



Principes d'analyse matricielle

M. J. Ville

► To cite this version:

M. J. Ville. Principes d'analyse matricielle. Annales de l'ISUP, 1955, IV (2-3), pp.141-218. hal-04095248

HAL Id: hal-04095248

<https://hal.science/hal-04095248>

Submitted on 11 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

PRINCIPES D'ANALYSE MATRICIELLE

par

M. J. VILLE

Professeur

CHAPITRE PREMIER

INTRODUCTION ALGÈBRIQUE

1.- NOTION DE NOMBRE

Aufur et à mesure du développement des mathématiques, on a observé une évolution de la notion de nombre, dans un sens de généralisation croissante.

Au début, nous trouvons les nombres dits "naturels", qui sont un, deux, trois, etc ... sans le zéro. Pour de pareils nombres, on sait définir l'addition et la multiplication, opérations toujours possibles. On sait définir aussi la soustraction et la division, avec ou sans reste, opérations qui ne sont pas toujours possibles. Etant donnés deux nombres a et b , il existe donc toujours des nombres x et y tels que :

$$a + b = x$$

$$ab = y$$

mais il n'existe pas toujours un nombre z tel que

$$a = b + z$$

Si un tel nombre existe, on a $a > b$, et on écrit conventionnellement

$$z = a - b.$$

Si $a > b$, il existe toujours un nombre q tel que

$$bq < a < b(q + 1)$$

ce nombre, qui est le quotient avec reste, n'est appelé quotient exact que si

$$a = bq$$

et on écrit alors conventionnellement

$$q = \frac{a}{b}.$$

2.- NOMBRES ENTIERS

Les généralisations de la notion de nombre sont intervenues pour rendre possibles des opérations qui ne l'étaient pas. Dans l'algèbre des entiers, qui sont les nombres formant la suite, infinie dans les deux sens :

$$\dots -2, -1, 0, 1, 2, \dots$$

on sait définir l'addition, la multiplication; la soustraction est toujours possible, puisque, quels que soient a et b , il existe toujours un nombre x tel que

$$a = b + x$$

L'écriture

$$x = a - b$$

a donc toujours un sens.

La division sans reste continue à ne pas être toujours possible; ce n'est qu'exceptionnellement que à a et b on peut associer un nombre q tel que

$$a = bq$$

3.- NOMBRES RATIONNELS

Pour donner à l'écriture $\frac{a}{b}$ un sens quels que soient a et b , on passe des entiers aux nombres rationnels, ou fractionnaires. $\frac{a}{b}$ a alors toujours un sens, sauf si $b = 0$. Les quatre opérations sont alors toujours possibles.

Les opérations impossibles dans l'algèbre des nombres rationnels sont les résolutions d'équations algébriques. Par exemple, l'équation

$$x^2 = 2$$

n'est satisfaite par aucun nombre rationnel, et le symbole $\sqrt{2}$ est donc dépourvu de sens dans l'algèbre des nombres rationnels. Seules ont un sens ses valeurs approchées.

4.- NOMBRES RÉELS

On est alors conduit à faire une théorie générale des valeurs approchées, qui peut se résumer comme suit :

Savoir calculer une valeur approchée d'un symbole, c'est savoir déterminer, pour un nombre rationnel quelconque, s'il est valeur approchée par défaut, ou par excès. C'est donc savoir associer à tout nombre rationnel r , sauf peut-être un, une valeur $f(r)$, égale à 0 si r est une valeur approchée par défaut, à 1 si r est une valeur approchée par excès. $f(r)$ est ainsi fonction, de valeur égale à 0 ou à 1, telle que si $r' < r''$ on a forcément

$f(r') < f(r'')$, puisque toute valeur plus petite qu'une valeur approchée par défaut est elle-même par défaut, et toute valeur plus grande qu'une valeur approchée par excès est elle-même par excès. $f(r)$ doit être définie pour tous les nombres rationnels sauf peut-être un parce que si le symbole définit un nombre rationnel r_0 , r_0 lui-même n'est approché ni par défaut, ni par excès, puisque c'est la valeur exacte. Si $f(r)$ est définie pour tous les nombres rationnels, le symbole dont $f(r)$ définit les valeurs approchées ne peut être un nombre rationnel. On dit alors qu'il représente un nombre irrationnel. Les nombres rationnels et irrationnels, dont l'ensemble constitue les nombres réels, ne sont donc autres du point de vue théorique que des séparations entre valeurs approchées par défaut et par excès.

5.- NOMBRES COMPLEXES

Les nombres réels permettent les quatre opérations fondamentales; ils permettent l'extraction d'une racine carrée, ou d'une racine d'ordre multiple. Dans l'algèbre des nombres réels, toute équation de la forme :

$$f(x) = 0$$

où $f(x)$ est une fonction continue, admet une racine dès que l'on a pu trouver deux nombres a et b tels que $f(a) \cdot f(b) < 0$. En particulier, toute équation algébrique de degré impair admet au moins une racine réelle. Mais il existe des équations très simples, telles que

$$x^2 + 1 = 0$$

qui n'admettent pas de racine. On est amené alors à introduire pour la généralité des énoncés la notion de nombre complexe, c'est-à-dire de nombres $x + yi$, sur lesquels les opérations sont les mêmes que sur les nombres réels, avec la restriction que $i^2 = -1$. Dans ces conditions, toute équation algébrique de degré n admet exactement n racines.

6.- NOMBRES HYPERCOMPLEXES

La généralisation se poursuit maintenant en observant qu'un nombre complexe $x + yi$ avec la règle

$$i^2 = -1$$

apparaît comme un cas particulier d'une expression

$$q = x_1 i_1 + x_2 i_2 + \dots + x_n i_n$$

$x_1, \dots, x_n$ étant des nombres réels, et les i_m étant des symboles obéissant à des règles d'opération de la forme

$$i_s i_t = \sum_{(m)} \alpha_{st}^{(m)} i_m \quad \left(\alpha_{st}^{(m)} \text{ réels} \right)$$

De telles expressions sont appelées nombres hypercomplexes. On peut montrer alors que si on considère les deux règles :

R_1 Commutativité de la multiplication :

$$xy = yx$$

R_2 Inexistence de diviseurs de zéro :

$$\text{Si } xy = 0, \text{ alors } x = 0 \text{ ou } y = 0$$

prises à part parmi les règles bien connues de l'algèbre des nombres complexes ordinaires, on obtient, selon la ou les règles auxquelles on renonce, les généralisations qui suivent.

7.- QUATERNIONS

Si on renonce à R_1 et à R_2 seulement, on montre que les nombres hypercomplexes ont forcément quatre unités hypercomplexes, c'est-à-dire sont de la forme :

$$q = x_1 i_1 + x_2 i_2 + x_3 i_3 + x_4 i_4$$

et que l'on peut, par des changements d'unités convenables n'affectant pas la généralité, les mettre sous la forme :

$$q = a + xi + yj + zk$$

a, x, y, z , étant réels, i, j, k obéissant aux règles

$$ijk = i^2 = j^2 = k^2 = -1$$

Ces nombres sont les quaternions de Hamilton; leur emploi est assez limité.

Dans les opérations sur quaternions, on a en général : $q_1 q_2 \neq q_2 q_1$ mais si

$$q_1 q_2 = 0$$

on a forcément soit $q_1 = 0$, soit $q_2 = 0$.

8.- MATRICES

Pour augmenter au delà de quatre le nombre des unités hypercomplexes, il faut renoncer à R_2 . On montre alors que les unités hypercomplexes peuvent se mettre sous la forme i_{st} , c'est-à-dire qu'on peut les affecter de deux indices, de manière que les règles d'opération soient :

$$(4) \quad i_{st} i_{uv} = \delta_{tu} \cdot i_{sv}$$

δ_{tu} étant l'indice de Kronecker, c'est-à-dire un nombre égal à 1 si $t = u$, à 0, si $t \neq u$. Un nombre hypercomplexe est alors une expression de la forme

$$(5) \quad q = \sum_{st} a_{st} i_{st}$$

On voit que si on procède au produit de (5) et d'un nombre

$$q' = \sum_{st} c_{st} i_{st}, \text{ on obtient } qq' = \sum_{st} c_{st} i_{st}$$

avec

$$c_{st} = \sum_h a_{sh} b_{ht}$$

Au lieu d'écrire q sous la forme (5), on peut l'écrire sous la forme

$$q = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

qui se prête mieux aux applications. Un tel tableau est appelé une matrice; l'algèbre hypercomplexe la plus générale (à un nombre fini d'unités hyper-complexes) est ainsi l'algèbre de ces tableaux, que l'on appelle l'algèbre matricielle. Lorsque l'on traite de l'algèbre matricielle directement sur des expressions telles que (5), avec les règles (4), on parle alors d'algèbre matricielle totale.

NOTATIONS VECTORIELLES ET MATRICIELLES

Nous aurons à manipuler des symboles mathématiques d'abord en tableaux rectangulaires, le plus souvent carrés. Les symboles d'un tableau rectangulaire à m lignes et n colonnes seront affectés de deux indices, le premier indiquant le rang de la ligne, le second le rang de la colonne dans laquelle figure le symbole considéré. Le tableau lui-même sera appelé une matrice à m lignes et n colonnes, ou, plus brièvement, une matrice (m, n) . Un pareil tableau sera encadré de crochets, conformément au schéma suivant :

$$(1) \quad \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix}$$

Un pareil tableau sera souvent représenté par une seule lettre majuscule; pour rappeler la nature des éléments qui y figurent, nous écrirons l'élément général entre deux crochets, comme suit :

$$(2) \quad A = [a_{ij}] \quad (m, n)$$

Les deux indices figurent en sus à droite rappelleant qu'il s'agit de m lignes et n colonnes. Quand aucune confusion ne sera possible, ces deux indices seront omis.

Il nous arrive souvent de considérer des tableaux réduits à une seule ligne ou à une seule colonne, tels que les tableaux suivants :

$$(3) \quad X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \quad U = [u_1 \ u_2 \ \dots \ u_n]$$

Dans ce cas nous affecterons chaque élément d'un seul indice, nous écrirons alors les notations abrégées :

$$X = [x_i] \quad U = [u_j]$$

Nous aurons à considérer des sommes et des produits de tableaux, la justification des définitions apparaissant par la suite.

La somme de deux tableaux A et B n'a de sens que si ces tableaux ont même nombre de lignes et même nombre de colonnes; c'est le tableau obtenu

CHAPITRE II

NOTATIONS VECTORIELLES ET MATRICIELLES

Nous aurons à manipuler des symboles mathématiques disposés en tableaux rectangulaires, le plus souvent carrés. Les symboles d'un tableau rectangulaires à m lignes et n colonnes seront affectés de deux indices, le premier indiquant le rang de la ligne, le second le rang de la colonne dans laquelle figure le symbole considéré. Le tableau lui-même sera appelé une matrice à m lignes et n colonnes, ou, plus brièvement, une matrice (m, n) . Un pareil tableau sera encadré de crochets, conformément au schéma suivant :

$$(1) \quad \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

Un pareil tableau sera souvent représenté par une seule lettre majuscule; pour rappeler la nature des éléments qui y figurent, nous écrivons l'élément général entre deux crochets, comme suit :

$$(2) \quad A = \begin{bmatrix} a_{ij} \end{bmatrix} (m, n)$$

Les deux indices figurant en bas à droite rappellent qu'il s'agit de m lignes et n colonnes. Quand aucune confusion ne sera possible, ces deux indices seront omis.

Il nous arrivera souvent de considérer des tableaux réduits à une seule ligne ou à une seule colonne, tels que les tableaux suivants :

$$(3) \quad X = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_m \end{bmatrix} \quad U = [u_1 \ u_2 \ \dots \ u_n]$$

Dans ce cas nous affecterons chaque élément d'un seul indice. Nous utiliserons alors les notations abrégées :

$$X = [x_i] \quad U = \{u_i\}$$

Nous aurons à considérer des sommes et des produits de tableaux, la justification des définitions apparaissant par la suite.

La somme de deux tableaux A et B n'a de sens que si ces tableaux ont même nombre de lignes et même nombre de colonnes; c'est le tableau obtenu

en faisant la somme des éléments occupant les mêmes emplacements dans A et B. Par conséquent, si

$$A = [a_{ij}]_{(m, n)} \quad B = [b_{ij}]_{(m, n)} \quad C = [c_{ij}]_{(m, n)}$$

l'équation

$$(4) \quad C = A + B$$

signifie que

$$c_{ij} = a_{ij} + b_{ij}$$

pour toutes les valeurs de indices i et j .

Le produit de deux tableaux A et B pris dans cet ordre n'a de sens que si le nombre de colonnes de A est égal au nombre de lignes de B.

Cette condition est facile à se mettre en mémoire si l'on observe qu'en écrivant que A est une matrice (m, n) et B une matrice (n, p) , on voit apparaître la suite

$$(m, n) \times (n, p)$$

où le second nombre de la première parenthèse est égal au premier nombre de la seconde. Cette condition préliminaire étant supposée respectée, le produit AB est une matrice (m, p) dont l'élément général c_{ij} a pour expression

$$(6) \quad c_{ij} = \sum_{h=1}^n a_{ih} b_{hj}$$

La formule (6) est tout à fait fondamentale, et doit être tout à fait familière. On observera que c_{ij} s'obtient en faisant la somme de tous les produits $a_{ih} b_{hj}$, le premier indice de a_{ih} étant le premier indice de c_{ij} , le second indice de b_{hj} étant le second indice de c_{ij} , le second indice de a_{ih} étant le premier indice de b_{hj} ; c'est par rapport à ce troisième indice que nous opérons la sommation. On observera également que la formule (6) suppose essentiellement que le second indice de a_{ij} peut prendre autant de valeurs que le premier indice de b_{ij} , ce qui revient à dire que A a autant de colonnes que B a de lignes. On dit quelquefois que c_{ij} est le produit de la i ème ligne de A par la j ème colonne de B.

Le produit de deux matrices dépend essentiellement de leur ordre. AB est en général différent de BA; il se peut que AB existe et que BA n'existe pas.

Il est bon de se familiariser avec les produits de matrices particulières. Soient par exemple les deux matrices :

$$(7) \quad X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = [x_i] \quad U = [u_1 \ u_2 \ \dots \ u_n] = \{u_i\}$$

qui sont des matrices $(n, 1)$ et $(1, n)$.

UX et XU ont toutes deux un sens, mais sont tout à fait différentes, UX est une matrice (1, 1), réduite par conséquent à un seul élément :

$$(8) \quad UX = u_1 x_1 + u_2 x_2 + \dots + u_n x_n$$

Au contraire, XU est une matrice (n, n), donc une matrice carrée

$$(9) \quad \begin{bmatrix} x_1 u_1 & x_1 u_2 & x_1 u_n \\ x_2 u_1 & x_2 u_2 & x_2 u_n \\ x_n u_1 & x_n u_2 & x_n u_n \end{bmatrix}$$

Nous voyons que les conventions précédentes nous obligent à certaines précautions lorsque nous avons à faire l'opération pourtant simple de multiplier tous les éléments d'une matrice par un même nombre. Soit en effet λ ce nombre. En ce qui concerne les matrices (7), nous pourrions considérer λ comme une matrice (1, 1), ce qui nous conduira à :

$$(10) \quad X\lambda = \begin{bmatrix} x_1 \lambda \\ x_2 \lambda \\ x_n \lambda \end{bmatrix} \quad \lambda U = \left\{ \lambda u_1, \lambda u_2, \dots, \lambda u_n \right\}$$

On observera l'emplacement du facteur λ , qui se met à droite de X, et à gauche de U. Cette convention n'est pas essentielle; on ne doit s'y astreindre que dans les recherches tout à fait théoriques.

En ce qui concerne une matrice (m, n), il est impossible de multiplier tous les éléments de cette matrice par λ en respectant les règles posées ci-dessus. Pour pouvoir faire cette opération correctement, il faut faire intervenir l'indice de Kronecker, qui est un nombre δ_{ij} , à deux indices, qui a pour valeur 0 lorsque les indices sont différents, et 1 lorsque les indices sont égaux. Une matrice dont les éléments sont des indices de Kronecker sera dite une matrice unité, et représentée par la lettre E.

Une matrice rectangulaire unité sera donc telle que

$$E = \begin{bmatrix} \delta_{ij} \end{bmatrix} \quad (m, n)$$

Une telle matrice jouit, lorsqu'elle est carrée, de la propriété de laisser toute matrice inchangée par multiplication. Soit en effet, une matrice A (m, n); formons le produit

$$B (m, n) = A (m, n) E (n, n)$$

Nous aurons

$$b_{ij} = \sum_h a_{ih} \delta_{hj}$$

Dans la sommation par rapport à h, seul intervient le δ_{hj} à indices égaux, donc δ_{jj} ; il en résulte que

$$b_{ij} = a_{ih}$$

On arriverait à la même conclusion en considérant le produit :

$$E (m, m) A (m, n)$$

On voit facilement que pour changer $A = a_{ij}$ en λa_{ij} , il faut multiplier A par la matrice $\begin{bmatrix} \lambda \delta_{ij} \end{bmatrix}$, à droite ou à gauche indifféremment. Quand nous écrivons donc

$$\lambda A = A\lambda = \begin{bmatrix} \lambda a_{ij} \end{bmatrix} (m, n)$$

il ne faudra pas perdre de vue que dans ces équations, λ est une notation abrégée pour la matrice $\begin{bmatrix} \lambda \delta_{ij} \end{bmatrix} (m, m)$ ou $\begin{bmatrix} \lambda \delta_{ij} \end{bmatrix} (n, n)$, selon que λ est à gauche ou à droite de A .

Si cette convention est bien comprise, la notation abrégée ne présente aucun inconvénient.

CHAPITRE III

RÉDUCTION DES MATRICES - RANG

1.- RAPPEL DES RÈGLES D'OPÉRATIONS RELATIVES AUX MATRICES

Nous rappelons qu'étant données deux matrices (m, n) , soient A et B , on sait définir leur somme $A + B$, et qu'étant données deux matrices, (m, n) (n, p) , on sait définir leur produit AB , qui est une matrice (m, p) . On sait par ailleurs définir le produit d'une matrice par un scalaire.

On vérifie sans peine les règles suivantes :

$$(1) \quad \begin{cases} \lambda (A + B) &= \lambda A + \lambda B \\ C (A + B) &= C A + C B \\ (A + B) C &= A C + B C \\ (A B) C &= A (B C) \end{cases}$$

Nous retrouverons les règles usuelles de l'algèbre, sauf la commutativité du produit.

On peut se demander quels sont les éléments qui jouent le rôle du zéro et de l'unité. Une matrice N (m, n) sera dite nulle si, quelle que soit A (m, n) , on a

$$(2) \quad A + N = N + A = A$$

Une telle matrice N a forcément tous ses éléments égaux à zéro.

De même, nous définirons une matrice unité E par la condition que multipliée par A , elle donne encore A comme résultat de la multiplication. Si E est une matrice (m, n) , nous devons avoir, quelle que soit la matrice A (n, p)

$$(3) \quad E A = A$$

Mais le produit EA est une matrice (m, p) . Il ne peut être égal à A que si $m = n$. Une matrice unité est donc forcément une matrice carrée. L'équation (3), si nous posons

$$E = \begin{bmatrix} e_{ij} \end{bmatrix}_{n, n} \quad A = \begin{bmatrix} a_{ij} \end{bmatrix}_{m, n}$$

s'écrit

$$(4) \quad a_{ij} = \sum_{h=1}^n e_{ih} a_{hj}$$

Les équations (4) devant avoir lieu identiquement par rapport aux a_{ij} , nous en déduisons

$$(5) \quad e_{ij} = \delta_{ij} \quad (\text{indice de Kronecker})$$

Quelles que soient les matrices A et B, nous aurons alors :

$$EA = A$$

$$BE = B$$

L'introduction de la matrice unité (1) va nous permettre de définir l'inverse d'une matrice A, comme étant une matrice qui, multipliée par A, donne une matrice unité.

2.- PROPRIÉTÉS GÉNÉRALES DES MATRICES INVERSES

Soit donnée une matrice A (m, n). Comme le produit n'est pas commutatif, nous serons amenés à distinguer une inverse droite B et une inverse gauche C telles que

$$(6) \quad AB = E(m, m) \quad CA = E(n, n)$$

Supposons que B et C existent toutes deux, B ne peut être qu'une matrice (n, m), C une matrice (n, m). D'après la propriété d'associativité du produit nous voyons que

$$(CA)B = C(AB)$$

En comparant à (6), nous obtenons

$$E(n, n)B(n, m) = C(n, m)E(m, m)$$

d'où

$$B = C.$$

Il apparaît ainsi que lorsqu'une matrice admet une inverse droite et une inverse gauche, ces deux inverses sont identiques. Il en résulte évidemment que la matrice inverse est unique.

