dit φ prend la même valeur pour les trois premiers éléments de J ; et il est nul pour les 8 derniers éléments de J .

a) déterminer a et b pour que sur l'ensemble J muni de la loi f_J , la fonction φ^J ait moyenne nulle et variance un.

b) calculer les fonctions ψ^I et θ^J définies par transition à partir de φ^J :

$$\psi^I = \varphi^J \circ f_J^I ; \quad \text{i.e. } \psi^i = \sum \{ \varphi^j f_j^i \mid j \in J \} ; \quad \text{et de même :}$$

$$\theta^J = \psi^I \circ f_I^J ;$$

c) déduire des calculs qui précèdent que, pour une valeur λ_α convenable, la correspondance k_{IJ} possède un facteur (F_α^I, G_α^J) tel que G_α^J soit nul pour les huit derniers éléments de l'ensemble J . Préciser les valeurs λ_α et les facteurs F_α , G_α (de variance λ_α).

1.3 Dans cette question, on considère un tableau de forme générale, construit comme nous l'expliquons ci-dessous. On note

S : un ensemble ou alphabet, formé de p signes : $\text{Card } S = p$; un élément de S sera noté s , ou s' , ou s'' ...

$J = S_1 \cup S_2 \cup \dots \cup S_q$: la réunion de q exemplaires de l'alphabet S (q entier fixé) : un élément J sera noté comme un signe s (ou s' , s'') de S suivi d'un nombre h pouvant varier de 1 à q ; on a donc :

$$J = \{sh \mid s \in S ; h = 1, \dots, q\}.$$

$M = S_1 \times S_2 \times \dots \times S_q$: le produit de q exemplaires de l'alphabet S : autrement dit, M est l'ensemble des mots de q lettres formés sur l'alphabet S . Pour tout élément m de M , et pour tout entier h compris entre 1 et q , on note : $m(h)$, la lettre de rang h du mot m ; on a donc :

$$m = m(1) m(2) \dots m(q) ;$$

de même, on a pour l'ensemble des mots un tableau de description booléen, k_{MJ} ; avec :

$$k(m, sh) = \begin{cases} 1 & \text{si } m(h) = s \\ 0 & \text{sinon} \end{cases}$$

I : une partition de l'ensemble M ; chaque élément i de I est une partie de M (partie non-vide : formée d'au moins un mot m ; et éventuellement de plusieurs mots : m' , m'' ...) ; à deux éléments i et i' distincts correspondent des parties de M , d'intersection vide :

$$\forall i, i' \in I : i \neq i' \Rightarrow i \cap i' = \emptyset ;$$

enfin M est la réunion des i : $M = \cup \{i \mid i \in I\}$.

A la partition I est associé un tableau de correspondance k_{IJ} défini par :

$$k(i, j) = \sum \{k(m, j) \mid m \in i\}.$$

(autrement dit, $k(i, sh)$ est le nombre des mots m qui appartiennent à i et sont tels que leur lettre de rang h est s)

Ceci posé, on dit que la colonne $sh \in J$, possède la propriété P relativement au tableau k_{IJ} si l'on a :

$$\forall i \in I, \forall s' \in S : s' \neq s \Rightarrow k(i, sh) \times k(i, s'h) = 0$$

a) si sh possède la propriété P est-il possible de trouver dans I un i qui contienne à la fois un mot m dont la lettre de rang h est s , et un mot m' ayant dans la position h une autre lettre $s'h$ ($s' \neq s$) ?

Quel est le total d'une ligne m du tableau k_{MJ} ?

Quel est le total d'une colonne sh du tableau k_{IJ} ? Quelle est la loi marginale f_j du tableau k_{IJ} .

b) soit $h, h' \in \{1, \dots, q\}$; $s, s' \in S$. Exprimer en fonction des entiers p et q le nombre

$$N(sh, s'h') = \text{Card}\{m \in M ; m(h) = sh ; m(h') = s'h'\}.$$

(on distinguera trois cas: $h \neq h'$; et $h = h'$ avec $s \neq s'$ ou $s = s'$).

c) on suppose que sh possède la propriété P relativement au tableau k_{IJ} .

Soit $h' \neq h$; calculer en fonction des entiers p et q :

$$L(sh, s'h') = \Sigma\{k(i, s'h') \mid i \in I ; k(i, sh) = 0\}$$

$$R(sh, s'h') = \Sigma\{k(i, s'h') \mid i \in I ; k(i, sh) \neq 0\}$$

d) le tableau k_{IJ} des questions 1.1 et 1.2 a été en fait construit suivant le procédé expliqué en tête de la question 1.3. Quels sont pour ce tableau k_{IJ} (de 1.1 et 1.2) :

l'ensemble S ; l'entier p ; l'entier q .

Préciser dans les trois cas $i = \text{TRP}$, $i' = \text{ILE}$, $i'' = \text{TAU}$ quelles sont les parties correspondantes i, i', i'' de M (i.e. quels sont le ou les mots correspondant à i, i', i'').

Chercher dans le tableau k_{IJ} (de 1.1 et 1.2) s'il existe des colonnes s_1, s_2 ou s_3 possédant la propriété P .

1.4 On se place dans le cas d'un tableau k_{IJ} général construit suivant le procédé expliqué en tête de la question 1.3 ; et on suppose que la colonne sh possède la propriété P .

a) construire une fonction φ^J satisfaisant aux conditions suivantes :

$\forall s' \in S, \forall h' \neq h : \varphi^{s'h'} = 0$ (i.e. φ est nulle en dehors du bloc Sh de J)

$\forall s', s'' \in S - \{s\} : \varphi^{s'h} = \varphi^{s''h}$ (i.e. φ est constante sur $Sh - \{sh\}$)

φ a moyenne nulle et variance 1 sur J muni de la loi marginale f_J du tableau k_{IJ} . On exprimera en fonction des entiers p et q les valeurs φ en sh et en $s'h$ ($s' \neq s$) : on pourra pour fixer les notations supposer que $\varphi^{sh} < 0$.

b) calculer la fonction ψ^I image de la fonction φ^J par la transition f_J^I associée au tableau k_{IJ} : $\psi^I = \varphi^J \circ f_J^I$; on distinguera deux cas pour le calcul de ψ^i : d'une part si $k(i, sh) \neq 0$; d'autre part si $k(i, sh) = 0$.

c) calculer la fonction θ^J image de ψ^I par transition : $\theta^J = \psi^I \circ f_I^J$ dans ce calcul, on utilisera le résultat de 1.3 c)

d) déduire des calculs qui précèdent que pour une valeur propre λ_α convenable, la correspondance k_{IJ} possède un facteur (F_α^I, G_α^J) tel que G_α^J soit nul sur tout segment Sh' ($h' \neq h$) et constant sur $Sh - \{sh\}$. Préciser la valeur λ_α et les facteurs F_α, G_α (pour $F_\alpha(i)$ on distinguera deux cas selon que $k(i, sh)$ est nul ou non)

1.5 On reste dans le cadre général de la question 1.3.

a) on considère le tableau de correspondance k_{MJ} , (ou tableau booléen de description des mots) ; on demande de calculer la transition :

$\pi_J^J = f_J^M \circ f_M^J$ obtenue en composant les deux transitions associées au tableau k_{MJ} . Pour l'expression des éléments $\pi_{j,i}^J$, en fonction des entiers p et q , on distinguera trois cas (cf 1.3 b) :

$$\pi_{s'h'}^{sh} \quad (h \neq h') ; \pi_{s'h}^{sh} \quad (s \neq s') ; \pi_{sh}^{sh}.$$

On pourra dessiner le tableau carré π_J^J dans le cas où $p = 4$ et $q = 3$

b) soit φ^J une fonction sur J , ayant moyenne nulle sur chacun des blocs Sh , i.e.

$$\forall h \in \{1, \dots, q\}: \sum \{\varphi^J(sh) \mid s \in S\} = 0 ;$$

exprimer $\varphi^J \circ \pi_J^J = \varphi^J \circ f_J^M \circ f_M^J$;