Nous donnerons une place particulière à ces matrices qui admettent une inverse droite et une inverse gauche, cette inverse étant d'ailleurs la même et unique; nous les appellerons matrices non singulières. Une matrice singulière, conformément à cette définition :

ou bien n'admettra pas d'inverse droite

ou bien n'admettra pas d'inverse gauche

ou bien n'admettra pas d'inverse du tout.

L'inverse d'une matrice non singulière A sera représentée par la notation A^{-1} .

Supposons que A_1 et A_2 admettent chacune une inverse droite, soient B_1 et B_2 ces inverses. Nous constatons que

$$A_1 A_2 B_2 B_1 = E$$

Donc le produit $A_1 A_2$ admet comme inverse droite le produit $B_2 B_1$.

Si A_1 et A_2 admettaient les inverses gauches C_1 et C_2 , le produit $A_1 A_2$ admettrait $C_2 C_1$ comme inverse gauche.

(1) Nous avons déjà vu que toute matrice carrée de la forme (5) laissait invariante toute autre matrice par multiplication; il restait à montrer, ce que nous avons fait, la réciproque.

D'une manière générale, on voit que si $A_1, A_2, \dots, A_n$ admettent les inverses droites $B_1, B_2, \dots, B_n$, le produit $A_1 A_2 \dots A_n$ admettra l'inverse droite $B_n B_{n-1} \dots B_1$; une proposition analogue aura lieu pour les inverses gauches.

Il résulte de ce qui précède que la condition nécessaire et suffisante pour que le produit de n matrices $A_1, A_2, \dots, A_n$ soit une matrice singulière est que l'une au moins des matrices $A_1, A_2, \dots, A_n$ soit singulière. Cette proposition est la correspondante de la proposition : La condition nécessaire et suffisante pour qu'un produit de nombres soit nul est que l'un au moins des facteurs du produit soit nul.

La proposition : "Si un produit de matrices est nul, l'une au moins des matrices constituant le produit est nulle", est une proposition fausse.

Nous allons maintenant démontrer la proposition fondamentale :

Une matrice non singulière est forcément une matrice carrée. La démonstration sera basée sur la remarque évidente : si on multiplie une matrice A par une matrice B non singulière, on obtient une matrice AB qui est de même nature que A , c'est-à-dire singulière si A est singulière, non singulière si A est non singulière. Cette remarque s'applique également à BA .

3.- OPÉRATIONS ÉLÉMENTAIRES SUR LES MATRICES

Etant donnée une matrice A , on appelle opérations élémentaires les opérations qui consistent à :

- Multiplier une ligne ou une colonne de A par un scalaire non nul.
- Intervertir deux lignes ou deux colonnes.
- Ajouter aux éléments d'une ligne (ou d'une colonne) les éléments d'une autre ligne (ou d'une autre colonne) multipliés par un même scalaire.

Nous allons montrer que chacune de ces opérations revient à multiplier A par une matrice d'un type particulier, non singulière, que nous appellerons matrice élémentaire.

La matrice déduite de A en y remplaçant a_{ji} par λa_{ji} est le produit de A par la matrice

$$\begin{bmatrix} \lambda & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

à droite

La matrice $\begin{bmatrix} \lambda \end{bmatrix}$ admet l'inverse

$$\begin{bmatrix} 1/\lambda & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Soit maintenant $\bar{A}$ la matrice déduite de A en y intervertissant les lignes de rang s et t , de sorte que

$$A = [a_{ij}] ; \bar{A} = [\bar{a}_{ij}] ; \bar{a}_{ij} = a_{ij} \quad (i \neq s, t),$$

$$\bar{a}_{sj} = a_{tj} \quad \bar{a}_{tj} = a_{sj}.$$

Mettons $\bar{A}$ sous la forme

$$\bar{A} = X A$$

X étant une matrice $[x_{ij}]$ à déterminer. Nous aurons

$$\bar{a}_{ij} = \sum_h x_{ih} a_{hj}$$

Pour $i \neq s, t$, nous aurons ainsi $x_{ih} = \delta_{ih}$ (indice de Kronecker).

Pour $i = s$, nous aurons

$$x_{sh} = 0 \quad \text{pour } h \neq t$$

$$x_{st} = 1$$

et pour $i = t$

$$x_{th} = 0 \quad \text{pour } h \neq s$$

$$x_{ts} = 1$$

La matrice X est ainsi déterminée par les conditions suivantes :

Dans chaque ligne figure un élément non nul et un seul. Cet élément a pour valeur 1. Il a pour rang, d'une manière générale i dans la i° ligne, à l'exception des deux lignes de rang s et t .

Dans la ligne de rang s , il a pour rang t , et dans la ligne de rang t , il a pour rang s .

La matrice X est non singulière, puisqu'elle est sa propre inverse. Il est évident en effet que

$$X^2 = X X = E$$

On voit sans peine que pour intervertir deux colonnes, il suffit de multiplier A à droite par une matrice analogue à X .

Considérons maintenant la matrice carrée

$$P = \begin{bmatrix} 1 & k & 0 & . & . \\ 0 & 1 & 0 & . & . \\ 0 & 0 & 1 & . & . \\ . & . & . & . & . \end{bmatrix}$$

dont tous les éléments sont nuls, sauf ceux de la diagonale principale, et l'élément $(1, 2)$, qui est égal à k . Nous avons ainsi

$$p_{ij} = \delta_{ij}$$

$$p_{12} = k$$

Dans ces conditions, si $B = AP$, nous voyons que :

$$b_{ij} = \sum_h a_{ih} p_{hj} = a_{ij} \quad (j \neq 2)$$

$$b_{i2} = \sum_h a_{ih} p_{h2} = k a_{i1} + a_{i2}$$

B se déduit de A en ajoutant à la deuxième colonne (a_{i2}) les éléments de la première colonne (a_{i1}) multipliés par k. P est une matrice non singulière, puisque son inverse est évidemment la matrice

$$P^{-1} = \begin{bmatrix} 1 & -k & 0 & . & . & . \\ 0 & 1 & 0 & . & . & . \\ 0 & 0 & 1 & . & . & . \\ . & . & . & . & . & . \end{bmatrix}$$

Pour ajouter aux éléments de la deuxième ligne les éléments de la première ligne multipliés par k, il suffira de multiplier A à gauche par la matrice P' déduite de P en intervertissant lignes et colonnes :

$$P' = \begin{bmatrix} 1 & 0 & 0 & . & . & . \\ k & 1 & 0 & . & . & . \\ 0 & 0 & 1 & . & . & . \\ . & . & . & . & . & . \end{bmatrix}$$

Pour ajouter à une ligne quelconque les éléments d'une autre ligne multipliés par k, soit, pour fixer les idées, les éléments de la 3^e ligne, multipliés par k, à ceux de la 5^e ligne, nous pourrions intervertir les lignes 1 et 3, puis les lignes 2 et 5, multiplier à gauche par k, puis intervertir à nouveau les lignes 1 et 3, puis les lignes 2 et 5. Ces opérations successives sont toujours des opérations pouvant s'exprimer en multipliant A par des matrices des types X et P; on multiplie ainsi A par une matrice non singulière.

Ces principes étant posés, nous voyons qu'étant donnée A absolument quelconque, on peut, par des opérations élémentaires :

changer a_{11} en 1.

annuler $a_{12}, a_{13}, \dots, a_{1n}, a_{21}, a_{31}, \dots, a_{m1}$.

changer a_{22} en 1.

annuler $a_{23}, a_{24}, \dots, a_{2n}, a_{32}, a_{42}, \dots, a_{m2}$,

et ainsi de suite.

Cette suite d'opérations sera arrêtée seulement lorsque A aura pris la forme suivante

$$\begin{bmatrix} 1 & 0 & 0 & . \\ 0 & 1 & 0 & . \\ 0 & 0 & 1 & . \\ . & . & . & . \end{bmatrix} = E(m; r; n)$$

dite matrice unité tronquée, où tous les éléments situés en dehors de la première diagonale seront nuls, et où, dans cette première diagonale, tous les éléments seront nuls à l'exception des r premiers qui sont égaux à 1.

On peut ainsi, à toute matrice $A_{(m, n)}$, associer deux matrices carrées, admettant des inverses, soient B et C , telles que

$$C_{(m, m)} A_{(m, n)} B_{(n, n)} = E_{(m; r; n)}$$

La détermination de C et de B peut se faire d'un grand nombre de manières : les opérations élémentaires que nous venons de décrire permettent d'obtenir C et B par des calculs rationnels. En multipliant l'égalité trouvée par C^{-1} à gauche et B^{-1} à droite, nous obtenons :

$$A = C^{-1} E_{(m; r; n)} B^{-1}$$

4.- MATRICES UNITÉS TRONQUÉES

Une matrice unité tronquée $E_{(m; r; n)}$ étant comme nous l'avons vu, une matrice rectangulaire (m, n) dont tous les éléments sont nuls, sauf ceux de la diagonale principale, lesquels sont égaux à 1 pour les r premiers, et nuls en ce qui concerne les autres, nous pouvons chercher la forme du produit de deux pareilles matrices. Soit donc le produit

$$E_{(m; r; n)} \cdot E_{(n; r'; p)}$$

Ce produit est une matrice rectangulaire à m lignes et p colonnes. Les éléments qui sont situés sur la diagonale principale sont nuls sauf si leur rang est inférieur ou égal à r et à r' . Nous en déduisons que :

$$(1) \quad E_{(m; r; n)} \cdot E_{(n; r'; p)} = E_{(m; \rho; p)} \quad \rho = \text{Min} \{ r, r' \}$$

La règle de multiplication (1) peut s'interpréter simplement en examinant l'opération

$$(2) \quad Y = E_{(m; r; n)} X$$

où X est une matrice $(n, 1)$, et Y une matrice $(m, 1)$.

On voit que le passage de X (qui contient n éléments) à Y (qui contient m éléments) se fait en prenant pour r premiers éléments de Y les r premiers éléments de X , et en complétant à m avec $m - r$ zéros. La formule (1) est alors évidente.

On peut, à partir de la formule (1), établir les formules

$$E_{(m; r; r)} \cdot E_{(r; r; n)} = E_{(m; r; n)}$$

$$E_{(r; r; m)} E_{(m; r; n)} E_{(n; r; r)} = E_{(r; r; r)}$$

dans lesquelles il ne faut pas perdre de vue que $r \leq m$, $r \leq n$.

L'intérêt de la matrice unité tronquée est qu'elle représente, à des changements de variables près, la matrice singulière la plus générale. Une relation telle que (2) s'écrit en effet :

$$y_1 = x_1 \quad y_2 = x_2 \dots \quad y_r = x_r \quad y_{r+1} = \dots = y_m = 0$$

On voit que Y ne dépend que de $x_1, x_2, \dots, x_r$, et que $y_{r+1}, \dots, y_m$ sont nuls. Il en résulte que si on veut calculer X à partir de Y supposé donné, ce sera en général impossible, puisque $x_{r+1}, \dots, x_n$ restent indéterminés. Si $n = r$, nous avons évidemment

$$X = E(n; n; m) Y$$

ce qui correspond à l'égalité :

$$E(n; n; m) E(m; n; n) = E(n; n; n)$$

Donc $E(m; n; n)$ admet l'inverse gauche $E(n; n; m)$.

De même $E(m; m; n)$ admet l'inverse droite $E(n; m; m)$.

Ces inverses ne sont pas forcément uniques.

Il est bon, pour étudier complètement la nature des matrices $E(m; r; n)$ de former les produits de $E(m; r; n)$ avec des matrices $A = A(n, p)$ et $B = B(p, m)$. On trouve

$$E(m; r; n) A(n, p) = \begin{bmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & 0 \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix} \begin{bmatrix} a_{n1} & \dots & a_{np} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{np} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ a_{r1} & a_{r2} & \dots & a_{rp} \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \quad \begin{matrix} \uparrow \\ (m \text{ lignes}) \\ \downarrow \end{matrix}$$

$$B(p, m) E(m, r, n) = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1m} \\ b_{21} & b_{22} & \dots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{p1} & b_{p2} & \dots & b_{pm} \end{bmatrix} \begin{bmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & 0 \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1r} & 0 & \dots & 0 \\ b_{21} & b_{22} & \dots & b_{2r} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ b_{p1} & b_{p2} & \dots & b_{pr} & 0 & \dots & 0 \end{bmatrix}$$

On voit ainsi que $E(m; r; n)$ ne peut admettre d'inverse droite que si $r = m$, et que sont alors inverses droite toutes les matrices de la forme :

$$A(n, r) = \begin{bmatrix} E(r, r, r) \\ a(n-r, r) \end{bmatrix}$$

c'est-à-dire les matrices obtenues en ajoutant au dessous de la matrice $E(r, r, r)$ (matrice unité d'ordre r) $n-r$ lignes comprenant r éléments arbitraires.

5.- UNE PROPRIÉTÉ DES MATRICES RECTANGULAIRES

Soit A une matrice rectangulaire "large", c'est à dire possédant plus de colonnes que de lignes

$$A = A_{(m, n)} \quad m < n$$

Il existe alors des matrices à une colonne

$$X = X_{(n, 1)}$$

telles que AX soit nulle, sans que X le soit. Pour montrer ceci, commençons par examiner le cas d'une matrice $E_{(m, r, n)}$.

Si X a les éléments $x_1, x_2, \dots, x_n$, la matrice $E_{(m, r, n)} X$ a les composantes $x_1, x_2, \dots, x_r, 0, 0, \dots, 0$. Il suffit donc de prendre pour X la matrice dont les r premiers éléments sont 0, les $n-r$ autres 1, que nous notons

$$X = \begin{bmatrix} 0_r \\ 1_{n-r} \end{bmatrix}$$

Pour obtenir l'égalité

$$E_{(m, r, n)} X = 0$$

Soit maintenant A . Nous pouvons mettre cette matrice sous la forme

$$A = B E_{(m, r, n)} C \quad B = B_{(m, m)}, \quad C = C_{(n, n)}$$

B et C étant non singulières. Pour obtenir $AX = 0$, il suffit de choisir pour X la matrice

$$X = C^{-1} \begin{bmatrix} 0_r \\ 1_{n-r} \end{bmatrix}$$

laquelle n'est pas nulle, puisque si elle était nulle on obtiendrait en multipliant par C à gauche

$$0 = \begin{bmatrix} 0_r \\ 1_{n-r} \end{bmatrix}$$

On vérifie bien que

$$AX = B E_{(m, r, n)} C X = B E_{(m, r, n)} \begin{bmatrix} 0_r \\ 1_{n-r} \end{bmatrix} = B \cdot 0 = 0$$

La proposition :

"L'équation $AX = 0$ admet toujours, si $m < n$, des solutions en X différentes de zéro",

n'est autre que la proposition bien connue :

"Un système de m équations homogènes à n inconnues admet toujours si $m < n$, des solutions différentes de zéro".

On peut, à l'aide des résultats ci-dessus, traiter des problèmes tels que le suivant.

Problème I

Montrer qu'il est impossible de mettre une matrice unité $E_{(n, n, n)}$ sous la forme

$$E = B C \quad B = B_{n, p} \quad C = C_{p, n} \quad (p < n)$$

c'est-à-dire sous la forme d'un produit de deux matrices rectangulaires dont la première est "haute" (davantage de lignes que de colonnes) et la seconde "large".

Il suffit pour montrer cela de faire intervenir une matrice X telle que

$$X \neq 0 \quad C X = 0$$

Nous savons que de telles matrices existent. De l'égalité

$$E X = B C X$$

on déduit alors

$$X = 0$$

ce qui est contradictoire.

6.- RANG D'UNE MATRICE

Nous sommes maintenant en mesure de définir le rang d'une matrice. Ce sera, pour une matrice $A_{(m, n)}$, le plus petit entier r tel que l'on puisse mettre A sous la forme

$$A = B C \quad B = B_{m, r} \quad C = C_{r, n}$$

La première chose à faire est de vérifier que le rang de $E_{(m, r, n)}$ est r . Soit $\bar{r}$ ce rang. Nous voyons tout d'abord que

$$E_{(m, r, n)} = E_{(m, r, r)} E_{(r, r, n)}$$

ce qui montre que $\bar{r} \leq r$. Supposons maintenant que l'on ait mis $E_{(m, r, n)}$ sous la forme

$$E_{(m, r, n)} = B_{(m, \bar{r})} C_{(\bar{r}, n)}$$

Nous aurons alors

$$E_{(r, r, m)} E_{(m, r, n)} E_{(n, r, r)} = E_{(r, r, r)} = \begin{bmatrix} E_{(r, r, m)} \\ B_{(m, \bar{r})} \end{bmatrix} \begin{bmatrix} C_{(\bar{r}, n)} & E_{(n, r, r)} \end{bmatrix} = K_{(r, \bar{r})} \cdot H_{(\bar{r}, r)}$$

D'après les résultats trouvés à propos du problème I, nous voyons que $\bar{r} \geq r$. Donc $r = \bar{r}$.

Pour déterminer le rang d'une matrice quelconque A , il suffit de procéder à sa réduction par les opérations élémentaires.

Par de telles opérations, on peut en effet mettre $A(m, n)$ sous la forme

$$(3) \quad A = B E(m; r; n) C$$

B et C étant non singulières. Le nombre r qui figure dans (3) est le rang de A . D'une part en effet, la formule (3) peut s'écrire :

$$(4) \quad A = \begin{bmatrix} B(m, m) & E(m; r; r) \end{bmatrix} \begin{bmatrix} E(r; r; n) & C(n; n) \end{bmatrix}$$

où nous voyons apparaître une décomposition de A montrant que le rang de A est au plus égal à r .

D'autre part, si l'on a mis A sous la forme

$$A = B_1(m; \bar{r}) C_1(\bar{r}, n) \quad \bar{r} \leq r$$

on en déduit

$$E(m, r, n) = (B^{-1} B_1) (C_1 C^{-1})$$

d'où

$$\bar{r} = r$$

le rang est donc bien r .

Il résulte de la définition même du rang que le rang d'un produit est au plus égal au rang de celui des facteurs qui a le plus petit rang. En effet, toute décomposition d'un facteur en un produit de deux matrices conduit à une décomposition du produit tout entier en un produit de deux matrices.

7.- RELATION ENTRE LE RANG ET L'EXISTENCE DES INVERSES

Toute matrice A admettant un rang bien déterminé, nous appellerons matrice $(m; r; n)$ une matrice à m lignes, n colonnes, de rang r .

Une matrice $(m; r; n)$ admet une inverse droite unique ou non si $r = m$ et dans ce cas seulement. Soit en effet A une telle matrice, et A^{-1} son inverse droite (ou une de ses inverses droite). A^{-1} est une matrice $(n, \bar{r}, m)$. Nous avons alors

$$(5) \quad E(m; m; m) = A(m; r; n) A^{-1}(n; \bar{r}; m)$$

Le rang de E est m . Donc le rang de A et de A^{-1} doit être au moins m . Comme il ne peut dépasser m , nous avons $r = \bar{r} = m$. Seules les matrices $A(m; m; n)$ ($n \geq m$) peuvent donc admettre une inverse droite. Recherchons une telle inverse, et mettons A sous la forme

$$A(m; m; n) = B(m, m) E(m; m; n) C(n; n)$$

B et C étant non singulières. Cette transformation peut toujours se faire, avons nous vu, à l'aide de transformations élémentaires. Cherchons une inverse droite de $E(m; m; n)$. Ce sera une matrice $(n; m; m)$; soit

$$\begin{bmatrix} e_{ij} \end{bmatrix} \quad \text{cette matrice. Comme} \\ E(m; m; n) = \begin{bmatrix} \delta_{ij} \end{bmatrix} \quad (m, n)$$

le produit de $E(m; m; n)$ par la matrice cherchée aura pour terme général

$$\delta_{ij} = \sum_{h=1}^n \delta_{ih} e_{hj} = e_{ij} \quad \begin{matrix} 1 \leq i \leq m \\ 1 \leq j \leq m \end{matrix}$$

Nous aurons ainsi

$$e_{ij} = \delta_{ij} \quad 1 \leq i \leq m, \quad 1 \leq j \leq m$$

$$e_{ij} \text{ arbitraire pour } m+1 \leq i \leq n$$

Il existe donc une infinité d'inverses droite de $E(m, m, n)$, dépendant de $m(n-m)$ paramètres arbitraires. Soit ξ une quelconque de ces inverses droite. La matrice

$$C^{-1} \xi B^{-1}$$

sera alors une inverse droite de A .