Si φ^J a variance 1, quelle est la variance de $\varphi^J \circ f_J^M$?

c) Quels sont sur J les facteurs non constants issus de la correspondance k_{MJ} et relatifs à une valeur propre λ_α non nulle? Que peut être λ_α ?

d) soit φ^J une fonction de moyenne nulle sur J ; on note :

$$\psi^I = \varphi^J \circ f_J^I ; W^M = \varphi^J \circ f_J^M$$

exprimer $\psi^I(i)$ en fonction de $Card i$ et des valeurs $W(m)$ pour $m \in i$; comparer la variance de ψ^I à celle de W^M ; qu'en déduit-on pour les valeurs propres de la correspondance k_{IJ} .

1.6 On revient au tableau particulier k_{IJ} des nos 1.1 et 1.2.

a) d'après 1.2c, 1.2d, 1.4d et 1.5d, donner la représentation des ensembles I et J dans le plan des deux premiers axes.

b) combien le tableau k_{IJ} admet-il de facteurs non triviaux (i.e. non constants et relatifs à des valeurs propres non nulles)? (On tiendra compte des relations linéaires existant entre les colonnes du tableau k_{IJ})

2 Correction de l'épreuve de Mars 1979

2.1 a) Le total $k(j)$ de chaque colonne valant 16, le total général du tableau k_{IJ} vaut $16 \times 12 = 192$, et donc le tableau f_{IJ} est égal à $k_{IJ}/192$, tandis que f_j est la mesure de probabilité uniforme sur $J : \forall j \in J : f_j = 1/\text{Card } J = 1/12$.

La loi marginale sur I s'écrit :

$$f_I = (1/64) (1, 3, 2, 2, 4, 6, 4, 1, 2, 2, 4, 6, 4, 6, 4, 2, 2, 2, 2, 3)$$

b) Puisque $k(j)$ est constant et égal à 16, le tableau f_I^J est identique au facteur $1/16$ près au tableau k_{IJ} .

c) Le tableau f_J^I s'écrit au facteur $1/18$ près :

	U ₁	C ₁	A ₁	G ₁	U ₂	C ₂	A ₂	G ₂	U ₃	C ₃	A ₃	G ₃
TRP	6							6				6
ILE			6		6				2	2	2	
TYR	6						6		3	3		
PHE	6				6				3	3		
PRO		6				6			1,5	1,5	1,5	1,5
LEU	2	4			6				1	1	2	2
VAL				6	6				1,5	1,5	1,5	1,5
MET			6		6							6
LYS			6				6				3	3
CYS	6							6	3	3		
ALA				6		6			1,5	1,5	1,5	1,5
ARG		4	2					6	1	1	2	2
THR			6			6			1,5	1,5	1,5	1,5
SER	4		2			4		2	2	2	1	1
GLY				6				6	1,5	1,5	1,5	1,5
HIS		6					6		3	3		
ASP				6			6		3	3		
ASN			6				6		3	3		
GLU				6			6				3	3
GLN		6					6				3	3
TAU	6						4	2			4	2

2.2 a) φ^J sera de moyenne nulle et de variance 1 si :

$$\begin{aligned} 3a + b &= 0 \\ (3a^2 + b^2)/12 &= 1 \end{aligned}$$

d'où, en choisissant a positif : $a = 1 ; b = -3$

b) $\psi^i = 1/3$ si $k_{IJ}(i, G1) = 0$

$\psi^i = -1$ si $k_{IJ}(i, G1) \neq 0$

$\theta^j = 1/3$ si $j \in \{U1, C1, A1\}$

$\theta^j = -1$ si $j = G1$

$\theta^j = 0$ si $j \in J - \{U1, C1, A1, G1\}$

c) On a : $\theta^J = \varphi^J \circ f_J^I \circ f_I^J = (1/3) \varphi^J$

Donc ψ^J est facteur de variance 1, relatif à la valeur propre $\lambda_\alpha = 1/3$ de k_{IJ} , facteur qui est nul pour les huit derniers éléments de J . Les facteurs F_α et G_α , de variance λ_α s'écrivent alors :

$$G_\alpha(j) = (1/\sqrt{3}; 1/\sqrt{3}; 1/\sqrt{3}; -\sqrt{3}; 0; 0; 0; 0; 0; 0; 0; 0; 0)$$

$$F_\alpha(i) = \psi^i = 1/3 \text{ si } k_{IJ}(i, G1) = 0$$

$$F_\alpha(i) = \psi^i = -1 \text{ si } k_{IJ}(i, G1) \neq 0$$

2.3 a) Si sh possède la propriété P , et s'il existait un élément de I , et deux éléments m et m' de i tels que $m(h) = sh$, $m'(h) = s'h \neq sh$, on aurait $k(i, sh) \geq 1$ et de même $k(i, s'h) \geq 1$, ce qui est impossible puisque $k(i, sh) \times k(i, s'h) = 0$. Si donc sh possède la propriété P , et si $k(i, sh) \neq 0$, on a $\forall m \in i : m(h) = sh$, et $k(m, s'h) = 0$ si $s'h$ est différent de sh .

Le total $k(m) = \sum\{k(m, j) \mid j \in J\}$ d'une ligne m de k_{MJ} est égal à q , puisque, quand j décrit S_h , $k(m, j)$ n'est non-nul et égal à 1 que si $j = m(h)$, h variant de 1 à q .

Le total $k(sh)$ d'une colonne sh de k_{IJ} s'écrit :

$$\begin{aligned} k(sh) &= \sum\{k(i, sh) \mid i \in I\} = \sum\{k(m, sh) \mid m \in i, i \in I\} \\ &= \sum\{k(m, sh) \mid m \in M\} \end{aligned} \quad (1)$$

Il est donc égal, résultat évident, au total de la colonne sh de k_{MJ} , total égal à $pq-1$, puisque les termes non-nuls et égaux à 1 dans (1) sont tels que $m(h)$ est fixé, et égal à sh , $m(h')$ pour $h' \neq h$, pouvant prendre l'une des p valeurs de l'alphabet S .

La loi marginale f_J de k_{IJ} (qui est aussi celle de k_{MJ}) est donc la loi uniforme sur $J : \forall j \in J : f_j = 1/\text{Card } J = 1/(pq)$, puisque J a un cardinal égal à pq .

$$b) N(sh, sh) = \text{Card}\{m \mid m \in M ; m(h) = sh\} = p^{q-1}$$

$$N(sh, s'h) = \text{Card}\{m \mid m \in M ; m(h) = sh ; m(h) = s'h\} = 0 \quad (s \neq s')$$

$$N(sh, s'h) = \text{Card}\{m \mid m \in M ; m(h) = sh ; m(h') = s'h'\} = p^{q-2} \quad (h \neq h')$$

Remarque : k_{MJ} peut être considéré comme un tableau disjonctif complet où chaque possibilité (q -uple) de réponses possibles a été prise une fois et une seule, S_1, S_2, \dots, S_q étant alors les ensembles de modalités associées à chacune des q questions.

N_{JJ} est alors le tableau de Burt associé à k_{MJ} .

$$\begin{aligned} c) L(sh, s'h) &= \sum\{k(i, s'h') \mid i \in I ; k(i, sh) = 0\} \\ &= \sum\{k(m, s'h') \mid m \in i ; i \in I ; k(i, sh) = 0\} \\ &= \sum\{k(m, s'h') \mid m \in M ; k(m, sh) = 0\} \end{aligned} \quad (2)$$

La dernière égalité résultant du fait que sh possédant la propriété P , si $k(m, sh) = 0$, et si i est la classe à laquelle appartient m , $k(i, sh) = 0$. En effet si $k(m, sh) = 0$ il existe $s' \in S$, $s' \neq s$, tel que $k(m, s'h) = 1$, auquel cas $k(i, s'h)$ est non-nul, et donc $k(i, sh) = 0$, puisque $k(i, sh) \times k(i, s'h) = 0$. Notons qu'on aurait pu aussi démontrer le résultat précédent à l'aide du résultat du § 2.3 a).