On montrerait de même que la condition nécessaire et suffisante pour que A admette des inverses gauche est que $r = n$. Dans ce cas, A admet une infinité d'inverses gauche, dépendant de $n(n-m)$ paramètres arbitraires.

En ce qui concerne les matrices carrées, nous constatons alors que la condition nécessaire et suffisante pour qu'une matrice carrée (n, n) admette une inverse est que $r = n$. L'inverse est alors à la fois droite et gauche; elle est unique.

8.- DISCUSSION D'UN SYSTÈME D'ÉQUATIONS LINÉAIRES

Soit $A(m; r; n)$ une matrice. Nous pouvons lui associer le système d'équations linéaires

$$(6) \quad A X = Y$$

aux inconnues $x_1, x_2, \dots, x_n; y_1, y_2, \dots, y_m$ sont les seconds membres des équations (6). Sous forme explicite, nous obtenons :

$$(6') \quad \begin{cases} a_{11} x_1 + a_{12} x_2 + \dots + a_{1n} x_n = y_1 \\ a_{21} x_1 + a_{22} x_2 + \dots + a_{2n} x_n = y_2 \\ \dots \\ a_{m1} x_1 + a_{m2} x_2 + \dots + a_{mn} x_n = y_m \end{cases}$$

Décomposons A , par des opérations élémentaires, sous la forme

$$A = B E(m; r; n) C$$

et posons

$$\bar{X} = C X \quad \bar{Y} = B^{-1} Y$$

Le système (6) prend la forme

$$(7) \quad E(m; r; n) \bar{X} = \bar{Y}$$

qui est équivalente à (6), puisque la correspondance entre X et $\bar{X}$ et entre Y et $\bar{Y}$, est biunivoque. Le système (7) n'est autre que :

$$(7') \quad \bar{x}_1 = \bar{y}_1 \quad \bar{x}_2 = \bar{y}_2 \dots \bar{x}_r = \bar{y}_r$$

$$(7'') \quad \bar{y}_{r+1} = \bar{y}_{r+2} = \dots = \bar{y}_m = 0$$

$$(7''') \quad \bar{x}_{r+1}, \bar{x}_{r+2}, \dots, \bar{x}_n \text{ arbitraires.} \quad (11)$$

Le système (6) n'admet donc de solution que si Y satisfait à $m-r$ conditions; ces conditions étant satisfaites, la solution X contient $n-r$ paramètres arbitraires.

9.- MATRICES DE PERMUTATIONS

On appelle matrice de permutations une matrice produit de matrices élémentaires du type X (1). Si on multiplie à droite une matrice A par une matrice de permutation, on opère une permutation des colonnes: si on la multiplie à gauche, on opère une permutation de lignes. Les matrices de permutation sont non singulières. Elles sont carrées. Dans chaque ligne figure un élément, et un seul, égal à 1, les autres éléments étant nuls; il en est de même pour les colonnes.

Soit une matrice quelconque A. Nous pouvons par des permutations portant sur les lignes et les colonnes, et par des additions portant uniquement sur les lignes, annuler tous les éléments se trouvant au-dessous de la première diagonale, et faire en sorte que les r premiers éléments de la première diagonale soient non nuls, tandis que les éléments des lignes de rang plus grands que r sont tous nuls.

Nous mettons ainsi A sous la forme

$$(8) \quad A = B \bar{A} P$$

où P est une matrice de permutation, B une matrice non singulière, et $\bar{A}$ une matrice telle que

$$\bar{A} = \begin{bmatrix} \bar{a}_{ij} \end{bmatrix} \quad \bar{a}_{ij} = 0 \quad \text{pour } i \leq j$$

$$a_{ss} = 0, a_{s+1,s+1} = P \text{ et } a_{sj} = 0$$

$$(j = 1, 2, \dots, n)$$

La matrice $\bar{A}$ peut évidemment se mettre sous la forme

$$\bar{A} = E(m; r; n) C$$

puisque par des opérations sur les colonnes on peut y annuler tous les éléments situés en dehors de la diagonale principale. r est donc le rang de la matrice primitive A.

10.- SYSTÈMES HOMOGÈNES

La décomposition (8) permet de discuter commodément un système homogène. Soit en effet le système :

$$(9) \quad A X = 0$$

qui est équivalent au système

$$(10) \quad \bar{A} \bar{X} = \bar{A} P X = 0$$

(1) cf p. 13.

qui s'écrit

$$(11) \quad \begin{cases} \bar{a}_{11} \bar{x}_1 + \bar{a}_{12} \bar{x}_2 + \dots + \bar{a}_{1r} \bar{x}_r + \bar{a}_{1,r+1} \bar{x}_{r+1} + \dots + \bar{a}_{1n} \bar{x}_n = 0 \\ \bar{a}_{22} \bar{x}_2 + \dots + \bar{a}_{2r} \bar{x}_r + \bar{a}_{2,r+1} \bar{x}_{r+1} + \dots + \bar{a}_{2n} \bar{x}_n = 0 \\ \bar{a}_{rr} \bar{x}_r + \bar{a}_{r,r+1} \bar{x}_{r+1} + \dots + \bar{a}_{r,n} \bar{x}_n = 0 \end{cases}$$

Les $m-r$ dernières équations sont omises, les coefficients $\bar{a}_{ij}$ correspondants étant tous nuls.

Dans les équations (11), tous les $\bar{a}_{ij}$ sont différents de zéro. Il en résulte que $\bar{x}_{r+1}, \bar{x}_{r+2}, \dots, \bar{x}_n$ peuvent être pris arbitrairement, et que $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_r$ sont alors déterminés d'une manière unique. Si nous remarquons que X ne diffère de $\bar{X}$ par la numérotation des inconnues, nous concluons :

Dans le système homogène (8), où A est une matrice $(m; r; n)$; il existe $n-r$ inconnues que l'on peut se donner arbitrairement. Les r autres s'en déduisent d'une manière univoque.

Revenons maintenant au système (6). Ce système n'a de solutions que si les équations (7'') sont satisfaites; or ces équations (7'') constituent un système, homogène, de $m-r$ équations à m inconnues. Ce système a une matrice de rang $m-r$, qui n'est autre que

$$B^{-1} \left[E(m; m; m) - E(m; r; m) \right] = F^{-1} Q_1 E(m; m-r; m) Q_2$$

où Q_1 et Q_2 sont des matrices de permutation.

Il existe donc, parmi les y_i , r nombres y_j qui peuvent être pris arbitrairement, les $m-r$ autres s'en déduisant d'une manière univoque.

Nous voyons ainsi que nous pouvons, par une permutation convenable des x_i et des y_j faire en sorte que les r premiers y_j et les $n-r$ derniers x_i soient arbitraires.

Supposons cette permutation faite, et considérons les équations (6) où nous supposons que les x_i sont des variables indépendantes. Quels que soient les x_i , les $y_{r+1}, y_{r+2}, \dots, y_m$ s'exprimeront linéairement en fonction de $y_1, y_2, \dots, y_r$. Il en résulte que les $m-r$ dernières lignes de A s'expriment en fonction des r premières.

Si donc une matrice A est de rang r , il existe r lignes en fonction desquelles on peut exprimer toutes les autres par combinaison linéaire.

Nous verrons plus tard que le rang d'une matrice n'est pas affecté par l'intervention du rôle des lignes et des colonnes. Il en résulte que la proposition ci-dessus a son équivalente (avec la même valeur de r) en ce qui concerne les colonnes.

CHAPITRE IV

ALGÈBRE DES MATRICES CARRÉES

1.- L'algèbre matricielle est surtout attachée aux matrices carrées, parce que seules de pareilles matrices peuvent être élevées à des puissances, et permettent donc de généraliser la notion de polynôme.

Nous aurons principalement dans ce qui suit à nous occuper d'expressions ayant l'une des formes suivantes :

1) Des polynômes à variable numérique et à coefficients matriciels, c'est-à-dire des expressions telles que

$$(1) \quad A_0 \lambda^m + A_1 \lambda^{m-1} + \dots + A_m = f(\lambda)$$

où λ est un nombre, $A_0, A_1, A_2, \dots, A_m$ des matrices carrées de même ordre.

2) Des polynômes à coefficients numériques et à variable matricielle, c'est-à-dire des expressions de la forme

$$(2) \quad a_0 A^m + a_1 A^{m-1} + \dots + a_m E = g(A)$$

où $a_0, a_1, \dots, a_m$ sont des nombres, et A une matrice.

On sait qu'en général le produit de deux matrices dépend de l'ordre des facteurs. Ceci est une gêne considérable dans des calculs portant sur des expressions telles que (1) ou (2). Dans ce chapitre, nous supposons donc, sauf mention contraire, que nous avons affaire à des matrices permutables, c'est-à-dire telles que

$$AB = BA$$

Les calculs que l'on peut faire sur de pareilles matrices sont assez généraux, d'après les propositions suivantes :

Etant donnée une matrice A , les matrices A^m et A^n sont permutables. Ceci est évident; nous avons

$$A^m \cdot A^n = A^n \cdot A^m = A^{m+n}$$

Si A et B sont permutables, et si f(A), g(B) sont deux polynômes en A et B à coefficients numériques, f(A) et g(B) sont permutables. La démonstration résulte du fait que le produit f(A) g(B) ne fait intervenir que des monômes de la forme

$$A^m B^n$$

qui peuvent s'écrire

$$A \dots A.B \dots B$$

et dans lesquels, par permutation répétée de deux facteurs consécutifs, on peut, sans changer le produit, placer les facteurs dans un ordre quelconque.

Si A et B sont permutables, et si B admet une inverse B⁻¹, A et B⁻¹ sont permutables. Partons en effet de l'égalité

$$A B = B A$$

Multiplions à droite et à gauche par B⁻¹, nous obtenons

$$B^{-1} A = A B^{-1}$$

D'une manière tout à fait générale, on peut considérer des polynômes à plusieurs variables matricielles f(A, B, C) et même des expressions

$$(3) \quad \frac{f(A, B, C)}{g(A, B, C)}$$

représentant le produit de la matrice f(A, B, C) par l'inverse de la matrice g(A, B, C). Si A, B, C sont permutables, toutes les expressions de la forme (3) sont définies sans ambiguïté (si g⁻¹ existe), c'est-à-dire peuvent s'écrire indifféremment f g⁻¹ ou g⁻¹ f. De plus toutes ces expressions de la forme (3) sont permutables, avec A, B, C et entre elles, et peuvent par conséquent entrer dans la formation de nouvelles expressions rationnelles.

Nous allons montrer, sur un exemple, la nature des problèmes que l'on peut ainsi traiter. Considérons la matrice du 2e ordre

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Elevons-le au carré; nous obtenons

$$A^2 = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a^2 + bc & b(a+d) \\ c(a+d) & bc + d^2 \end{bmatrix}$$

Nous voyons apparaître les éléments b(a+d), qui sont les éléments correspondants de A multipliés par a+d. Ceci nous conduit à former la différence

$$A^2 - (a+d) A = \begin{bmatrix} a^2 + bc - a(a+d) & 0 \\ 0 & bc + d^2 - d(a+d) \end{bmatrix} = (bc - ad) E$$

La matrice A est donc racine de l'équation

$$(4) \quad A^2 - (a + d) A + (ad - bc) E = 0$$

obtenue en formant le polynome

$$f(\lambda) = \lambda^2 - \lambda(a + d) + ad - bc$$

en y remplaçant λ par A , et en égalant à zéro. On remarque que l'équation (4) permet immédiatement de former l'inverse de A , puisque, si $ad - bc \neq 0$:

$$A \begin{bmatrix} \frac{a+d}{ad-bc} E - \frac{A}{ad-bc} \end{bmatrix} = E$$

d'où il apparaît que

$$\begin{aligned} A^{-1} &= \frac{1}{ad-bc} \begin{bmatrix} (a+d)E - A \end{bmatrix} \\ &= \frac{1}{ad-bc} \begin{bmatrix} a+d & 0 \\ 0 & a+d \end{bmatrix} - \begin{bmatrix} a & b \\ c & d \end{bmatrix} \\ &= \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \end{aligned}$$

Le dénominateur $ad - bc$, terme constant de $f(\lambda)$, n'est autre que le déterminant de la matrice A , et $f(\lambda)$ lui-même n'est autre que le déterminant de la matrice

$$\begin{bmatrix} a - \lambda & b \\ c & d - \lambda \end{bmatrix} = A - \lambda E$$

Nous avons ainsi montré pour une matrice d'ordre 2, que si on forme le déterminant de $A - \lambda E$

$$f(\lambda) = \begin{bmatrix} A - \lambda E \end{bmatrix}$$

et si on remplace λ par A , on obtient une matrice nulle.

2.- INVERSE D'UNE MATRICE CARRÉE

Rappelons un résultat classique sur les déterminants. Nous appellerons $|A|$ le déterminant formé avec les éléments d'une matrice carrée

$$A = \begin{bmatrix} a_{ij} \end{bmatrix}$$

Nous appellerons $\bar{a}_{ij}$ le cofacteur de a_{ij} dans le développement de $|A|$, c'est à dire le coefficient de a_{ij} dans l'expression du déterminant; on sait que ce cofacteur est

$$\bar{a}_{ij} = (-1)^{i+j} \Delta_{ij}$$

en appelant Δ_{ij} le déterminant du tableau obtenu en supprimant, dans le tableau des éléments de A , la i ème ligne et la j ème colonne. On sait de plus que

$$\sum_j a_{ij} \bar{a}_{ij} = |A|, \text{ quel que soit } i.$$

$$\sum_i a_{ij} \bar{a}_{ij} = |A|, \text{ quel que soit } j.$$

$$\sum_j a_{ij} \bar{a}_{hj} = 0, \text{ quel que soit } i, \text{ si } h \neq i.$$

$$\sum_i a_{ij} \bar{a}_{ih} = 0, \text{ quel que soit } j, \text{ si } h \neq i.$$

La matrice

$$\bar{A} = [\bar{a}_{ij}]$$

est la matrice adjointe de A .

Considérons la matrice transposée de $\bar{A}$, que nous désignerons par $\hat{A}$ c'est à dire la matrice déduite de $\bar{A}$ par échange des lignes et colonnes. Elle sera de la forme

$$(4) \quad \hat{A} = [\hat{a}_{ij}] \quad \hat{a}_{ij} = (-1)^{i+j} \Delta_{ji}$$

On vérifie alors que

$$A\hat{A} = \hat{A}A = |A| \cdot E$$

c'est à dire que le produit d'une matrice par son adjointe transposée est une matrice dont tous les éléments sont nuls, sauf ceux de la diagonale principale, qui sont égaux à $|A|$.

Si donc $|A| \neq 0$, nous avons l'expression de l'inverse de A (dont nous savons que si elle existe elle est unique) :

$$(5) \quad A^{-1} = \frac{\hat{A}}{|A|}.$$

Si $|A| = 0$, A ne peut pas avoir d'inverse. En effet, les règles de multiplication des déterminants montrent que

$$|A \cdot B| = |A| \cdot |B|$$

d'où

$$|A| \cdot |A^{-1}| = |E| = 1.$$

Si $|A| = 0$, cette dernière égalité est impossible. Donc la condition $A \neq 0$ est nécessaire et suffisante pour l'existence de l'inverse, qui a toujours, quand elle existe, la forme (5).

3.- INVERSE D'UNE MATRICE $A - \lambda E$

Considérons une matrice A , d'ordre n , et la famille des matrices $A - \lambda E$, dépendant du paramètre numérique λ . Le déterminant de $A - \lambda E$ est

un polynome de degré n , que nous notons

$$f(\lambda) = |A - \lambda E|$$

La matrice adjointe transposée de $A - \lambda E$ est une matrice dont chacun des éléments est un polynome de degré $n-1$ au plus, puisque chaque élément est lui-même le déterminant d'un tableau à $n-1$ lignes et $n-1$ colonnes, dans lequel λ figure au premier degré.

Si nous ordonnons en λ cette adjointe transposée, nous obtenons un polynome en λ à coefficients matriciels

$$g(\lambda) = B_1 \lambda^{n-1} + B_2 \lambda^{n-2} + \dots + B_n$$

Le produit de $g(\lambda)$, matrice adjointe de $A - \lambda E$, par $A - \lambda E$ est égal à $f(\lambda) E$, puisque $f(\lambda)$ est le déterminant de $A - \lambda E$. D'où

$$(A - \lambda E) [B_1 \lambda^{n-1} + B_2 \lambda^{n-2} + \dots + B_n] = f(\lambda) E$$

Développons le membre de gauche en respectant l'ordre des facteurs, précaution nécessaire puisque nous ne connaissons pas la nature exacte des matrices B , et ordonnons en λ

$$(6) \quad -B_1 \lambda^n + (A B_1 - B_2) \lambda^{n-1} + (A B_2 - B_3) \lambda^{n-2} + \dots + A B_n = f(\lambda) E$$

Cette égalité matricielle, d'après la définition de l'égalité de deux matrices, équivaut à n^2 égalités entre nombres. Considérons l'élément occupant la place (e_{ij}) dans la matrice située à gauche du signe égale; et l'élément occupant la même place dans la matrice de droite. Ils sont égaux; ce sont deux polynômes en λ ; l'égalité a lieu quel que soit λ ; donc les coefficients des puissances semblables de λ sont égaux. Cette égalité des puissances semblables de λ ayant lieu dans chaque élément, il en résulte évidemment qu'il a lieu si on considère les coefficients de λ dans (6). Explicitons $f(\lambda)$:

$$f(\lambda) = (-1)^n [\lambda^n - \mu_1 \lambda^{n-1} + \mu_2 \lambda^{n-2} - \dots + (-1)^n \mu_n]$$

Nous obtenons :

$$B_1 = (-1)^{n-1} E$$

$$A B_1 - B_2 = (-1)^n \mu_1 E$$

$$A B_2 - B_3 = (-1)^{n-1} \mu_2 E$$

Nous voyons que $B_1, B_2, B_3, \dots$ sont des polynomes en A , de degrés respectifs $0, 1, 2, \dots$ et que par conséquent le polynome matriciel $g(\lambda)$ ne contient que des coefficients permutable entre eux et avec A . Nous sommes dans le cas favorable où nous n'avons pas à nous préoccuper de l'ordre des facteurs

Considérons maintenant le polynome $f(\lambda)$; si nous y remplaçons par A , nous obtenons la matrice $f(A)$. E (le facteur E intervenant pour que le terme constant de $f(\lambda)$, s'il y en a un, prenne forme matricielle). $f(\lambda) E$ peut s'écrire sous la forme (6). Nous en déduisons que :

$$f(A) \cdot E = -B_1 A^n + (A B_1 - B_2) A^{n-1} + \dots + A B_n$$

et par conséquent, puisque nous pouvons intervertir l'ordre des facteurs, nous avons $f(A) = 0$. *

Le polynome $f(\lambda)$ s'appelle polynome caractéristique de A . Nous avons donc montré que si

$$f(\lambda) = |A - \lambda E|$$

on a

$$f(A) = 0$$

et que la matrice inverse de $A - \lambda E$ est de la forme

$$(7) \quad (A - \lambda E)^{-1} = \frac{B_1 \lambda^{n-1} + B_2 \lambda^{n-2} + \dots + B_n}{f(\lambda)}$$

où $B_1, B_2, \dots, B_n$ sont des matrices pouvant s'exprimer par des polynomes en A , et qui sont donc permutables avec A et entre elles.

Si $f(0) \neq 0$, c'est-à-dire si $|A| \neq 0$, A admet une inverse, qui s'obtient en faisant simplement $\lambda = 0$ dans (7), d'où $A^{-1} = \frac{B_n}{f(0)}$.

Nous voyons par conséquent que des opérations rationnelles sur une matrice nous conduisent à des matrices permutables, et que par conséquent nous pouvons dans beaucoup de cas opérer comme en algèbre ordinaire. Proposons nous par exemple le problème suivant, où nous prenons délibérément les notations de l'algèbre ordinaire.