Les termes non-nuls et égaux à 1, dans le dernier membre de (2) étant tels que $m(h') = s'h'$ et $m(h) \neq sh$, on en déduit que $L(sh, s'h') = p^{q-2}(p-1)$.

$$\begin{aligned} \text{Comme } L(sh, s'h') + R(sh, s'h') &= \Sigma\{k(i, s'h') \mid i \in I\} \\ &= k(s'h') = p^{q-1} \end{aligned}$$

$$\text{on a : } R(sh, s'h') = p^{q-2}$$

$$d) S = \{U, C, A, G\}; p = 4; q = 3$$

TRP se réduit à l'élément m tel que : $m(1) = U; m(2) = G; m(3) = G$, ce que nous écrirons simplement TRP = $\{U_1 G_2 G_3\}$ ou $\{UGG\}$.

Pour ILE, il y a trois mots, puisque dans chacun des trois blocs de colonnes le total de la ligne ILE est 3. Ces trois mots commencent par A1, U2 et ne diffèrent que par leur 3^o signe : on a ILE = $\{AUU, AUC, AUA\}$.

De même TAU comprend trois mots : tous commencent par U1 ; deux se continuent par A2 et un par G2 ; il n'y a que deux possibilités pour la troisième lettre : A3 (deux fois) et G3 (une fois) ; d'où TAU = $\{UAA, UAG, UGA\}$

Les colonnes G1 et U2 possèdent la propriété P.

2.4 a) Posant $a = \varphi^{s'h}$ ($s' \neq s$) et $b = \varphi^{sh}$, φ^J sera de moyenne nulle et variance 1 si :

$$(p-1)a + b = 0 \quad (3)$$

$$((p-1)a^2 + b^2)/(pq) = 1$$

d'où l'on déduit : $a = (q/(p-1))^{1/2}$; $b = -((p-1)q)^{1/2}$

$$b) \text{ On a : } k(i)\psi^i = a\Sigma\{k(i, s'h) \mid s' \in S - \{s\}\} + bk(i, sh)$$

$$1^\circ \text{ cas : } k(i, sh) = 0$$

$$k(i)\psi^i = a\Sigma\{k(m, s'h) \mid s' \in S - \{s\}, m \in i\}$$

Comme $m(h) \neq sh$, $\forall m \in i$, pour m fixé, quand s' décrit $S - \{s\}$, $k(m, s'h)$ prend une fois et une seule la valeur 1, quand $s'h = m(h)$, sinon il est nul. On a donc :

$$k(i)\psi^i = a \text{Card } i.$$

$$\text{Comme : } k(i) = \Sigma\{k(i, j) \mid j \in J\} = \Sigma\{k(m, j) \mid m \in i, j \in J\}$$

$$= \Sigma\{k(m) \mid m \in i\} = q \text{Card } i$$

$$\text{on a , si } k(i, sh) = 0 : \psi^i = a/q$$

$$2^\circ \text{ cas : } k(i, sh) \neq 0$$

sh possédant la propriété P, on a $\forall s' \neq s : k(i, s'h) = 0$,

$$\text{et donc : } k(i)\psi^i = bk(i, sh) = b\Sigma\{k(m, sh) \mid m \in i\}$$

or comme $k(i, s'h) = 0$, $\forall s' \neq s$, $k(m, s'h) = 0$, $\forall m \in i$, $\forall s' \neq s$, ce qui implique que $m(h) = sh$, $\forall m \in i$, d'où :

$$k(i)\psi^i = b \text{Card } i$$

$$\text{et donc , si } k(i, sh) \neq 0 : \psi^i = b/q$$

c) On a :

$$k(j) \theta^j = (a/q) \Sigma \{k(i,j) | i \in I ; k(i,sh) = 0\} + \\ (b/q) \Sigma \{k(i,j) | i \in I ; k(i,sh) \neq 0\}$$

1° cas : $j = s'h'$, $h' \neq h$.

On a alors :

$$k(s'h') \theta^{s'h'} = (a/q) L(sh, s'h') + (b/q) R(sh, s'h') \\ = (p^{q-2}/q) (a(p-1) + b) = 0 \text{ d'après (3)}$$

$$\text{d'où } \theta^{s'h'} = \varphi^{s'h'} = 0 \text{ si } h' \neq h.$$

2° cas : $j = sh$;

$$k(sh) \theta^{sh} = (b/q) \Sigma \{k(i,sh) | i \in I ; k(i,sh) \neq 0\} \\ = (b/q) \Sigma \{k(i,sh) | i \in I\} = k(sh) b/q ;$$

$$\text{d'où } \theta^{sh} = b/q = \varphi^{sh}/q$$

3° cas : $j = s'h \neq sh$

Comme $k(i,sh) \neq 0$ implique que, $\forall s' \neq s$, $k(i,s'h) = 0$, on a :

$$k(s'h) \theta^{s'h} = (a/q) \Sigma \{k(i,s'h) | i \in I ; k(i,sh) = 0\} \\ = (a/q) \Sigma \{k(i,s'h) | i \in I\} = k(s'h) a/q ;$$

$$\text{d'où } \theta^{s'h} = a/q = \varphi^{s'h}/q$$

d) L'on déduit de 2.4c que φ^J est facteur de variance 1 de k_{IJ}^J , et associé à la valeur propre $\lambda_\alpha = 1/q$, puisque $\theta^J = \varphi^J \circ f_{J^I}^I \circ f_I^J = (1/q) \varphi^J$.

Le facteur G_α^J de variance λ_α , associé, est donc tel que :

$$\forall s' \in S, \forall h' \neq h : G_\alpha(s'h') = 0 ;$$

$$\forall s' \neq s : G_\alpha(s'h) = a q^{-1/2} = (p-1)^{-1/2} ;$$

$$G_\alpha(sh) = b q^{-1/2} = -(p-1)^{1/2} ;$$

Le facteur $F_{\alpha I}^I$, de variance λ_α , associé à G_α^J étant égal à $\varphi^J \circ f_I^I$ est identique à ψ^I . On a donc :

$$\text{si } k(i,sh) = 0 : F_\alpha(i) = a/q = ((p-1)q)^{-1/2} ;$$

$$\text{si } k(i,sh) \neq 0 : F_\alpha(i) = b/q = -((p-1)/q)^{1/2} ;$$

$$2.5 \text{ a) } \pi_j^{j'} = \Sigma \{f_j^m f_m^{j'} | m \in M\}$$

$$= \Sigma \{k(m,j) k(m,j') / k(m) k(j') | m \in M\}$$

$$= N(j, j') / (qp^{q-1})$$

$$= p^{q-2} / (qp^{q-1}) = (1/pq) \text{ si } j = sh, j' = s'h', h' \neq h.$$

$$= 1/q \text{ si } j = j'$$

$$= 0 \text{ si } j = sh, j' = s'h, s' \neq s.$$

Dans le cas où $p = 4$, $q = 3$, on a donc :

$$\begin{aligned}\pi_{j'}^{j'} &= 1/12 \text{ si } j = sh, j' = s'h', h' \neq h \\ \pi_{j'}^{j'} &= 1/3 \\ \pi_{s'h'}^{sh} &= 0 \text{ si } s \neq s'\end{aligned}$$

$$\begin{aligned}b) \varphi^J \circ \pi_J^{sh} &= \Sigma\{\varphi^{s'h'} \pi_{s'h'}^{sh} \mid s' \in S\} + & (4) \\ & \Sigma\{\Sigma\{\varphi^{s'h'} \pi_{s'h'}^{sh} \mid s' \in S\} \mid h' \in \{1, \dots, q\} - \{h\}\} \\ &= (1/q) \varphi^{sh} + (1/pq) \Sigma\{\Sigma\{\varphi^{s'h'} \mid s' \in S\} \mid h' \in \{1, \dots, q\} - \{h\}\} \\ &= (1/q) \varphi^{sh}\end{aligned}$$

$$d'où \quad \varphi^J \circ \Pi_J^J = (1/q) \varphi^J$$

φ^J est donc facteur de k_{MJ} associé à la valeur propre $1/q$, et donc si φ^J est de variance 1, $\varphi^J \circ f_J^M$ est de variance la valeur propre, i.e. $(1/q)$.

c) L'on déduit de (4), qui a été établi sans faire d'hypothèse sur φ^J , que φ^J est vecteur propre de k_{MJ} associé à la valeur propre λ si :

$$(1/(pq)) \Sigma\{\varphi(s'h') \mid s' \in S; h' = 1, \dots, q; h' \neq h\} = (\lambda - 1/q) \varphi^{sh}$$

d'où l'on déduit que si $\lambda \neq 1/q$, φ^{sh} ne dépend que de h .