4.- PROBLÈME II

A étant une matrice, et λ un nombre, est-il correct d'écrire

$$\frac{1}{A - \lambda} - \frac{1}{A} = \frac{\lambda}{A(A - \lambda)}?$$

Nous n'insistons pas sur le fait que $\frac{1}{A}$ est écrit au lieu de A^{-1} , $A - \lambda$ au lieu de $A - \lambda E$.

La formule est correcte.

Posons en effet

$$\begin{aligned} A - \lambda &= B^{-1} \\ A &= C^{-1} \\ A(A - \lambda E) &= D^{-1} \end{aligned}$$

La formule signifie que

$$B - C = \lambda D$$

ou encore que

$$B D^{-1} - C D^{-1} = \lambda E$$

* Il est d'usage d'omettre le facteur E dans $f(A)$. E , aucune confusion n'étant possible.

Or,

$$B D^{-1} = B A (A - \lambda E)$$

D'après la définition de B, nous avons

$$B A = E + \lambda B$$

d'où

$$B D^{-1} = (E + \lambda B) (A - \lambda E) = A - \lambda E + \lambda B A - \lambda^2 B = A$$

De même

$$C D^{-1} = C A (A - \lambda E) = A - \lambda E$$

La formule est bien correcte.

Nous pouvons, dans le même esprit, tirer de (7) les valeurs de B_1 , B_2 , ..., B_n :

5.- PROBLÈME III

Exprimer $B_1, B_2, \dots, B_n$ en fonction de A^{-1} (en utilisant la formule (7)).

La formule (7) peut s'écrire

$$B_1 \lambda^{n-1} + B_2 \lambda^{n-2} + \dots + B_n = \frac{f(\lambda)}{A - \lambda}$$

En faisant $\lambda = 0$, nous obtenons

$$B_n = \frac{f(0)}{A}$$

d'où

$$B_1 \lambda^{n-2} + B_2 \lambda^{n-3} + \dots + B_{n-1} = \frac{1}{\lambda} \left[\frac{f(\lambda)}{A - \lambda} - \frac{f(0)}{A} \right]$$

Faisons $\lambda = 0$ dans cette expression. Nous avons à chercher la limite de

$$\frac{A f(\lambda) - f(0) (A - \lambda)}{\lambda A (A - \lambda)}$$

laquelle est

$$\frac{f(0) + A f'(0)}{A^2}$$

d'où

$$B_{n-1} = \frac{f(0) + A f'(0)}{A^2}$$

Continuons ce processus

$$B_1 \lambda^{n-3} + B_2 \lambda^{n-4} + \dots + B_{n-2} = \frac{1}{\lambda^2} \left[\frac{f(\lambda) - f(0)}{A - \lambda} - \frac{f(0)}{A} \right] - \frac{f(0) + Af'(0)}{\lambda A^2}$$

Nous sommes conduits, en faisant $\lambda = 0$, à

$$B_{n-2} = \frac{1}{A^3} \left[f(0) + Af'(0) + \frac{A^2}{1 \cdot 2} f''(0) \right]$$

et par conséquent

$$B_{n-p} = \frac{1}{A^{p+1}} \left[f(0) + Af'(0) + \frac{A^2}{1 \cdot 2} f''(0) + \dots + \frac{A^p}{p!} f^{(p)}(0) \right]$$

Tous les B sont ainsi des polynômes en A^{-1} , de forme relativement simple. B_1 est alors de la forme

$$B_1 = \frac{1}{A^n} = \left[f(0) + Af'(0) + \frac{A^2}{1 \cdot 2} f''(0) + \dots + \frac{A^{n-1}}{n-1!} f^{(n-1)}(0) \right]$$

On peut remarquer alors que

$$f(\lambda) = f(0) + \lambda f'(0) + \frac{\lambda^2}{1 \cdot 2} f''(0) + \dots + \frac{\lambda^n}{n!} f^{(n)}(0)$$

et que par conséquent, puisque $f(A) = 0$

$$B_1 = -\frac{1}{A^n} \cdot \frac{A^n}{n!} f^{(n)}(0) = (-1)^{n-1} E$$

conformément à ce qui a déjà été vu.

6.- RELATION ENTRE LE DEGRÉ DU POLYNÔME CARACTÉRISTIQUE ET LE RANG D'UNE MATRICE CARRÉE

Une matrice carrée de rang r peut se mettre sous la forme

$$A = B^{-1} E (n; r; n) C^{-1} = B^{-1} e C^{-1}$$

où e est une matrice unité tronquée, B et C deux matrices non singulières, et admettant par conséquent des inverses. Considérons maintenant le polynôme caractéristique de A :

$$f(\lambda) = |A - \lambda E|$$

Nous pouvons l'écrire

$$f(\lambda) = |B^{-1} e C^{-1} - \lambda E|$$

ou encore

$$f(\lambda) = |B^{-1} (e - \lambda BC) C^{-1}|$$

d'où nous tirons

$$|e - \lambda BC| = |B| \cdot |C| \cdot f(\lambda)$$

Nous savons que $|B|$ et $|C|$ sont deux quantités non nulles. Les coefficients nuls de $f(\lambda)$ sont donc ceux de

$$f_1(\lambda) = |e - \lambda M|$$

M étant une matrice non singulière. Posant $\mu = \frac{1}{\lambda}$, nous voyons que

$$f_1(\lambda) = \lambda^n |\mu e - M|$$

Le déterminant $|M - \mu e|$ est un déterminant analogue à un polynôme caractéristique, avec la différence que la matrice unité E y est remplacée par une matrice unité tronquée.

Ce polynôme est manifestement de degré r . Donc $f(\lambda)$ contient λ^{n-r} en facteur. Donc :

Les matrices carrées de rang r sont caractérisées par le fait que le polynôme caractéristique a $n-r$ racines nulles.

7.- DIAGONALISATION D'UNE MATRICE CARRÉE QUAND $f(\lambda)$ a n RACINES DISTINCTES

Supposons que le polynôme caractéristique $f(\lambda)$ de A ait n racines distinctes $\lambda_1, \dots, \lambda_n$. Pour chacune de ces racines, nous avons :

$$|A - \lambda_i E| = 0$$

Le système d'équations linéaires

$$(A - \lambda_i E) X = 0$$

admet donc au moins une solution non nulle. Soit X_i une de ces solutions. Nous avons par conséquent

$$AX_i = \lambda_i X_i \quad (i = 1, 2, \dots, n)$$

Le vecteur X_i est un vecteur colonne. En juxtaposant ces colonnes, nous obtenons une matrice

$$B = [X_1, X_2, \dots, X_n]$$

qui est non singulière. Si en effet elle était singulière son déterminant serait nul, et il existerait entre les colonnes une relation linéaire de la forme

$$k_1 X_1 + k_2 X_2 + \dots + k_n X_n = 0$$

d'où l'on déduirait par multiplication par A .

$$k_1 \lambda_1 X_1 + k_2 \lambda_2 X_2 + \dots + k_n \lambda_n X_n = 0$$

et par nouvelle multiplication

$$k_1 \lambda_1^2 X_1 + k_2 \lambda_2^2 X_2 + \dots + k_n \lambda_n X_n = 0$$

et ainsi de suite. En éliminant les coordonnées de $X_1, X_2, \dots, X_n$ entre ces équations, on obtiendrait

$$\begin{vmatrix} 1 & 1 & \dots & 1 \\ \lambda_1 & \lambda_2 & \dots & \lambda_n \\ \lambda_1^{n-1} & \lambda_2^{n-1} & \dots & \lambda_n^{n-1} \end{vmatrix} = 0$$

ce qui est impossible, puisque ce déterminant, qui est de Vandermonde, ne peut s'annuler pour des λ_i distincts. La matrice B admet donc une matrice inverse B^{-1} , qui peut s'écrire sous la forme

$$B^{-1} = \{U_1, U_2, \dots, U_n\}$$

c'est à dire d'une juxtaposition de n vecteurs ligne, tels que

$$U_i X_j = \delta_{ij}$$

Si nous formons le produit

$$B^{-1} AB$$

nous voyons alors qu'il prend la forme

$$D = \begin{bmatrix} \lambda_1 & & \\ & \delta_{ij} & \\ & & \end{bmatrix}$$

d'une matrice dans laquelle tous les éléments sont nuls sauf ceux de la diagonale principale qui sont les λ_i . A peut donc se mettre sous la forme

$$A = B D B^{-1}$$

D étant une matrice diagonale. L'opération qui a consisté à chercher B de manière à mettre A sous la forme ci-dessus s'appelle diagonalisation de A . Cette opération est particulièrement simple quand $f(\lambda)$ a des racines distinctes.

Lorsque $f(\lambda)$ a des racines multiples, la situation est plus compliquée, comme le montre le problème:

8.- PROBLÈME IV

Mettre sous forme diagonale la matrice

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

en supposant que $f(\lambda)$ a une racine double.

$f(\lambda)$ a ici la forme

$$\lambda^2 - (a+d)\lambda + ad - bc.$$

Il y a une racine double si

$$(a + d)^2 - 4ad + 4bc = 0$$

c'est-à-dire si

$$(a - d)^2 + 4bc = 0$$

Soit λ_0 la racine double. Nous avons alors

$$a + d = 2\lambda_0$$

$$ad - bc = \lambda_0^2$$

Supposons $a \neq d$, auquel cas b et c sont tous deux différents de zéro. Nous pouvons poser

$$a = \lambda_0 + k \quad d = \lambda_0 - k$$

$$b = \mu k \quad c = -\frac{k}{\mu}$$

Nous avons alors

$$A = \lambda_0 E + k \begin{bmatrix} 1 & \mu \\ -\frac{1}{\mu} & -1 \end{bmatrix}$$

Soit B une matrice quelconque non singulière, que nous supposons de déterminant égal à 1, ce qui ne restreint évidemment pas la généralité. Nous aurons ainsi

$$B = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \quad B^{-1} = \begin{bmatrix} \delta & -\beta \\ -\gamma & \alpha \end{bmatrix} \quad \alpha\delta - \beta\gamma = 1$$

$$BAB^{-1} = \lambda_0 E + k \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \begin{bmatrix} 1 & \mu \\ -\frac{1}{\mu} & -1 \end{bmatrix} \begin{bmatrix} \delta & -\beta \\ -\gamma & \alpha \end{bmatrix}$$

$$\lambda_0 E + \begin{bmatrix} \alpha\delta + \beta\gamma - (\alpha\gamma\mu + \frac{\beta\delta}{\mu}) & \mu(\alpha - \frac{\beta}{\mu})^2 \\ -\mu(\gamma - \frac{\delta}{\mu})^2 & (\alpha\gamma\mu + \frac{\beta\delta}{\mu}) - \alpha\delta - \beta\gamma \end{bmatrix}$$

La réduction à la forme diagonale est impossible, puisqu'il faudrait avoir à la fois

$$\beta = \mu\alpha \quad \delta = \mu\gamma$$

ce qui est incompatible avec $\alpha\delta - \beta\gamma = 1$

On ne peut annuler qu'un élément dans la matrice réduite, soit le terme $\gamma - \frac{\delta}{\mu}$

Posons par conséquent :

$$\delta = \mu\gamma$$

Nous avons alors

$$B A B^{-1} = \begin{bmatrix} \lambda_0 & \frac{1}{\mu \gamma^2} \\ 0 & \lambda_0 \end{bmatrix}$$

Si $\mu > 0$, nous pouvons prendre $\gamma = \frac{1}{\sqrt{\mu}}$, ce qui conduit à la forme réduite

$$B A B^{-1} = \begin{bmatrix} \lambda_0 & 1 \\ 0 & \lambda_0 \end{bmatrix}$$

Si $a = d$, un des éléments b ou c est nul; la réduction est encore impossible, sauf si $b = c = 0$, auquel cas elle est déjà faite.

9.- RÉDUCTION D'UNE MATRICE SYMÉTRIQUE

Une matrice symétrique est une matrice telle que

$$a_{ij} = a_{ji}$$

Pour étudier commodément de telles matrices, on utilise l'opération de transposition, qui associe à toute matrice

$$A = [a_{ij}]$$

la matrice

$$A' = [a'_{ij}]$$

avec

$$a'_{ij} = a_{ji}$$

La transposition est simplement l'échange des lignes et des colonnes. On vérifie que

$$(A')^{-1} = (A^{-1})'$$

et que

$$(AB)' = B' A', (ABC)' = C' B' A', \dots$$

Il faut prendre garde que A' , quoique étroitement liée à A , n'est en général pas commutable avec A , c'est à dire que $AA' \neq A'A$.

Pour en revenir aux matrices symétriques, elles satisfont à la condition (nécessaire et suffisante) :

$$A = A'$$

Soit donc A une matrice symétrique, dont nous supposons d'abord que le polynôme caractéristique $f(\lambda)$ a toutes ses racines distinctes.

Soit X un vecteur propre associé à la racine λ_0 .

$$A X = \lambda_0 X$$

Nous pouvons supposer que la somme des carrés des éléments de X est égale à 1 ce qui s'exprime par

$$X'X = 1$$

Si X est réel, cela revient à multiplier tous les éléments par un même nombre, ce qui ne change rien à la propriété d'être un vecteur propre, non nul associé à λ_0 . Nous allons voir que l'on peut supposer X réel. Si en effet λ_0 ou X étaient complexes (A étant réelle), nous aurions en même temps

$$AX = \lambda_0 X$$

$$AX^* = \lambda_0^* X^*$$

l'astérisque désignant la substitution du nombre complexe conjugué à tout nombre complexe figurant dans l'expression.

Nous déduisons des égalités ci-dessus

$$X'^* A X = \lambda_0 X^* X$$

$$X' A X^* = \lambda_0^* X' X^*$$

Les expressions figurant au premier membre sont transposées l'une de l'autre; ce sont des matrices $(1, 1)$. Elles sont donc égales. De même

$$X'^* X = X' X^*$$

sont égales, comme transposées, puisque ce sont des matrices $(1, 1)$, c'est-à-dire des nombres. Il en résulte que $\lambda_0 = \lambda_0^*$. toutes les racines de $f(\lambda)$ sont réelles. Comme X est une solution du système homogène.

$$(A - \lambda_0 E) X = 0$$

à coefficients réels, nous pouvons supposer que X est réel. Nous pouvons ainsi, à la matrice symétrique A , associer n racines réelles $\lambda_1, \lambda_2, \dots, \lambda_n$ (que nous supposons distinctes) de $f(\lambda) = 0$, et n vecteurs propres réels X_i satisfaisant à

$$AX_i = \lambda_i X_i \quad X_i' X_i = 1$$

Nous allons évaluer $X_i' X_j$. Nous avons

$$AX_i = \lambda_i X_i$$

$$AX_j = \lambda_j X_j$$

d'où

$$X_j' A X_i - X_i' A X_j = \lambda_i X_j' X_i - \lambda_j X_i' X_j$$

Nous voyons, par transposition (A étant symétrique), que

$$X_j' A X_i = X_i' A X_j \quad (\text{égalités entre nombres})$$

$$X_i' X_j = X_j' X_i$$

d'où il résulte, puisque $\lambda_i \neq \lambda_j$, que

$$X_i' X_j = 0 \quad \text{si } i \neq j$$

Nous avons donc d'une manière générale

$$X_i^1 X_j = \delta_{ij}$$

Si donc nous posons

$$B = [X_{11}, X_{21}, \dots, X_{n1}]$$

la matrice inverse de B n'est autre que sa transposée :

$$B' = \{X_1^1, X_2^1, \dots, X_n^1\} = B^{-1}$$

Une matrice symétrique A peut donc être mise sous forme diagonale à l'aide d'une matrice B qui est identique à son inverse transposée.

$$B^{-1} A B = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \quad (BB' = E)$$

Nous avons vérifié cette propriété lorsque tous les λ_i sont distincts. Elle reste vraie même s'ils ne sont pas distincts, comme nous le verrons plus bas.

Les matrices B satisfaisant à $BB' = E$ jouent un rôle particulièrement important dans les applications du calcul matriciel. On les appelle matrices orthogonales.

10.- GÉNÉRALISATION DES NOTATIONS MATRICIELLES

Pour traiter des questions théoriques un peu compliquées, comme celles qui sont relatives à la diagonalisation d'une matrice quand le polynôme caractéristique admet des racines multiples, on peut être amené à considérer des matrices dont les éléments sont eux-mêmes des matrices, telles que des matrices de la forme

$$A = [a_{ij}]$$

a_{ij} étant lui-même une matrice. Cette matrice a_{ij} pourra être une matrice de nombres, ou une matrice de matrices. Examinons le cas où a_{ij} est une matrice de nombres. Considérons une autre matrice B, de même nature que A

$$B = [b_{ij}]$$

Le produit de A par B aura pour élément général la matrice

$$c_{ij} = \sum_b a_{ib} b_{bj}$$

ce qui exige que :

- 1)- a_{ih} soit multipliable par b_{hj} ;
 - 2)- que toutes les matrices $a_{ih} b_{hj}$ aient mêmes dimensions
- Appelons

$$m(i, j), n(i, j)$$

les dimensions (nombres de lignes et colonnes) de a_{ij} , et

$$m'(i, j), n'(i, j)$$

les dimensions de b_{ij} . Nous devons avoir

$$1) - n(i, h) = m'(h, j)$$

$$2) - m(i, h) \text{ indépendant de } h$$

$$n'(h, j) \text{ indépendant de } h.$$

Pour que la notation généralisée soit utilisable, nous voyons donc que nous sommes amenés à astreindre les a_{ij} à avoir des dimensions de la forme

$$m(i, j) = m(i)$$

$$n(i, j) = n(j)$$

ce qui revient à dire que les éléments a_{ij} sont obtenus en découpant dans une matrice A écrite in extenso les rectangles

	n_1	n_q
m_1	a_{11}	
m_p		a_{pq}

Dans ces conditions, étant donnée une matrice A formée de nombres

$$A = [a_{ij}] \quad (m, n)$$

si on pose

$$X_s = \begin{bmatrix} a_{1s} \\ a_{2s} \\ \vdots \\ a_{ms} \end{bmatrix}$$

on pourra écrire

$$A = [X_1 \ X_2 \ \dots \ X_n]$$

et si on pose

$$U_t = [a_{t1} \ a_{t2} \ \dots \ a_{tn}]$$

on pourra écrire

$$A = \begin{bmatrix} U_1 \\ \vdots \\ U_m \end{bmatrix} = \{U_1 \ U_2 \ \dots \ U_m\}$$

On remarquera que dans ces conditions, si on considère deux matrices mises sous la forme

$$\begin{aligned} A &= \{ U_1 \ U_2 \ \dots \ U_n \} \\ B &= \begin{bmatrix} Y_1 & Y_2 & \dots & Y_n \end{bmatrix} \end{aligned}$$

le produit AB se met sous la forme

$$AB = \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{bmatrix} \begin{bmatrix} Y_1 & Y_2 & \dots & Y_n \end{bmatrix} = \begin{bmatrix} U_i & Y_j \end{bmatrix}$$

En particulier, l'inverse de A (mise sous la forme d'une superposition de vecteurs lignes U_i) est une matrice B (mise sous la forme de juxtaposition de vecteurs colonnes Y_i) tels que

$$U_i Y_j = \delta_{ij} \quad (\text{indice de Kronecker})$$

11.- MULTIPLICATION ET DIVISION DES POLYNOMES DE MATRICES

Toute formule relative à des polynômes à variable scalaire qui se réduit à une identité algébrique reste vraie pour des polynômes de matrices.

En effet, les règles de calcul sont les mêmes dans les deux cas, sauf les exceptions suivantes, qu'il faut toujours avoir présentes à l'esprit :

Le produit n'est pas commutatif

Un produit peut être nul sans qu'aucun de ses facteurs le soit.