Posant $\varphi^{sh} = a_h$, les a_h doivent vérifier l'équation :

$$(1/q) \Sigma\{a_h \mid h' = 1, \dots, q\} = \lambda a_h$$

On retrouve ainsi le facteur trivial ($a_h = 1, \forall h = 1, \dots, q$) constant et égal à 1, associé à la valeur propre 1, et un espace propre de dimension $q-1$, relatif à la valeur propre 0, et engendré par les fonctions de moyenne nulle sur J , et constantes sur chaque S_h , résultat qui ne saurait nous étonner, puisque l'on a vu que k_{MJ} peut être considéré comme un tableau disjonctif complet (cf remarque du 2.3 b)).

Les facteurs non triviaux de k_{MJ} sont donc engendrés par des facteurs de moyenne nulle sur chaque S_h (résultat évident, k_{MJ} étant un tableau disjonctif complet). Comme ces facteurs sont associés à la même valeur propre $\lambda = 1/q$, ils engendrent un sous-espace propre de dimension $\text{Card } J - q = q(p-1)$.

On retrouve bien ainsi tous les facteurs de k_{MJ} .

d) On a :

$$\begin{aligned}k(i) \psi^i &= \Sigma\{k(i, j) \varphi^j \mid j \in J\} = \Sigma\{k(m, j) \varphi^j \mid m \in i, j \in J\} \\ &= \Sigma\{k(m) \Sigma\{\varphi^j k(m, j) / k(m) \mid j \in J\} \mid m \in i\} \\ &= \Sigma\{k(m) w^m \mid m \in i\}\end{aligned}$$

d'où puisque $k(m)/k(i) = 1/\text{Card } i$;

$$\psi^i = \Sigma\{w^m \mid m \in i\} / \text{Card } i$$

La variance de ψ^I étant la variance interclasse de W^M , associé à la partition I de M, est plus petite que la variance de W^M .

Si φ^J est le facteur de variance 1 de k_{IJ} associé à la plus grande valeur propre λ_1 issue de cette correspondance, on a :

$$\text{variance } (\psi^I) = \lambda_1 \leq \text{variance } (W^M)$$

Or la variance de W^M est inférieure ou égale à la plus grande valeur propre issue de la correspondance k_{MJ} , valeur propre qui est égale à $1/q$. On en déduit que toutes les valeurs propres issues de k_{IJ} sont inférieures ou égales à $1/q$, valeur limite qui est atteinte, s'il existe une colonne sh de k_{IJ} possédant la propriété P, le facteur correspondant ayant été défini en 2.4).*

Remarque : Soit plus généralement S' une partie de S ; on dit que le bloc de colonne S'h de k_{IJ} possède la propriété P si $\forall i \in I, \forall s' \in S', \forall s'' \in S-S' : k(i, s'h) \times k(i, s''h) = 0$; alors on a aussi un facteur φ , nul en dehors de Sh, constant sur S'h et sur Sh-S'h, relatif à la valeur propre $1/q$; et tous les facteurs de k_{IJ} relatifs à la v. p. ($1/q$) sont de cette forme.

2.6 a) Les colonnes G1 et U2 possédant la propriété P (cf 2.3 d)), on en déduit d'après 2.4 d) deux facteurs sur J de k_{IJ} , relatifs à la valeur propre $1/3$, facteurs orthogonaux puisque le premier a pour support S1 et le second pour support S2. D'après 2.5 d) ces facteurs sont associés à la plus grande valeur propre de k_{IJ} , puisque les valeurs propres issues de k_{IJ} sont inférieures ou égales à $1/3$.