La non commutativité du produit, dans le cas d'une seule variable matricielle, ne nous gênera pas, pour la raison évidente que le produit de deux monômes est indépendant de l'ordre des termes :

$$(4) \quad aA^p \cdot bA^q = ba^q \cdot aA^p = ab A^{p+q}$$

Le produit de deux polynômes $f(A)$ et $g(A)$ se fera d'après les règles usuelles. De même, nous pourrions diviser deux polynômes suivant les puissances ascendantes ou descendantes puisque les identités

$$\begin{aligned} (5) \quad f(\lambda) &= g(\lambda) q(\lambda) + r(\lambda) \\ f(\lambda) &= g(\lambda) q(\lambda) + \lambda^N r(\lambda) \end{aligned}$$

restent des identités si l'on substitue à λ une matrice carrée quelconque A . De même la formule de Taylor :

$$f(\lambda + h) = f(h) + \lambda f'(h) + \dots + \frac{\lambda^m}{m!} f^{(m)}(h) + \dots + \frac{\lambda^N}{N!} f^{(N)}(h) \text{ relative}$$

à un polynôme, qui est une identité en λ et h , reste une identité lorsqu'on remplace soit h , soit λ par une matrice A . Nous obtenons ainsi :

$$(6) \quad \begin{cases} f(A + h) = Ef(h) + Af'(h) + \dots + \frac{A^m}{m!} f^{(m)}(h) + \dots + \frac{A^N}{N!} f^{(N)}(h) \\ f(\lambda E + A) = f(A) + \lambda f'(A) + \dots + \frac{\lambda^m}{m!} f^{(m)}(A) + \dots + \frac{\lambda^N}{N!} f^{(N)}(A) \end{cases}$$

Remarquons que dans la deuxième formule, $f^{(N)}(A)$ est un polynôme de degré 0 en A , c'est-à-dire un scalaire multiplié par E .

Si dans les formules (6) nous faisons $h = \lambda$, nous obtenons l'identité :

$$(9) \quad \begin{cases} f(A) + \lambda f'(A) + \dots + \frac{\lambda^m}{m!} f^{(m)}(A) + \dots + \frac{\lambda^N}{N!} f^{(N)}(A) \\ = E f(\lambda) + A f'(\lambda) + \dots + \frac{A^m}{m!} f^{(m)}(\lambda) + \dots + \frac{A^N}{N!} f^{(N)}(\lambda) = f(A + \lambda E) \end{cases}$$

12. POLYNÔME TYPIQUE D'UN VECTEUR DANS UNE MATRICE

Soit A une matrice, et X un vecteur. Considérons la suite de vecteurs

$$X, AX, \dots, A^n X$$

Dans cette suite figurent $n+1$ vecteurs qui ne peuvent donc être indépendants. Il existe entre eux une combinaison linéaire

$$a_0 X + a_1 AX + \dots + a_n A^n X = 0$$

ce qui peut s'écrire

$$(1) \quad g(A) X = 0$$

$g(A)$ étant le résultat de la substitution de A à λ dans le polynôme $g(\lambda)$. Parmi tous les polynômes $g_1(\lambda)$ satisfaisant à (1), considérons celui, soit $g_1(\lambda)$, qui est de plus bas degré. Pour l'uniformité, nous supposons égal à un le coefficient du terme en plus haut degré en λ . Ce polynôme $g_1(\lambda)$ sera appelé polynôme typique du vecteur X dans la matrice A . Ce polynôme est unique. Si en effet $g_1(\lambda)$, $g_2(\lambda)$ satisfaisaient à (1), $g_1(\lambda) - g_2(\lambda)$ y satisferait aussi, de sorte que $g_1(\lambda)$ et $g_2(\lambda)$ ne pourraient être tous deux de degré minimum.

Tout polynôme $f(\lambda)$ tel que

$$f(A) X = 0$$

est un multiple du polynôme typique de X dans A . Il suffit pour le voir de procéder à la division

$$f(\lambda) = g(\lambda) q(\lambda) + r(\lambda)$$

d'où

$$r(A) X = 0$$

$g(\lambda)$ ne peut être de degré minimum que si $r(\lambda) = 0$

13. POLYNÔME TYPIQUE D'UNE MATRICE

Soit une matrice A , n vecteurs indépendants $X_1, X_2, \dots, X_n$, et leurs polynômes typiques respectifs $g_1, g_2, \dots, g_n$.

Comme tout vecteur est de la forme

$$X = \sum a_i X_i$$

il en résulte que pour tout X :

$$(2) \quad g_1(A) \quad g_2(A) \dots g_n(A) X = 0$$

En effet, les facteurs du premier membre de (2) peuvent être écrits dans un ordre quelconque.

Parmi tous les polynômes $g(\lambda)$ tels que

$$(3) \quad g(A) X = 0$$

quel que soit X , c'est-à-dire tels que

$$(4) \quad g(A) = 0$$

considérons celui de degré minimum, astreignons le à avoir pour la plus haute puissance de λ le coefficient 1. Ce polynôme sera appelé polynôme typique de A . Par les mêmes raisonnements que dans le paragraphe précédent, on montre qu'il est unique, et que tout polynôme satisfaisant à (4) est un multiple du polynôme typique. On constate de même que puisque, pour un vecteur quelconque X , le polynôme caractéristique de A satisfait à (3), c'est un multiple du polynôme typique de X dans A . Le polynôme caractéristique défini plus haut est donc un multiple du polynôme typique de A , et de tous les polynômes typiques de vecteurs dans A .

14.- POLYNÔME TYPIQUE D'UNE MATRICE NON SINGULIÈRE

Si A est non singulière, $g(0) \neq 0$, et réciproquement. Si en effet A est non singulière, si l'on avait $g(0) = 0$ on en déduirait que

$$g(A) A^{-1} = 0$$

$g(\lambda) : \lambda$ étant un polynôme $f_1(\lambda)$ tel que $f_1(A) = 0$. $f(\lambda)$ ne serait pas de degré minimum.

Réciproquement, si $g(0) \neq 0$, le polynôme.

$$(5) \quad g_1(\lambda) = \frac{g(0) - g(\lambda)}{g(0)}$$

définit une matrice $B = g_1(A)$ inverse de A , A est par conséquent non singulière.

15.- RÉDUCTION D'UNE MATRICE

Soit une matrice non singulière A , de polynôme typique

$$(6) \quad g(\lambda) = (\lambda - \lambda_1)^{\alpha_1} (\lambda - \lambda_2)^{\alpha_2} \dots (\lambda - \lambda_k)^{\alpha_k} = g_1(\lambda) g_2(\lambda) \dots g_k(\lambda)$$

en posant $g_i(\lambda) = (\lambda - \lambda_i)^{\alpha_i}$

Désignons d'une manière générale par S_i l'ensemble des vecteurs X_i satisfaisant à l'équation

$$(7) \quad g_i(A) X = (A - \lambda_i E)^{\alpha_i} X = 0$$

Ces systèmes de vecteurs sont disjoints; en effet si l'on avait à la fois

$$g_i(A) X = (A - \lambda_i E)^{\alpha_i} X = 0$$

$$g_j(A) X = (A - \lambda_j E)^{\alpha_j} X = 0$$

on pourrait, puisque $g_i(\lambda)$, $g_j(\lambda)$ sont des polynômes premiers entre eux, écrire une identité de la forme :

$$1 = P(\lambda) g_i(\lambda) + Q(\lambda) g_j(\lambda)$$

d'où l'on réduirait en remplaçant λ par A , puis multipliant à droite par X , que $X = 0$.

Par ailleurs si X appartient à S_i , il en est de même de AX , puisque A et $g_i(A)$ sont commutables.

Soit maintenant un vecteur X quelconque. Il existe deux polynômes P_1 et Q_1 tels que :

$$P_1(\lambda) g_1(\lambda) + Q_1(\lambda) g_2(\lambda) \dots g_k(\lambda) = 1^{(1)}$$

d'où

$$X = P_1(A) g_1(A) X + Q_1(A) g_2(A) \dots g_k(A) X$$

et par conséquent

$$g_1(A) [X - P_1(A) g_1(A) X] = 0$$

de sorte que nous pouvons mettre X sous la forme

$$X = X_1 + Y_2$$

X_1 appartenant à S_1 , Y_2 étant de la forme

$$Y_2 = g_1(A) Z_2 \quad (Z_2 = P_1(A) X)$$

Nous pouvons continuer ainsi, en considérant une identité

$$P_2(\lambda) g_2(\lambda) + Q_2(\lambda) g_3(\lambda) \dots g_k(\lambda) = 1$$

d'où

$$Y_2 = P_2(A) g_2(A) Y_2 + Q_2(A) \dots g_k(A) Y_2$$

et par conséquent

$$g_2(A) [Y_2 - P_2(A) g_2(A) Y_2] = 0$$

ce qui nous conduit à

$$Y_2 = X_2 + Y_3$$

(1) Nous utilisons encore le fait que $f_1(\lambda)$ d'une part, le produit $f_2(\lambda) f_3(\lambda) \dots f_k(\lambda)$ d'autre part, sont premiers entre eux.

X_2 appartenant à S_2 , Y_3 étant de la forme

$$Y_3 = g_1(A) g_2(A) Z_3$$

et ainsi de suite.

En continuant ainsi, on voit que tout vecteur X est de la forme

$$X = X_1 + X_2 + \dots + X_k$$

$X_1, X_2, \dots, X_k$ appartenant à $S_1, S_2, \dots, S_k$ respectivement.

Il ne peut exister aucune relation linéaire entre des vecteurs $X_1, X_2, \dots, X_k$. Soit en effet, si de pareilles relations existaient,

$$X_t + \dots + X_h = 0 \quad (t < h)$$

la plus courte d'entre elles (h minimum). En multipliant par $g_h(A)$, nous en déduirions une relation de la forme

$$X_t' + \dots + X_{h-s}' = 0 \quad s \geq 1$$

d'où contradiction.

Si X_0 est un vecteur propre attaché à une racine λ_0 du polynôme caractéristique $f(\lambda)$, nous pourrions mettre ce vecteur sous la forme

$$X_0 = X_1 + X_2 + \dots + X_k$$

Nous aurons donc

$$g(A) X_0 = 0$$

ce qui montre que le polynôme typique de X_0 divise $g(\lambda)$. Or ce polynôme typique est $\lambda - \lambda_0$. Donc $\lambda - \lambda_0$ divise $g(\lambda)$.

Nous concluons que :

$f(\lambda)$ est un multiple de $g(\lambda)$, mais ne contient pas d'autres facteurs binômes que $g(\lambda)$. En d'autres termes, $g(\lambda)$ étant de la forme (6), $f(\lambda)$ est de la forme (au signe près)

$$f(\lambda) = (\lambda - \lambda_1)^{\gamma_1} \dots (\lambda - \lambda_k)^{\gamma_k}$$

$$\gamma_1 + \gamma_2 + \dots + \gamma_k = n$$

$$\gamma_i \geq \alpha_i$$

Revenons maintenant à la représentation d'un vecteur quelconque :

Si nous appelons $\beta_1, \beta_2, \dots, \beta_k$ les nombres de vecteurs indépendants figurant dans $S_1, S_2, \dots, S_k$ respectivement, nous aurons

$$\beta_1 + \beta_2 + \dots + \beta_k = n$$

et nous pourrions former une matrice P , non singulière

$$P = [P_{ij}]$$

telle que ses β_1 premières colonnes représentant des vecteurs X_1 , les β_2 colonnes suivantes des vecteurs X_2 et ainsi de suite.

Les n vecteurs colonnes de P étant indépendants, P admet une inverse P^{-1} . Si on forme la matrice $P^{-1}AP$, on voit qu'elle est de la forme :

$$D = \left\{ \bar{T}_1 \bar{T}_2 \bar{T}_3 \dots \bar{T}_k \right\} \cdot A \left[\bar{S}_1 \bar{S}_2 \dots \bar{S}_k \right]$$

dans laquelle $\bar{S}_i$ est une matrice (n, β_i) composée de β_i colonnes, chacune des colonnes représentant un vecteur propre associé à λ_i . $\bar{T}$ représente alors une matrice (β_i, n) , à β_i lignes. La matrice $[S_1 S_2 \dots S_k]$ est la matrice carrée obtenue en juxtaposant, de gauche à droite, les matrices S_i ; la matrice $[\bar{T}_1 \bar{T}_2 \dots \bar{T}_k]$ inverse de la précédente, s'obtient en superposant, du haut vers le bas, les matrices T_i . Le produit

$$\left\{ \bar{T}_1 \bar{T}_2 \dots \bar{T}_k \right\} \left[\bar{S}_1 \bar{S}_2 \dots \bar{S}_k \right] = \left[\bar{T}_i \bar{S}_j \right]$$

est une matrice unité. Nous avons alors

$$\bar{T}_i \bar{S}_j = e_i$$

δ_{ij} étant l'indice de Kronecker, et e_i une matrice unité carrée d'ordre β_i . La matrice D peut alors s'écrire

$$D = \left\{ \bar{T}_1, \bar{T}_2, \dots, \bar{T}_k \right\} \left[A\bar{S}_1, A\bar{S}_2, \dots, A\bar{S}_k \right]$$

$$D = \left[\bar{T}_i A\bar{S}_j \right]$$

Or, d'après la définition des $\bar{S}_j$, $A\bar{S}_j$ est une matrice (n, β_j) dont toutes les colonnes appartiennent au système S_j .

Pour $i \neq j$, nous aurons

$$\bar{T}_i (A\bar{S}_j) = 0$$

puisque $\bar{T}_i \bar{S}_j = 0$. En effet, $\bar{T}_i \bar{S}_j = 0$ signifie que le produit par $\bar{T}_i$ d'une colonne quelconque de $\bar{S}_j$ est nul. Or chacune des colonnes de $A\bar{S}_j$ est un vecteur du système S_j , et par conséquent une combinaison linéaire des colonnes de $\bar{S}_j$.

D prend donc la forme

$$D = \left[\delta_{ij} A_i \right]$$

avec

$$A_i = \bar{T}_i A \bar{S}_i$$

A prend alors la forme

$$A = P \begin{bmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_k \end{bmatrix} P^{-1}$$

dans laquelle les A_i représentent des matrices (β_i, β_i) , et les zéros des matrices de zéros.

La réduction que nous venons de faire s'appelle première réduction à la forme canonique.

Cette réduction ne modifie pas le polynôme caractéristique. En effet :

$$f(\lambda) = |A - \lambda E| = |P [P^{-1} A P - \lambda E] P^{-1}| = |D - \lambda E|$$

puisque

$$|P| \cdot |P^{-1}| = |E| = 1.$$

On remarque par ailleurs que

$$f(\lambda) = |D - \lambda E| = |A_1 - \lambda e_1| |A_2 - \lambda e_2| \dots |A_k - \lambda e_k|$$

d'après une règle bien connue du calcul des déterminants.

Nous avons donc

$$f(\lambda) = f_1(\lambda) f_2(\lambda) \dots f_k(\lambda)$$

$f_1, f_2, \dots, f_k$ étant les polynômes caractéristiques de $A_1, A_2, \dots, A_k$ respectivement.

$A_1, A_2, \dots, A_k$ sont tels que

$$g_i(A_i) = 0$$

En effet :

$$g_i(A_i) = g_i(\bar{T}_i A \bar{S}_i) = \bar{T}_i g_i(A) \bar{S}_i = 0$$

$g_i(\lambda)$ divise donc $f_i(\lambda)$.

Comme $f_i(\lambda)$ ne peut contenir d'autres facteurs que ceux de $g_i(\lambda)$, nous aurons ainsi (au signe près) :

$$f_i(\lambda) = (\lambda - \lambda_i)^{\beta_i}$$

En résumé :

Une matrice A ayant un polynôme caractéristique de la forme (au signe près) :

$$f(\lambda) = \prod (\lambda - \lambda_i)^{\beta_i}$$

son polynôme typique est de la forme

$$g(\lambda) = \prod (\lambda - \lambda_i)^{\alpha_i} \quad \alpha_i \leq \beta_i$$

Si on considère les familles de vecteurs S_i définies par

$$(A - \lambda_i E)^{\alpha_i} X = 0$$

il existe dans chacune de ces familles β_i vecteurs indépendants. Ces familles sont indépendantes entre elles.

On peut mettre A sous la forme

$$A = P \begin{bmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & A_k \end{bmatrix} P^{-1}$$

les A_i étant des matrices carrées; A_i a pour polynôme typique $(\lambda - \lambda_i)^{\alpha_i}$ et pour polynôme caractéristique $(\lambda - \lambda_i)^{\beta_i}$.

Cette réduction s'appelle première réduction à la forme canonique. Pour aller plus loin, il faut procéder à la réduction des matrices A_i elles-mêmes (ce que nous ne ferons pas ici).

CHAPITRE V

MATRICES CARRÉES PARTICULIÈRES

1.- TRANSFORMATIONS DE MATRICES

Etant donnée une matrice A , une matrice de la forme :

$$(1) \quad B = P A Q \quad (P, Q \text{ non singulières})$$

est dite équivalente à A .

Nous rappelons qu'on appelle transposée de A la matrice obtenue en échangeant dans A les lignes et les colonnes; on représente cette matrice par A' , de sorte que :

$$A = [a_{ij}] \quad A' = [a'_{ij}] \quad a'_{ij} = a_{ji}$$

On montre sans peine que la règle de transposition d'un produit est :

$$(A_1 A_2 \dots A_n)' = A_n' A_{n-1}' \dots A_1'$$

d'où l'on déduit que :

$$(A^{-1})' = (A')^{-1}$$

Lorsque dans (1) on a $P Q = E$, la transformation est une similitude. A et B sont dites transformées l'une de l'autre. Nous avons ainsi :

$$(2) \quad B = P^{-1} A P \quad A = P B P^{-1}$$

La transformée d'un produit est égale au produit des transformées :

$$P^{-1} A_1 A_2 P = (P^{-1} A_1 P) \cdot (P^{-1} A_2 P)$$

Si dans (1) on a $P = Q'$, on dit que la transformation est congruente.

Nous avons alors :

$$(3) \quad B = Q' A Q$$

de sorte que

$$A = (Q^{-1})' B Q^{-1}$$

Une transformation à la fois congruente et de similitude est une transformation orthogonale.

Quand on a affaire à des éléments complexes, on définit les matrices A^* et A^+ par :

$$a_{ij}^* = \text{nombre conjugué de } a_{ij}$$

$$A^+ = (A^*)'$$

Une transformation telle que :

$$B = Q^+ A Q$$

est dite conjunctive.

Une transformation à la fois conjunctive et de similitude est dite unitaire.

2.- MATRICES PARTICULIÈRES

Nous donnons ci-dessous une liste de matrices de formes particulières qui reviennent souvent dans les applications :

Symétrie	$A = A'$
Symétrie gauche	$A = -A'$
Pseudosymétrie	$A + A' = 2E$
Réelle	$A = A^*$
Imaginaire pure	$A + A^* = 0$
Hermitienne	$A = A^+ (= A'^*)$
Hermitienne gauche	$A = -A^+$
Pseudohermitienne	$A + A^+ = 2E$
Orthogonale	$A' = A^{-1}$
Unitaire	$A = (A^+)^{-1}$

Dans le domaine réel, les matrices hermitiennes et unitaires deviennent symétriques et orthogonales.

3.- FORME GÉNÉRALE D'UNE MATRICE UNITAIRE

Etant donné un vecteur complexe :

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

on définit X^+ comme le transposé et conjugué

$$X^+ = [x_1^* \ x_2^* \ \dots \ x_n^*]$$

et le carré de la "longueur" de X comme

$$|X|^2 = X^+ X = \sum |x_i|^2$$

Ceci posé, on appelle transformation unitaire (de vecteurs) une transformation conservant les longueurs, c'est-à-dire telle que

$$\text{si } Y = A X \quad |Y|^2 = X^2$$

ou en utilisant les transpositions

$$(4) \quad Y^+ Y = X^+ X.$$

Nous déduisons de (4) que :

$$X^+ A^+ A X = X^+ X$$

quel que soit X ; c'est-à-dire en explicitant

$$(5) \quad \sum a_{hi}^* a_{hj} = \delta_{ij}, \text{ ou } A^+ A = E = A A^+$$

ce qui est conforme à une terminologie déjà introduite.

Nous allons montrer que toute matrice unitaire A est de la forme :

$$(6) \quad A = K^+ K^{-1}$$

et réciproquement.

Soit donc A unitaire, satisfaisant par conséquent à :

$$(7) \quad A A^+ = E$$

Cherchons à résoudre le système

$$(8) \quad A = K^+ K^{-1}$$

$$K + K^+ = 2E$$

Nous résolvons (8) en tirant K^+ de la deuxième équation, et portant dans la première, ce qui donne

$$A = (2E - K) K^{-1} = 2K^{-1} - E$$

d'où

$$(9) \quad K = 2(A + E)^{-1}$$

La solution du système (8), si elle existe, est donc (9). Il reste à vérifier que (9) satisfait à (8).