Nous désignerons par G_1^J le facteur sur J, de support S1, facteur déjà calculé en 2.2c) et par G_2^J le facteur sur J de support S2, que l'on peut calculer à partir de 2.4 d). Ces facteurs ainsi que les facteurs F_1^I et F_2^I associés sur I s'écrivent :

$$F_1^I(i) = -1 \text{ si } i \in \{\text{VAL, ALA, GLY, ASP, GLU}\} \\ = 1/3 \text{ sinon}$$

$$G_1^J = (1/\sqrt{3}) (1 ; 1 ; 1 ; -3 ; 0 ; 0 ; 0 ; 0 ; 0 ; 0 ; 0 ; 0)$$

$$F_2^I(i) = -1 \text{ si } i \in \{\text{ILE, PHE, LEU, VAL, MET}\} \\ = 1/3 \text{ sinon}$$

$$G_2^J = (1/\sqrt{3}) (0 ; 0 ; 0 ; 0 ; -3 ; 1 ; 1 ; 1 ; 0 ; 0 ; 0 ; 0)$$

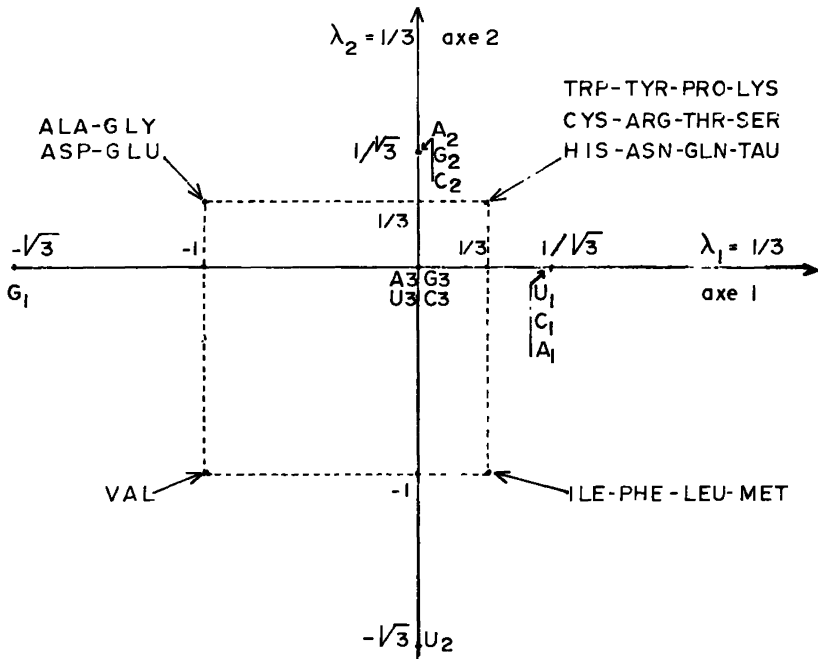
d'où la représentation dans le plan des axes 1-2 de I et J, donnée ci-dessous.

b) Les colonnes du tableau k_{IJ} vérifient les mêmes relations que celles du tableau k_{MJ} , à savoir que :

$$\Sigma\{k(i, j) | j \in S1\} = \Sigma\{k(i, j) | j \in S2\} = \Sigma\{k(i, j) | j \in S3\} = k(i)/3$$

* I étant une partition de M, la valeur propre de rang α issue de k_{IJ} est plus petite ou égale à la valeur propre de rang α issue de k_{MJ} (cf [Partition et v. propre] à paraître dans les Cahiers), ce qui constitue une autre démonstration du fait que les valeurs propres de k_{IJ} sont plus petites ou égales à $1/q$.

ce qui implique que tout facteur sur J non trivial est de moyenne nulle sur chaque Sh. Il y a donc au plus $\text{Card} J - q = q(p-1) = 9$ facteurs non triviaux pour k_{IJ} .



Comme les colonnes U_3 et C_3 sont identiques, il y a en fait au plus 8 facteurs non triviaux pour k_{IJ} .