Nous supposons $A + E$ non singulière. K et K^+ sont donc non singulières. Le système (8) est alors équivalent à :

$$(8') \quad \begin{cases} (K^+)^{-1} A = K^{-1} \\ E + A = 2K^{-1} \end{cases}$$

tandis que (9) donne :

$$(9') \quad \begin{cases} K^{-1} = \frac{1}{2}(A + E) \\ (K^+)^{-1} = \frac{1}{2}(A^+ + E) \end{cases}$$

Sous la forme (8') et (9'), la vérification que K , déterminé par (9'), satisfait à (8'), est évidente, tenu compte de (7).

Supposons maintenant A défini par (8). La deuxième équation (8) montre que K et K^+ sont permutables, puisque K s'exprime sous la forme :

$$K^+ = 2E - K$$

Les inverses de K et K^+ sont également permutables entre elles et avec K et K^+ . On voit alors immédiatement que :

$$A^+ A = (K^{-1})^+ K K^+ K^{-1} = (K^{-1})^+ K^+ K K^{-1} = E$$

Le raisonnement est en défaut si $A + E$ est une matrice singulière.

La forme

$$A = K^+ K^{-1} = K^{-1} K^+$$

où K est une matrice pseudohermitienne complètement arbitraire, est donc la forme générale des matrices unitaires.

Par exemple, la matrice unitaire d'ordre 2 la plus générale s'obtient comme suit :

$$K = \begin{bmatrix} 1 & a \\ -a & 1 \end{bmatrix} \quad K^{-1} = \begin{bmatrix} 1 & -a \\ a^* & 1 \end{bmatrix} \quad \frac{1}{1 + |a|^2}$$

$$A = K^+ K^{-1} = \begin{bmatrix} 1 & -aa^* & -2a \\ 2a^* & 1-aa^* \end{bmatrix} \quad \frac{1}{1 + |a|^2}$$

Elle dépend d'un paramètre complexe arbitraire a .

D'une façon générale, la matrice unitaire d'ordre n dépend de $\frac{n(n-1)}{2}$ paramètres arbitraires (complexes).

Dans le domaine réel, la théorie s'applique aux matrices orthogonales, dont la forme générale est :

$$A = K' K^{-1} \quad \text{avec } K + K' = 2E \text{ (pseudo-symétrique)}$$

A doit satisfaire à $AA' = E$, parce qu'une transformation orthogonale doit conserver les distances. La transformation orthogonale la plus générale entre les vecteurs (x, y, z) et (X, Y, Z) peut donc se définir par :

$$x = u - rv + qw$$

$$X = u + rv - qw$$

$$y = ru + v - pw$$

$$Y = -ru + v + pw$$

$$z = -qu + pv + w$$

$$Z = qu - pv + w$$

en éliminant u, v, w (p, q, r sont des paramètres arbitraires).

CHAPITRE VI

DES PROBABILITÉS CONDITIONNELLES

Nous indiquerons dans ce chapitre les principes de Calcul des Probabilités conditionnelles, de manière que le lecteur puisse s'y reporter à propos de la théorie des matrices de corrélation.

1. - PROBABILITÉ CONDITIONNELLE D'UN ÉVÈNEMENT

Toute probabilité peut être considérée comme conditionnelle en ce sens qu'elle est calculée sous la condition que certaines hypothèses sont satisfaites ; cela a été mis en évidence dans l'écriture :

$$\text{Pr. } \left\{ E/H \right\} = p$$

Mais il arrive souvent que dans le problème, on pose des hypothèses générales, que nous représenterons par exemple par la lettre H, et que l'on examine ce qui se passe lorsque ces hypothèses se trouvent modifiées. Si nous ne considérons que les hypothèses H d'une part, et les différents événements susceptibles de se produire $E_1, E_2, \dots$ d'autre part, la seule modification à H que nous pouvons faire intervenir est de supposer que l'un des événements $E_1, E_2, \dots$ s'est produit. Supposons par exemple que E_1 se soit produit ; l'information que nous possédons sur E_2 par exemple consiste en H et la connaissance de l'apparition de E_1 ; nous noterons cette information par :

$$H \text{ et } E_1, \text{ ou plus brièvement } HE_1, \quad (1)$$

et nous déduirons de cette nouvelle information, une nouvelle valeur de la probabilité de E_2 , que nous noterons :

$$\text{Pr. } \left\{ E_2 / HE_1 \right\} \quad (2)$$

Cette valeur sera en général différente de la valeur ancienne :

$$\text{Pr. } \left\{ E_2 / H \right\} \quad (3)$$

et sera appelée probabilité conditionnelle de E_2 si E_1 s'est produit.

On démontre dans les Traités de Calcul des Probabilités, la formule des probabilités composées :

$$\text{Pr. } \{ E_1 E_2 / H \} = \text{Pr. } \{ E_1 / H \} \times \text{Pr. } \{ E_2 / HE_1 \} \quad (4)$$

où $E_1 E_2$ est une abréviation pour "l'évènement E_1 et E_2 ".

La formule (3) permet de calculer l'une quelconque des probabilités qui y figurent connaissant les deux autres.

Si la probabilité (2) est beaucoup plus voisine de (1) que la probabilité (3), on peut être tenté de conclure que E_2 est une cause de E_1 ; cette conclusion n'a pas de fondement ; en effet, la probabilité de (2) fait intervenir l'apparition de E_1 , qui peut être négligée que l'existence d'une cause commune à E_1 et E_2 ; nous allons d'ailleurs procéder à un calcul qui montre quelle symétrie on peut introduire entre E_1 et E_2 .

Faisons le quotient de (2) par (3), nous obtenons, en tenant compte de (4), et en échangeant ensuite E_1 et E_2 .

$$\frac{\text{Pr. } \{ E_2 / HE_1 \}}{\text{Pr. } \{ E_2 / H \}} = \frac{\text{Pr. } \{ E_1 / HE_2 \}}{\text{Pr. } \{ E_1 / H \}} = \frac{\text{Pr. } \{ E_1 E_2 / H \}}{\text{Pr. } \{ E_1 \} \cdot \text{Pr. } \{ E_2 \}} = \Delta \quad (5)$$

Nous voyons que l'apparition de E_1 modifie la probabilité de E_2 dans la même proportion que l'apparition de E_2 modifie la probabilité de E_1 . Il n'y a donc pas lieu, même si Δ est grand, de considérer E_1 comme cause de E_2 plutôt que E_2 comme cause de E_1 , à moins naturellement que d'autres considérations ne viennent appuyer cette affirmation.

Si $\Delta > 1$, comme l'apparition d'un des évènements E_1 ou E_2 favorise l'apparition de l'autre, nous dirons qu'il existe entre E_1 et E_2 une corrélation positive ; si $\Delta < 1$, l'apparition d'un évènement défavorise l'apparition de l'autre, nous dirons que la corrélation est négative.

Si $\Delta = 0$, les évènements sont indépendants, et la corrélation est dite nulle. Il ne faut pas perdre de vue que la grandeur et le signe de la corrélation peuvent dépendre essentiellement de H .

2. - ÉVÈNEMENTS INDÉPENDANTS

Dans le cas d'indépendance de E_1 et E_2 , la formule (4) se réduit à :

$$\text{Pr. } \{ E_1 E_2 / H \} = \text{Pr. } \{ E_1 / H \} \times \text{Pr. } \{ E_2 / H \} \quad (6)$$

On peut envisager l'indépendance de deux points de vue ; d'un point de vue abstrait et formel, où elle s'exprime par la formule (6), ou d'un point concret : deux évènements sont indépendants quand il n'existe aucune cause commune ou aucune liaison matérielle entre eux. Ces deux points de vue coïncident quand les probabilités sont supposées déterminées ; ils peuvent diverger quand ce n'est pas le cas, comme le montre l'exemple suivant.

Soit une pièce de monnaie telle que l'apparition de face, si on jette la pièce, ait une probabilité p , qui puisse prendre les valeurs p_1 et p_2 avec les probabilités respectives $\bar{\omega}_1$ et $\bar{\omega}_2$.

La probabilité d'obtenir face après un premier jet est :

$$\omega_1 p_1^2 + \omega_2 p_2 \quad (7)$$

La probabilité d'obtenir face deux deux jets successifs est :

$$\omega_1 p_1^2 + \omega_2 p_2^2 \neq (\omega_1 p_1 + \omega_2 p_2)^2$$

Nous voyons que les deux coups ne peuvent être considérés comme indépendants. La probabilité d'obtenir face au deuxième coup est en effet égale à (7) si l'on ne sait pas le résultat du premier coup, et à :

$$\frac{\omega_1 p_1^2 + \omega_2 p_2^2}{\omega_1 p_1 + \omega_2 p_2}$$

si le premier coup a donné face. Il existe une corrélation entre les événements, bien qu'il n'existe entre eux aucune liaison matérielle, et encore moins de relation de cause à effet. Cette corrélation tient à ce que le résultat du premier jet nous apporte une information nouvelle avant de porter des jugements de causalité.

3. - DISTRIBUTIONS CONDITIONNELLES DE VARIABLES ALÉATOIRES

Considérons maintenant des variables aléatoires, c'est-à-dire des grandeurs numériques $X_1, X_2, \dots, X_n$, telles que, quelles que soient les valeurs $X_1, X_2, \dots, X_n$, l'événement E qui a lieu si les inégalités :

$$X_1 < 1, X_2 < X_2, \dots, X_n < X_n \quad (8)$$

ont lieu a une probabilité bien déterminée :

$$\text{Pr. } \{ E \} = \text{Pr. } \{ X_1 < X_1, \dots \} = p(x_1, x_2, \dots, x_n) \quad (9)$$

Cette probabilité est une fonction des variables x_i elle n'est naturellement définie que dans un champ d'hypothèses H , que nous sous-entendons, mais n'inscrivons pas dans nos formules pour ne pas surcharger l'écriture. La fonction P s'appelle fonction de distribution cumulative des variables X_i . Si cette fonction peut se mettre sous la forme :

$$P = \int^{x_1} \dots \int^{x_n} p(\xi_1, \dots) d\xi_1 \dots d\xi_n \quad (10)$$

la fonction p sera appelée fonction de distribution différentielle. Cette fonction s'interprète par l'inégalité :

$$\text{Pr. } \{ x_1 < X_1 < x_1 + dx_1, \dots \} = p(x_1, \dots) dx_1 \dots dx_n \quad (11)$$

Nous omettrons par la suite les épithètes cumulative et différentielle ; pour éviter toute confusion, nous ferons suivre quand le doute sera possible les fonctions de distribution différentielle du signe différentiel.

4. - DISTRIBUTIONS RELATIVES A DEUX VARIABLES CONTINUES

Dans le cas de deux variables continues, nous aurons, à partir de la

fonction de distribution :

$$\text{Pr. } \left\{ x_1 < X_1 < x_1 + dx, x_2 < X_2 < x_2 + dx_2 \right\} = p_{12}(x_1, x_2) dx_1 dx_2 \quad (12)$$

1° - les fonctions de distribution marginales :

$$\left\{ \begin{array}{l} \text{Pr. } \left\{ x_1 < X_1 < x_1 + dx_1 \right\} = dx_1 \times \int p_{12} dx_2 \\ \qquad \qquad \qquad = p_1(x_1) dx_1 \\ \text{Pr. } \left\{ x_2 < X_2 < x_2 + dx_2 \right\} = dx_2 \times \int p_{12} dx_1 \\ \qquad \qquad \qquad = p_2(x_2) dx_2 \end{array} \right. \quad (13)$$

2° - les fonctions de distribution conditionnelles :

$$\left\{ \begin{array}{l} \text{Pr. } \left\{ x_1 < X_1 < x_1 + dx_1 / X_2 = x_2 \right\} = \frac{p_{12}}{p_2} dx_1 \\ \qquad \qquad \qquad = p_{1,2}(x_1; x_2) dx_1 \\ \text{Pr. } \left\{ x_2 < X_2 < x_2 + dx_2 / X_1 = x_1 \right\} = \frac{p_{12}}{p_1} dx_2 \\ \qquad \qquad \qquad = p_{2,1}(x_2; x_1) dx_2 \end{array} \right. \quad (14)$$

La corrélation entre deux valeurs x_1 et x_2 est indiquée par le rapport :

$$\Delta(x_1, x_2) = \frac{p_{12}}{p_1} = \frac{p_{2,1}}{p_2} = \frac{p_{12}}{p_1 p_2} \quad (15)$$

Les variables sont indépendantes si $\Delta = 1$ quelles que soient les valeurs x_1 et x_2 ; p_{12} s'exprime alors en fonction des seules distributions marginales :

$$p_{12} = p_1 p_2 \quad (16)$$

5. - DISTRIBUTIONS RELATIVES A TROIS VARIABLES

Pour plus de deux variables, les formules se compliquent un peu ; commençons par le cas de trois variables, que nous supposons à distribution continue (les formules relatives à des distributions discontinues se déduisent des formules que nous allons obtenir en remplaçant les $\int$ par des $\sum$).

Appelant $p_{123} dx_1 dx_2 dx_3$ la fonction de distribution de X_1, X_2, X_3 ;

$$\text{Pr. } \left\{ x_1 < X_1 < x_1 + dx_1 \dots \right\} = p_{123} dx_1 dx_2 dx_3 \quad (17)$$

nous obtiendrons :

1° - La distribution marginale de X_1 :

$$\text{Pr. } \left\{ x_1 < X_1 < x_1 + dx_1 \right\} = dx_1 \iint p_{123} dx_2 dx_3 = p_1(x_1) dx_1 \quad (18)$$

2° - La distribution marginale de X_1 et X_2 :

$$\begin{aligned} \text{Pr. } \left\{ x_1 < X_1 < x_1 + dx_1, x_2 < X_2 < x_2 + dx_2 \right\} \\ = dx_1 dx_2 \int p_{123} dx_1 = p_{12}(x_1, x_2) dx_1 dx_2 \end{aligned} \quad (19)$$

3° - La distribution conditionnelle de X_1 connaissant X_2 :

$$p_{1.2}(x_1; x_2) dx_1 = \frac{p_{12}}{p_2} dx_1 \quad (20)$$

(où p_2 est la distribution marginale de X_2).

4° - La distribution conditionnelle de X_1 connaissant X_2 et X_3 :

$$p_{1.23}(x_1; x_2, x_3) dx_1 = \frac{p_{123}}{p_{23}} dx_1 \quad (21)$$

où p_{23} est la distribution marginale de X_2 et X_3 .

5° - La distribution conditionnelle de X_1 et X_2 connaissant X_3 :

$$p_{12.3}(x_1, x_2, x_3) dx_1 dx_2 = \frac{p_{123}}{p_3} dx_1 dx_2 \quad (22)$$

où p_3 est la distribution marginale de X_3 .

Les cinq catégories de distributions pour chacune desquelles nous venons de donner un exemple, comprennent chacune un certain nombre de cas, à savoir :

1° - trois distributions marginales à une variable (X_1, X_2, X_3).

2° - trois distributions marginales à deux variables : ($X_2 X_3; X_3 X_1; X_1 X_2$).

3° - six distributions conditionnelles à une variable supposant connue une autre variable.

4° - trois distributions conditionnelles à une variable supposant les deux autres connues.

5° - trois distributions conditionnelles à deux variables supposant la troisième connue.

Cela fait en tout six distributions marginales et douze distributions conditionnelles.

Pour distinguer ces différentes distributions nous conviendrons d'adopter les conventions suivantes :

- une distribution marginale est affectée des indices des variables qui y figurent.

- une distribution conditionnelle est affectée de deux groupes d'indices, séparés par un point ; les indices du premier groupe sont ceux des variables auxquelles la distribution est relative ; les indices du deuxième

groupe sont ceux des variables dont la valeur est supposée connue.

La distribution relative aux trois variables X_1, X_2, X_3 étant p_{123} , nous obtenons les distributions marginales par intégration :

$$\begin{aligned} p_1 &= \int p_{123} dx_2 dx_3 & p_2 &= \int p_{123} dx_3 dx_1 \dots \\ p_{12} &= \int p_{123} dx_3 & p_{23} &= \int p_{123} dx_1 \dots \end{aligned} \quad (23)$$

et les distributions conditionnelles par formation de quotients :

$$\begin{aligned} p_{12} &= \frac{p_{12}}{p_2} & p_{1,3} &= \frac{p_{13}}{p_3} \dots \\ p_{1,23} &= \frac{p_{123}}{p_{23}} & p_{2,31} &= \frac{p_{123}}{p_{31}} & p_{12,3} &= \frac{p_{123}}{p_3} \end{aligned} \quad (24)$$

6. - DISTRIBUTIONS RELATIVES A UN NOMBRE QUELCONQUE DE VARIABLES

Nous sommes maintenant préparés à traiter le cas d'un nombre quelconque n de variables aléatoires.

Soient $X_1, X_2, \dots, X_n$ ces variables et :

$$p_\mu (x_1, x_2, \dots, x_n) dx_1, \dots, dx_n \quad \mu = 123 \dots n \quad (25)$$

leur fonction de distribution, μ étant une abréviation pour l'exemple des indices 1 à n .

Soit α une combinaison quelconque d'indices pris parmi l'ensemble :

$$\alpha = (i_1 i_2 \dots i_m)$$

La distribution marginale des variables correspondantes sera :

$$p_\alpha (dx)_\alpha = (dx)_\alpha \int p_\mu dx_\mu \quad \alpha \quad (27)$$

(dx) représentant le produit des différentielles :

$$dx_{i_1} dx_{i_2} \dots dx_{i_m} \quad (28)$$

et $(dx)_\mu \alpha$ représentant le produit des différentielles de celles parmi les variables X_i dont l'indice ne figure pas dans la combinaison α .

Quant aux distributions conditionnelles, voici comment on les obtient.

Soient α et β deux combinaisons d'indices pris dans l'ensemble ; ces deux combinaisons sont supposées ne présenter aucun indice commun.

La fonction de distribution marginale des variables affectées d'indices figurant dans α ou β est :

$$P_{\alpha\beta} = \int p_\mu dx_\mu \quad \alpha\beta \quad (29)$$

Si nous appelons $p_{\alpha,\beta}(dx)_\alpha$ la distribution conditionnelle des variables dont les indices appartiennent à α , quand les variables dont les indices appartiennent à β , sont supposées connues, nous obtiendrons, d'après le théorème des probabilités composées :

$$p_{\alpha,\beta} = \frac{P_{\alpha\beta}}{P_\beta} \text{ et par conséquent } p_{\alpha\beta} = \frac{P_{\alpha\beta}}{P_\beta} \quad (30)$$

Telle est la formule générale donnant les distributions conditionnelles. Les formules (32) fournissent des exemples d'application de cette formule générale. Donnons encore un exemple pour fixer les idées : soient les cinq variables X_1, X_2, X_3, X_4, X_5 , et soit à trouver la distribution conditionnelle de X_2 et X_4 connaissant X_3 et X_5 . Partant de la fonction de distribution :

$$P(x_1, \dots, x_5) dx_1 \dots dx_5$$

(nous omettons $\mu = 12345$ pour simplifier l'écriture), nous aurons d'après (29) :

$$P_{2345} = \int p dx_1 \quad P_{35} = \int p dx_1 dx_2 dx_4$$

et d'après (31)

$$P_{24.35} = \frac{P_{2345}}{P_{35}}$$

$P_{24.35}$ ayant pour signification :

$$\text{Pr. } \left\{ x_2 < X_2 < x_2 + dx_2, x_4 < X_4 < x_4 + dx_4 / X_3 = x_3, X_5 = x_5 \right\} \\ = P_{24.35} dx_2 dx_4$$

Les indications précédemment données suffisent à déterminer les distributions conditionnelles dans le cas d'un nombre quelconque de variables. Considérons maintenant la question de la corrélation ; d'une manière générale, soient n variables ; dans l'ensemble des indices, considérons deux groupes distincts α et β : nous nous proposons d'étudier la corrélation entre les variables dont les indices appartiennent à α , que nous appellerons brièvement les variables du groupe α , ou plus brièvement encore, les variables de α et les variables du groupe β .

a Formons le rapport de la distribution conditionnelle des variables du groupe α (quand les variables du groupe β sont connues) à leur distribution marginale. Nous obtenons ainsi :

$$\Delta_{\alpha,\beta} = \frac{P_{\alpha,\beta}}{P_\alpha} = \frac{P_{\alpha\beta}}{P_\alpha P_\beta} \quad (32)$$

$\Delta_{\alpha,\beta}$ caractérisera la corrélation multiple entre les deux groupes de variables α et β . $\Delta_{\alpha,\beta}$ est une fonction des groupes des variables $(x)_\alpha$ et $(x)_\beta$.

Considérons maintenant les variables n'appartenant ni au groupe α ni au groupe β ; elles forment un groupe γ (qui peut ne comprendre aucune variable si μ se partage en α et β). Si nous formons le quotient de la distribution conditionnelle des variables du groupe α (connaissant les valeurs

des variables β et γ) à leur distribution conditionnelle connaissant seulement la valeur des variables γ , nous obtenons le rapport :

$$\delta_{\alpha,\beta}(\gamma) = \frac{P_{\alpha,\beta\gamma}}{P_{\alpha,\gamma}} = \frac{P_{\mu} P_{\gamma}}{P_{\beta\gamma} P_{\alpha\gamma}} \quad (33)$$

qui caractérise la corrélation partielle entre les groupes α et β .

Nous remarquerons que dans (41) comme dans (40) on peut permuter les groupes d'indices α et β , par suite de la symétrie de ces expressions.

7. - INDÉPENDANCE MARGINALE, INDÉPENDANCE PARTIELLE DE DEUX GROUPES DE VARIABLES

Deux groupes de variables α et β sont considérés comme indépendants si, quelles que soient les valeurs supposées connues des variables du groupe complémentaire γ , il n'existe aucune corrélation entre les deux groupes α et β . L'indépendance s'exprime donc en égalant à 1 le quotient δ , ce qui conduit à :

$$P_{\mu} = \frac{P_{\beta\gamma} P_{\alpha\gamma}}{P_{\gamma}} = P_{\gamma} P_{\beta,\gamma} P_{\alpha,\gamma} \quad (34)$$

La distribution p_{μ} est alors de la forme d'un produit de deux fonctions des variables du groupe α figurant dans la première, les variables du groupe β dans la seconde.

Remarquons que le fait d'égaliser à 1 le quotient Δ conduit à un résultat différent :

$$P_{\alpha\beta} = P_{\alpha} P_{\beta} \quad (35)$$

et n'exprime l'indépendance des groupes α et β que dans la distribution marginale de cet ensemble de variables.

Nous pourrions dire que (34) caractérise l'indépendance partielle des groupes α et β (dans le groupe qui les comprend tous deux) tandis que (35) caractérise l'indépendance marginale.

L'indépendance marginale et l'indépendance partielle correspondent à deux formes différentes de p_{μ} . La formule (34) nous fournit déjà la décomposition :

$$P_{\mu} = P_{\gamma} P_{\beta,\gamma} P_{\alpha,\gamma} \text{ (indépendance partielle)} \quad (36)$$

Quant à la formule (35), elle nous conduit à la décomposition :

$$P_{\mu} = P_{\alpha} P_{\beta} P_{\gamma,\alpha,\beta} \text{ (indépendance marginale)} \quad (37)$$

Or, de toute manière, nous pourrions écrire pour p_{μ} les décompositions suivantes :

$$P_{\mu} = P_{\gamma} P_{\beta,\gamma} P_{\alpha,\beta,\gamma} \quad (38)$$

$$P_{\mu} = P_{\alpha} P_{\beta,\alpha} P_{\gamma,\alpha,\beta}$$

Comparant (36), (37), (38), nous obtenons des conditions nécessaires et suffisantes d'indépendance :

$$\text{Indépendance partielle : } \begin{matrix} p_{\alpha, \gamma} = p_{\alpha, \beta \gamma} \\ \text{(ou)} \\ p_{\beta, \gamma} = p_{\beta, \alpha \gamma} \end{matrix} \quad (39)$$

$$\text{Indépendance marginale : } \begin{matrix} p_{\beta} = p_{\beta, \alpha} \\ \text{(ou)} \\ p_{\alpha} = p_{\alpha, \beta} \end{matrix} \quad (40)$$

Si l'on cherche à reconnaître sur la forme de p s'il y a indépendance marginale ou partielle, on peut utiliser les propositions suivantes :

La condition nécessaire et suffisante pour que les variables α et β soient indépendantes partiellement est que la fonction de distribution soit de la forme :

$$p = f_{\alpha \gamma} g_{\beta \gamma} \quad (41)$$

où les variables α ne figurent que dans $f_{\alpha \gamma}$, les variables β ne figurent que dans $g_{\beta \gamma}$.

La condition nécessaire et suffisante pour que les variables α et β soient indépendantes marginalement est que la fonction de distribution soit de la forme :

$$p = f_{\alpha} g_{\beta} h_{\alpha \beta \gamma} \quad (42)$$

où f_{α} ne dépend que des variables α , g_{β} ne dépend que des variables β et où h est une fonction telle que son intégrale prise par rapport aux variables γ soit une constante (indépendante des variables α et β).

Que les conditions ci-dessus énoncées soient nécessaires, ceci est évident au seul examen de (36) et (37). Nous allons démontrer qu'elles sont suffisantes dans le cas de trois variables ;

Le cas de plus de trois variables se traite de la même manière, en substituant le cas échéant des intégrations multiples aux intégrales simples.

Soient donc trois variables telles que :

$$p(x_1, x_2, x_3) = f(x_1, x_3) g(x_2, x_3) \quad (43)$$

Nous avons :

$$p_{13} = f \int f dx_2 \quad p_{23} = g \int f dx_1 \quad p_3 = \int f x_1 \int g dx_2$$

L'inégalité (34) est vérifiée. Il y a bien indépendance partielle de X_1 et X_2 .

Soient maintenant trois variables telles que :

$$p(x_1, x_2, x_3) = f(x_1) g(x_2) h(x_1, x_2, x_3) \quad (44)$$

avec :

$$\int h dx_3 = K \quad (K \text{ constante indépendante de } x_1 \text{ et } x_2).$$

Nous voyons immédiatement que :

$$p_1 = K f \int f dx_2$$

$$p_2 = K g \int f dx_1$$

$$p_{12} = K f g \int \int \int p dx_1 dx_2 dx_3 = K \int f dx_1 \int g dx_2 = 1$$

d'où l'égalité (35).

Considérons pour fixer les idées, une distribution de la forme :

$$p = A \exp. \left\{ - (x_1^2 + x_2^2 + x_3^2 - 2 a_1 x_2 x_3 - 2 a_2 x_3 x_1 - 2 a_3 x_1 x_2) \right\} \quad (45)$$

Cette distribution appartient à une classe de distribution à laquelle nous consacrerons une étude particulière, sous le nom de distributions normales. A est une constante dont la valeur est choisie de manière que :

$$\int p \, dx_1 \, dx_2 \, dx_3 = 1$$

La forme quadratique entre crochets est positive quelles que soient les valeurs de x_1, x_2, x_3 (ce qui impose à a_1, a_2, a_3 des conditions sur lesquelles nous reviendrons plus tard).

Nous voyons sur (45) que la condition pour que X_1 et X_2 soient partiellement indépendantes est :

$$a_3 = 0 \quad (46)$$

p est alors de la forme (43).

Si nous intégrons (45) en x_3 , nous obtenons, comme distribution marginale de x_1 et x_2 :

$$p_{12} = B \exp. \left\{ - \left[x_1^2 + x_2^2 - 2 a_3 x_1 x_2 - (a_1 x_2 + a_2 x_1)^2 \right] \right\}$$

où B est une nouvelle constante. Nous avons là une expression de la forme $f(x_1) g(x_2)$ si :

$$a_1 a_2 + a_3 = 0 \quad (47)$$

Telle est la condition d'indépendance marginale, qui est tout à fait différente de la condition (46). Nous vérifions facilement que si (47) est vérifiée, p peut s'écrire :

$$p = A \exp. \left\{ - (1 - a_2^2) x_1^2 \right\} \exp \left\{ - (1 - a_1^2) x_2^2 \right\} \exp \left\{ - \left[x_3 - (a_1 x_2 + a_2 x_1) \right]^2 \right\}$$

décomposition qui est bien de la forme (44).

8. - INDÉPENDANCE COMPLÈTE

Etant donnés les deux groupes de variables α et β , nous pourrions envisager une troisième sorte d'indépendance caractérisée par le fait que p est de la forme :

$$p_{\mu} = p_{\alpha} p_{\beta} p_{\gamma}$$

Nous dirons alors qu'il y a indépendance complète entre les variables α et les variables β . L'indépendance complète entraîne naturellement l'indépendance partielle et l'indépendance marginale. Il est à remarquer que s'il y a indépendance partielle entre les groupes α et β , les groupes β et γ , et les groupes γ et α , il y a indépendance complète entre ces trois groupes. Nous ne ferons la démonstration que dans le cas de trois variables.

Supposons donc que X_1 et X_2 soient partiellement indépendantes, ainsi que X_2 et X_3 d'une part et X_1 et X_3 d'autre part, nous aurons ainsi :

$$P = P_3 P_{1,3} P_{2,3} = P_1 P_{2,1} P_{3,1} \quad (48)$$

Intégrons en x_2 nous obtenons :

$$P_3 P_{1,3} = P_1 P_{3,1} = P_{13}$$

et par conséquent :

$$P_{2,3} = P_{2,1} = P_{2,31} \quad (49)$$

L'égalité (49) montre que $P_{2,31}$ ne dépend ni de x_1 , ni de x_3 ; nous avons ainsi :

$$P = P_{13} P_{2,31}(x_2)$$

et en intégrant en x_1 et x_3 :

$$P_{2,31} = P_2(x_2) = P_{2,3}$$

les égalités (48) deviennent alors :

$$P = P_2 P_3 P_{1,3} = P_1 P_2 P_{3,1} \quad (50)$$

Tenons compte maintenant de l'indépendance partielle entre X_1 et X_2 . Nous en déduisons les égalités dérivées de (49) par permutation circulaire des indices, en particulier :

$$P_{3,1} = P_3$$

ce qui, nous reportant à (50), nous conduit à l'égalité cherchée.

Dans le cas de la distribution normale, la remarque que nous venons de faire est évidente ; si $a_1 = a_2 = a_3 = 0$, les trois variables sont indépendantes. Dans le cas particulier de la distribution normale, on vérifie que l'indépendance marginale des trois couples : $X_1 X_2$, $X_2 X_3$, $X_3 X_1$ assure l'indépendance complète (par suite des restrictions $|a_1| < 1$, $|a_2| < 1$, $|a_3| < 1$ imposées à a_1 , a_2 , a_3).

CHAPITRE VII

NOTIONS SUR LA THÉORIE DE L'INFORMATION

Considérons un système E qui peut prendre n états $E_1, E_2, \dots, E_n$ avec des probabilités $p_1, p_2, \dots, p_n$. Si ces probabilités sont égales, et ont par conséquent pour valeur commune $1/n$, les n états sont placés exactement sur le même plan, et on peut donc considérer que E peut prendre effectivement ces n états. Si les probabilités sont inégales, on est amené à considérer que E ne prend effectivement qu'un nombre moindre d'états, que l'on peut évaluer de la manière suivante. Supposons que $n = 2$ pour simplifier l'exposé. Le système E peut ainsi prendre deux états avec les probabilités p et q ; supposons $p > q$, et cherchons à définir le nombre effectifs d'état, soit μ que peut prendre E. Pour cela procédons à N expériences.

Les résultats des expériences peuvent prendre 2^N formes différentes dont μ^N effectivement prises. Pour déterminer ce nombre μ supposons que l'on néglige, parmi les 2^N formes possibles, les moins probables, de manière que les probabilités cumulées sur ces formes soient ε . La probabilité que dans N épreuves, il y en ait r où E prend la forme E_1 et $N - r$ où il prend la forme E_2 est :

$$\frac{N!}{r! (N-r)!} p^r q^{N-r}$$

Si nous négligeons les événements les moins probables, de probabilité ε , nous déterminons une limite inférieure pour r , soit R , par la condition :

$$\sum_{r=0}^R \frac{N!}{r! (N-r)!} p^r q^{N-r} = \varepsilon \quad (1)$$

Le nombre des cas sur lesquels se trouvent accumulés les probabilités restantes, de somme $1 - \varepsilon$, est donc donné par la formule :

$$\mu^N = \sum_{r=R+1}^N \frac{N!}{r! (N-r)!} \quad (2)$$

Pour N grand, on peut montrer que $\log \mu$ a la valeur asymptotique :

$$\log \mu = p \log \frac{1}{p} + q \log \frac{1}{q}$$

La même théorie peut s'appliquer à $n > 2$, auquel cas, si n éventualités sont possibles avec probabilités :

$$p_1, p_2, \dots, p_n$$

le nombre effectif d'éventualités réalisables est μ , défini asymptotiquement par :

$$H = \log \mu = - \sum p_i \log p_i$$

La quantité H est appelée entropie du système considéré, ou quantité d'information (*) nécessaire pour connaître l'état du système. Le nom de quantité d'information est choisi à partir des considérations suivantes :

Pour connaître l'état d'un système qui peut prendre deux états également probables, il faut une information.

Pour connaître l'état d'un système qui peut prendre quatre états également probables, il faut deux informations, dont la première choisit entre (1 ou 2) et (3 ou 4), la seconde achevant la détermination.

Pour connaître l'état d'un système qui peut prendre trois états de probabilités $\frac{1}{2}, \frac{1}{3}, \frac{1}{4}$, il faut une information et demie : la première choisit entre (1) et (2 ou 3), la seconde entre (2 ou 3) ; comme la seconde information est inutile une fois sur deux en moyenne, cela fait bien 1,5 unité d'information.

On constate, pour les cas simple examinés, que le nombre d'unités d'information nécessaires est donné par la formule :

$$I = - \sum p_i \log_2 p_i$$

On montre dans le développement de la théorie que ce résultat est général.

La notion d'entropie permet de définir d'une manière simple la dépendance entre deux systèmes. Soient deux systèmes tels qu'il y ait la probabilité p_{ij} pour que le premier soit dans l'état i et le second dans l'état j . La probabilité pour le premier système, considéré isolément, d'être dans l'état i est :

$$p_i = \sum_j p_{ij}$$

La probabilité pour le second d'être dans l'état j est, s'il est considéré isolément :

$$q_j = \sum_i p_{ij}$$

(*) Dans la théorie de l'information, on adopte généralement la base 2 pour les logarithmes.

Le nombre d'états effectivement pris par le premier système a pour logarithme :

$$h_1 = - \sum p_i \log p_i$$

En ce qui concerne le second système, ce logarithme est :

$$h_2 = - \sum q_i \log q_i$$

En ce qui concerne les deux systèmes considérés simultanément, le nombre de couples d'états effectivement réalisable a pour logarithme :

$$H = - \sum p_{ij} \log p_{ij}$$

On montre que, quels que soient les p_{ij}

$$H < h_1 + h_2$$

et que l'on a :

$$H = h_1 + h_2$$

uniquement si p_{ij} est de la forme :

$$p_{ij} = p_i q_j$$

c'est-à-dire si les deux systèmes sont indépendants.

L'indépendance étant caractérisée par :

$$h_1 + h_2 - H = 0$$

on est amené à caractériser la dépendance par la quantité positive :

$$h_1 + h_2 - H$$

Appliquée à deux variables aléatoires X_1, X_2 on est ainsi amené à considérer l'intégrale :

$$\int p_{12} \log \frac{p_{12}}{p_1 p_2} dx_1 dx_2 = \mathcal{N} \log \Delta_{12}$$

Δ_{12} ayant la signification donnée à la page 7.

De même on sera amené à associer à des groupes de variables les coefficients :

$$\mathcal{N} \log \Delta_{\alpha, \beta}, \quad \mathcal{N} \log \delta_{\alpha, \beta}(\gamma)$$

qui sont des coefficients de corrélation multiple et partielle. Ces coefficients sont essentiellement positifs, et s'annulent uniquement dans les cas d'indépendance marginale ou partielle.

CHAPITRE VIII

MATRICES ET CORRÉLATION

1. - CORRÉLATION SIMPLE

Soient deux variables aléatoires X_1 et X_2 ; nous supposons l'existence des valeurs moyennes :

$$\mathcal{M} X_i = a_i \quad (1)$$

$$\mathcal{M} X_i X_j = m_{ij}$$

Si nous substituons à X_1, X_2 les variables centrées :

$$X'_i = X_i - a_i$$

nous obtenons, au lieu de m_{ij} quatre moments centrés.

$$\mu_{ij} = \mathcal{M}' X'_i \times X'_j \quad (2)$$

que nous pouvons ranger dans une matrice carrée symétrique :

$$M_{12} = \begin{bmatrix} \mu_{11} & \mu_{12} \\ \mu_{21} & \mu_{22} \end{bmatrix} \quad (3)$$

Cette matrice est dite matrice des moments attachés aux variables X_1, X_2 ; la forme quadratique :

$$Q_{12}(u_1, u_2) = \mu_{11} u_1^2 + 2 \mu_{12} u_1 u_2 + \mu_{22} u_2^2$$

est la valeur moyenne du carré :

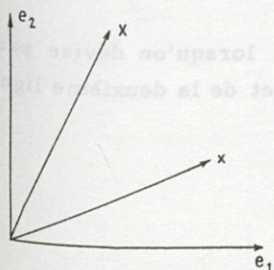
$$(u_1 X'_1 + u_2 X'_2)^2 \quad (4)$$

elle est donc essentiellement positive. Elle ne peut s'annuler que si, pour certaines valeurs u_1, u_2 l'expression (4) est nulle, ce qui entraîne l'existence d'une relation linéaire entre les variables X'_1 et X'_2 . Il s'agit ainsi d'une matrice définie positive.

Pour interpréter cette matrice, supposons donnée une origine dans un plan, et un système de référence constitué par deux vecteurs $\vec{e}_1, \vec{e}_2$, orthogonaux, de longueur égale à l'unité. Le vecteur :

$$\vec{X} = \vec{e}_1 X_1 + \vec{e}_2 X_2$$

a pour extrémité un point aléatoire dans le plan. Soit x un vecteur non aléatoire, de composantes x_1, x_2 .



Nous pouvons lui associer le vecteur y :

$$\vec{y} = \vec{e}_1 y_1 + \vec{e}_2 y_2$$

y_1 et y_2 étant liés à x_1, x_2 par :

$$\begin{cases} y_1 = \mu_{11} x_1 + \mu_{12} x_2 \\ y_2 = \mu_{21} x_1 + \mu_{22} x_2 \end{cases}$$

La transformation ainsi définie, qui change x en y , reste invariante par rotation des axes $\vec{e}_1, \vec{e}_2$ soit en effet un troisième vecteur $\vec{z}$.

$$\vec{z} = \vec{e}_1 z_1 + \vec{e}_2 z_2$$

Formons le produit scalaire $y \times z$, Nous obtenons :

$$\begin{aligned} y \times z &= y_1 z_1 = \mu_{11} x_1 z_1 + \mu_{12} (x_1 z_2 + z_1 x_2) + \mu_{22} x_2 z_2 \\ &= \mathcal{M} \begin{bmatrix} X_1^2 & x_1 z_1 + X_1' X_2' (x_1 z_2 + z_1 x_2) + X_2^2 & x_2 z_2 \end{bmatrix} \\ &= \mathcal{M} \begin{bmatrix} X_1' & x_1 + X_2' x_2 \end{bmatrix} \begin{bmatrix} X_1' & z_1 + X_2' z_2 \end{bmatrix} \end{aligned}$$

d'où l'équation indépendante de $\vec{e}_1$ et $\vec{e}_2$:

$$\vec{y} \times \vec{z} = \mathcal{M} \{ (\vec{X} \cdot \vec{x}) (\vec{X} \cdot \vec{z}) \} \quad (5)$$

La relation (5) quand $\vec{x}$ est donné, doit avoir lieu quel que soit $\vec{z}$. Elle définit ainsi $\vec{y}$ en fonction de $\vec{x}$, ce qui conduit à la transformation :

$$\vec{y} = \mathcal{C} x \quad (6)$$

qui ne dépend pas du système $\vec{e}_1, \vec{e}_2$ choisi, à condition que ce système soit constitué par des vecteurs orthogonaux et de longueur unité (*).

La forme quadratique Q peut également être définie comme fonction du vecteur $\vec{x}$, puisque :

$$Q(x_1, x_2) = \vec{x} \mathcal{C} \vec{x} \quad (7)$$

(*) Il est bien connu que c'est dans de pareils systèmes de référence que la propriété, pour une matrice, d'être symétrique, est invariante.

Considérons maintenant la matrice :

$$S_{12} = \begin{bmatrix} 1 & r_{12} \\ r_{21} & 1 \end{bmatrix}$$

où :

$$r_{12} = r_{21} = \frac{\mu_{12}}{\sqrt{\mu_{11} \mu_{22}}}$$

Cette matrice se confond avec la matrice M lorsqu'on divise par $\sqrt{\mu_{11}}$ $\sqrt{\mu_{12}}$ les éléments de la première ligne et de la deuxième ligne respectivement, ce qui conduit à :

$$\begin{bmatrix} \sqrt{\mu_{11}} & \frac{\mu_{12}}{\sqrt{\mu_{11}}} \\ \frac{\mu_{21}}{\sqrt{\mu_{22}}} & \sqrt{\mu_{22}} \end{bmatrix}$$

puisque l'on fait subir la même opération aux colonnes. On voit ainsi que :

$$M_{12} = \begin{bmatrix} \sqrt{\mu_{11}} & 0 \\ 0 & \sqrt{\mu_{22}} \end{bmatrix} \quad S_{12} = \begin{bmatrix} \sqrt{\mu_{11}} & 0 \\ 0 & \sqrt{\mu_{22}} \end{bmatrix}$$

La matrice S_{12} , dite matrice de corrélation entre X_1 et X_2 ne se rattache pas à une transformation indépendante de e_1 et e_2 . Si nous posons en effet :

$$\omega = \vec{e}_1 \omega_1 + \vec{e}_2 \omega_2$$

avec

$$\omega_1 = x_1 + r_{12} x_2$$

$$\omega_2 = r_{12} x_1 + x_2$$

Nous voyons que :

$$\vec{\omega} = \vec{x} + \vec{u}$$

$$\vec{u} = \vec{e}_1 u_1 + \vec{e}_2 u_2$$

$$u_1 = r_{12} x_2$$

$$u_2 = r_{12} x_1$$

Si la relation entre $\vec{x}$ et $\vec{\omega}$ était indépendante des axes, il en serait de même de la relation entre $\vec{x}$ et $\vec{u}$. Cherchons les vecteurs propres de cette dernière transformation. Ils sont définis par :

$$\frac{u_1}{u_2} = \frac{x_1}{x_2} = \frac{x_2}{x_1}$$

d'où :

$$x_1 = \pm x_2$$

Les vecteurs propres $\vec{x}$ sont bissectrices des axes, donc liés à ces axes.

2. - CORRÉLATIONS A n DIMENSIONS

Soient maintenant n variables aléatoires $X_1, X_2, \dots, X_n$ que nous représentons par un vecteur $\vec{X}$

$$\vec{X} = \vec{e}_1 X_1 + \vec{e}_2 X_2 + \dots + \vec{e}_n X_n$$

dans un espace à n dimensions rapporté à n vecteurs de référence orthogonaux et de longueur 1. Si nous supposons que :

$$\mathcal{M} X_i = 0$$

ce qui peut toujours être obtenu par un changement d'origine, nous pouvons associer à un vecteur aléatoire X la matrice des moments :

$$M_{12\dots n} = [\mu_{ij}] \quad \mu_{ij} = \mathcal{M} X_i X_j \quad (10)$$

Cette matrice peut également s'écrire :

$$M_{12\dots n} = \mathcal{M} [X_i X_j] \quad (11)$$

en introduisant la notion de valeur moyenne d'une matrice (matrice formée en prenant les valeurs moyennes des éléments d'une matrice aléatoire).

La matrice aléatoire :

$$M = [X_i X_j]$$

définit une transformation linéaire aléatoire, qui à tout vecteur $\vec{x}(x_1, x_2, \dots, x_n)$ associe le vecteur aléatoire $\vec{Y}(Y_1, Y_2, \dots, Y_n)$ tel que :

$$Y_i = \sum_h X_i X_h x_h = X_i (\vec{X} \cdot \vec{x})$$

Nous voyons ainsi apparaître la transformation :

$$\vec{Y} = \vec{X} (\vec{X} \cdot \vec{x}) \quad (12)$$

qui associe à un vecteur certain $\vec{x}$ un vecteur aléatoire $\vec{Y}$, colinéaire à $\vec{X}$. Il apparaît clairement sur (12) que la transformation est indépendante des vecteurs $\vec{e}_i$ choisis comme système de référence. Ces vecteurs $\vec{e}_i$ interviennent simplement pour donner un sens précis au produit scalaire; pourvu qu'ils soient orthogonaux et de longueur 1, peu importe, en ce qui concerne (12), leur position dans l'espace. Si nous prenons la valeur moyenne de (12), nous obtenons une transformation, non aléatoire :

$$\vec{y} = \mathcal{M} [\vec{X} (\vec{X} \cdot \vec{x})] = \mathcal{C} \vec{x} \quad (12')$$

de signification indépendante des axes. Cette transformation s'exprime par la matrice des moments. La forme de l'équation (12) montre qu'il suffit, pour que cette transformation ait un sens, que l'on ait défini le produit scalaire de deux vecteurs. La matrice des corrélations $S_{12\dots n}$ se définit comme

$$S_{12\dots n} = [r_{ij}] \quad r_{ij} = \frac{\mu_{ij}}{\sqrt{\mu_{ii} \mu_{jj}}}$$

d'où l'on tire :

$$M_{12 \dots n} = \text{diag.} \left[\sqrt{\mu_{ii}} \right] S_{12 \dots n} \cdot \text{diag.} \left[\sqrt{\mu_{ii}} \right] \quad (14)$$

$$\left| M_{12 \dots n} \right| = \mu_{11} \mu_{22} \dots \mu_{nn} \left| S_{12 \dots n} \right|$$

La forme quadratique :

$$Q(\vec{x}) = \sum_{ij} \mu_{ij} x_i x_j$$

est définie positive, comme ayant pour valeur :

$$Q(\vec{x}) = \vec{x} \cdot \vec{x} = \mathcal{N}(\vec{X} \cdot \vec{x})^2 \quad (15)$$

3. - VARIABLES NORMALES

On appelle variable normale toute variable X dont la distribution à la forme de la loi de Laplace-Gauss :

$$\frac{1}{\sigma \sqrt{2\pi}} \exp. \left\{ -\frac{(x-a)^2}{2\sigma^2} \right\} dx \quad (16)$$

Nous appellerons variables normales $X_1, X_2, \dots, X_n$ tout système de variables aléatoires telles que toute combinaison linéaire de ces variables soit elle-même une variable normale.

Nous pourrions supposer, sans diminuer la généralité de la question, que les variables considérées ont pour valeur moyenne zéro ; si ce n'était pas le cas, il suffirait de leur substituer les variables :

$$X_i' = X_i - \mathcal{N} X_i \quad (17)$$

Rappelons que si une variable normale X a pour distribution :

$$f(x) dx = \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ -\frac{x^2}{2\sigma^2} \right\} dx \quad (18)$$

sa fonction caractéristique est :

$$\varphi(t) = \mathcal{N} e^{itx} = \exp \left\{ -\frac{\sigma^2 t^2}{2} \right\} \quad (19)$$

et que, réciproquement, toute variable qui a pour fonction caractéristique la fonction (20) a pour distribution la fonction (19).

Soient maintenant les variables X_i , centrées, supposées normales et soit leur fonction caractéristique :

$$\begin{aligned} \varphi(t_1, t_2, \dots, t_n) &= \mathcal{N} \exp \left\{ i(t_1 X_1 + \dots + t_n X_n) \right\} \\ &= \mathcal{N} \exp \left\{ i \vec{t} \cdot \vec{X} \right\} = \varphi(\vec{t}) \end{aligned} \quad (20)$$

Introduisons une variable auxiliaire λ :

$$\varphi(\lambda \vec{t}) = \mathcal{N} \exp \left\{ i \lambda (t_1 X_1 + \dots + t_n X_n) \right\}$$

Si $\vec{t}$ reste fixe, $\varphi(\lambda \vec{t})$ considérée comme fonction de λ n'est autre que la fonction caractéristique de la variable $t_1 X_1 + \dots + t_n X_n$. Cette variable combinaison linéaire des X_i , est par hypothèse une variable normale : donc $\varphi(\lambda \vec{t})$ est de la forme (20) :

$$\varphi(\lambda \vec{t}) = \exp \left\{ - \frac{\lambda^2 \sigma^2 (t_1, t_2, \dots, t_n)}{2} \right\} \quad (22)$$

σ^2 n'est autre que la variance de $\sum t_i X_i$. Or :

$$\mathfrak{M}(\sum t_i X_i)^2 = \mathfrak{M}(\vec{t} \cdot \vec{X})^2 = \sum_{ij} \mu_{ij} t_i t_j$$

Faisons $\lambda = 1$, nous obtenons sans peine :

$$\varphi(\vec{t}) = \exp \left\{ - \frac{\sum \mu_{ij} t_i t_j}{2} \right\} \quad (23)$$

Procédons à une rotation des axes de manière que la matrice des variances devienne diagonale. Le vecteur $\vec{t}$ aura les composantes t_i le vecteur $\vec{x}$ les composantes x_i , nous obtiendrons :

$$\varphi(\vec{t}) = \exp \left\{ - \sum_i \frac{\sigma_i^{-2} t_i^2}{2} \right\}$$

La fonction caractéristique étant le produit de n fonctions de $\vec{t}_1, \vec{t}_2, \dots, \vec{t}_n$ respectivement, les variables X_i sont indépendantes ; X_i a pour distribution :

$$\frac{1}{\sigma_i \sqrt{2\pi}} \exp. \left\{ - \frac{x_i^2}{2 \sigma_i^2} \right\} d\bar{x}_i$$

et la fonction de distribution des $\bar{X}_i$ peut s'écrire (*) :

$$\frac{1}{\sigma_1 \sigma_2 \dots \sigma_n} \frac{1}{(\sqrt{2\pi})^n} \exp. \left\{ - \frac{1}{2} \vec{x} \mathcal{C}^{-1} \vec{x} \right\} d\bar{x}_1 \dots d\bar{x}_n \quad (24)$$

en effet la transformation $\mathcal{C}$ de la formule (12') s'écrit ici :

$$y = \mathcal{C} \vec{x} \quad y_i = \sigma_i^{-2} x_i$$

et la transformation inverse $\mathcal{C}^{-1}$ est $x_i = y_i / \sigma_i^{-2}$.

4. - DISTRIBUTION NORMALE - CORRÉLATION NORMALE

Si nous revenons aux variables $x_1, x_2, \dots, x_n$, comme $\mathcal{C}$ et $\bar{X}$ ont des significations invariantes, nous obtenons la distribution :

$$\frac{1}{\bar{\sigma}_1 \bar{\sigma}_2 \dots \bar{\sigma}_n} \frac{1}{(\sqrt{2\pi})^n} \exp. \left\{ - \frac{1}{2} \vec{x} \mathcal{C}^{-1} \vec{x} \right\} dx_1 \dots dx_n \quad (25)$$

(*) Par $\vec{x} \mathcal{C}^{-1} \vec{x}$ nous entendons le produit scalaire du vecteur $\vec{x}$ par le vecteur $\mathcal{C}^{-1} \vec{x}$ déduit de $\vec{x}$ par la transformation $\mathcal{C}^{-1}$ inverse de $\mathcal{C}$.

Il ne nous reste qu'à exprimer d'une manière invariante les $\bar{\sigma}_i$. Considérons le déterminant de $\mathcal{C}$; il a pour expression :

$$|\mathcal{C}| = (\bar{\sigma}_1 \bar{\sigma}_2 \dots \bar{\sigma}_n)^2$$

nous obtenons donc comme type de la distribution normale :

$$|\mathcal{C}|^{-\frac{1}{2}} (2\pi)^{-\frac{n}{2}} \exp. \left\{ -\frac{1}{2} \bar{x} \mathcal{C}^{-1} \bar{x} \right\} dx_1 \dots dx_n \quad (26)$$

où $\mathcal{C}$ est la transformation dont la matrice a pour élément général $\mathcal{M}_{X_i X_j}$.

Quant à la fonction caractéristique, elle s'écrit :

$$\varphi(\vec{t}) = \exp. \left\{ -\frac{1}{2} \vec{t} \mathcal{C} \vec{t} \right\} \quad (27)$$

Nous avons trouvé en (26) la forme nécessaire d'une distribution normale. Il est aisé de démontrer réciproquement que toute combinaison linéaire :

$$Y = a_1 X_1 + a_2 X_2 + \dots + a_n X_n$$

est une variable normale. La fonction caractéristique de Y est en effet :

$$\psi(t) = \mathcal{M} e^{itY} = \varphi(\vec{t}a) = \exp \left\{ -\frac{t^2}{2} \vec{a} \mathcal{C} \vec{a} \right\}$$

ce qui est bien la forme (20).

Quand des variables ont une distribution normale, on dit aussi qu'elles sont en corrélacion normale.

5. - DISTRIBUTION NORMALE A DEUX OU TROIS PARAMÈTRES

Nous allons donner deux applications des formules précédentes. Soient d'abord deux variables X_1 et X_2 que nous supposons centrées et normées. Soit :

$$\mathcal{M}_{X_1 X_2} = r_{12}$$

La matrice des μ_{ij} est ici :

$$M_{12} = \begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix}$$

La matrice inverse est :

$$N_{12} = \frac{1}{1 - r_{12}^2} \begin{bmatrix} 1 & -r_{12} \\ -r_{12} & 1 \end{bmatrix}$$

D'où la distribution :

$$\frac{1}{2\pi} \frac{1}{\sqrt{1-r_{12}^2}} \exp \left\{ -\frac{1}{2(1-r_{12}^2)} [X_1^2 - 2r_{12}x_1x_2 + x_2^2] \right\} dx_1 dx_2 \quad (28)$$

Soient maintenant trois variables X_1, X_2, X_3 centrées et normées; pour simplifier l'écriture, nous adopterons les notations :

$$r_{23} = \mathfrak{M} X_2 X_3 = \rho_1 \quad r_{31} = \mathfrak{M} X_3 X_1 = \rho_2 \quad r_{12} = \mathfrak{M} X_1 X_2 = \rho_3$$

La matrice des moments est ici :

$$S_{123} = M_{123} = \begin{bmatrix} 1 & \rho_3 & \rho_2 \\ \rho_3 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{bmatrix}$$

tandis que la matrice inverse est :

$$\frac{1}{1-(\rho_1^2 + \rho_2^2 + \rho_3^2) + 2\rho_1\rho_2\rho_3} \begin{bmatrix} 1-\rho_1^2; \rho_1\rho_2-\rho_3; \rho_1\rho_3-\rho_2 \\ \rho_1\rho_2-\rho_3; 1-\rho_2^2; \rho_2\rho_3-\rho_1 \\ \rho_1\rho_3-\rho_2; \rho_2\rho_3-\rho_1; 1-\rho_3^2 \end{bmatrix}$$

Par conséquent le vecteur $\mathfrak{C}^{-1} \bar{x}$ a pour composantes :

$$\frac{1}{1-(\rho_1^2 + \rho_2^2 + \rho_3^2) + 2\rho_1\rho_2\rho_3} \begin{cases} (1-\rho_1^2)x_1 + (\rho_1\rho_2-\rho_3)x_2 + (\rho_1\rho_3-\rho_2)x_3 \\ (\rho_1\rho_2-\rho_3)x_1 + (1-\rho_2^2)x_2 + (\rho_2\rho_3-\rho_1)x_3 \\ (\rho_1\rho_3-\rho_2)x_1 + (\rho_2\rho_3-\rho_1)x_2 + (1-\rho_3^2)x_3 \end{cases}$$

et la distribution est de la forme :

$$(2\pi)^{-\frac{3}{2}} [1-(\rho_1^2 + \rho_2^2 + \rho_3^2) + 2\rho_1\rho_2\rho_3]^{-\frac{1}{2}}$$

$$\exp \left\{ -\frac{1}{[2 \ 1-(\rho_1^2 + \rho_2^2 + \rho_3^2) + 2\rho_1\rho_2\rho_3]} \right\} (1-\rho_1^2)x_1^2 + (1-\rho_2^2)x_2^2 + (1-\rho_3^2)x_3^2 + 2(\rho_2\rho_3-\rho_1)x_2x_3 + 2(\rho_3\rho_1-\rho_2)x_3x_1 + 2(\rho_1\rho_2-\rho_3)x_1x_3 \Big\}$$

Si l'on désire obtenir les distributions relatives à des variables qui ne sont ni centrées ni normées, il suffira de substituer à x_i l'expression :

$$(x_i - a_i) \sigma_i$$

6. - DISTRIBUTIONS CONDITIONNELLES ATTACHÉES A UNE DISTRIBUTION NORMALE A DEUX OU TROIS VARIABLES

Nous avons exploité p_{12} et p_{123} , par les formules (28) et (29) dans le cas d'une corrélation normale. Il nous est possible d'expliciter les distributions conditionnelles. Nous trouvons ainsi :

$$p_{1.2} = \frac{p_{12}}{p_2} = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{1 - r_{12}^2}} \exp \left\{ - \frac{(x_1 - r_{12} x_2)^2}{2(1 - r_{12}^2)} \right\} \quad (30)$$

$$p_{1.23} = \frac{p_{123}}{p_{23}} = \frac{1}{\sqrt{2\pi}} \times \frac{1}{\sigma_{1.23}} \exp \left\{ - \frac{(x_1 - b_{12.3} x_2 - b_{13.2} x_3)^2}{2\sigma_{1.23}^2} \right\} \quad (31)$$

$$\sigma_{1.23}^2 = \frac{1 - (r_{23}^2 + r_{31}^2 + r_{12}^2) + 2 r_{12} r_{23} r_{31}}{1 - r_{23}^2} \quad (32)$$

$$b_{1.23} = \frac{r_{12} - r_{23} r_{31}}{1 - r_{23}^2}$$

et enfin :

$$p_{12.3} = \frac{p_{123}}{p_3} = \frac{1}{\sqrt{2\pi} \sigma_{123}} \quad (33)$$

$$\exp \left\{ - \frac{(1 - r_{23}^2)(x_1 - r_{13} x_3)^2 + (1 - r_{13}^2)(x_2 - r_{23} x_3)^2 - 2(r_{12} - r_{13} r_{23})(x_1 - r_{12} x_3)(x_2 - r_{23} x_3)}{2\sigma_{123}^2} \right\}$$

7. - NOTION DE VARIANCE COLLECTIVE D'UN SYSTÈME DE VARIABLES

Nous sommes amenés par comparaison des formules (31) (32), (33) avec une distribution normale à une variable, à considérer les dénominateurs tels que σ_{123} comme des variances de système de variables. Soit d'une manière générale un système de n variables normales, centrées, indépendantes :

$$X_1, X_2 \dots X_n$$

et soit $\vec{X}$ le vecteur représentatif. Considérons n vecteurs indépendants ayant même distribution que $\vec{X}$:

$$\vec{X}^{(1)} \vec{X}^{(2)} \dots \vec{X}^{(n)}$$

Le parallélépipède construit sur ces vecteurs a pour volume la valeur du déterminant des coordonnées des $X^{(i)}$, c'est-à-dire :

$$V = \sum \pm X_{i_1}^{(1)} X_{i_2}^{(2)} \dots X_{i_n}^{(n)}$$

Le carré de ce volume est la somme de $(n!)^2$ termes de la forme :

$$X_{i_1}^{(1)} X_{i_2}^{(2)} \dots X_{i_n}^{(n)} X_{j_1}^{(1)} X_{j_2}^{(2)} \dots X_{j_n}^{(n)} \quad (35)$$

Prenons la valeur moyenne de V^2 . Seuls auront une valeur moyenne différente de 0, les termes (35) où $j = i_k$ ($k = 1, 2, \dots, n$) par conséquent :

$$\mathbb{M} \{ V^2 \} = \sum \mathbb{M} [X_{i_1}^{(1)} X_{i_2}^{(2)} \dots X_{i_n}^{(n)}]^2 - n! (\sigma_1 \sigma_2 \dots \sigma_n)^2$$

Si nous convenons donc de considérer comme variance collective des X_i la valeur moyenne du carré du volume du parallélépipède construit sur les vecteurs $X^{(i)}$, divisée par $n!$ nous obtiendrons pour cette variance l'expression :

$$\sigma_{12 \dots n} = \sigma_1 \sigma_2 \dots \sigma_n = | M_{12 \dots n} |^{1/2}$$

(En effet, on peut toujours par un changement d'axes se ramener au cas où les composantes de $\vec{x}$ sont indépendantes).

$|M|$ représente le déterminant de la matrice des moments. Comme la valeur de $|M|$ reste invariante par rotation d'axes, nous obtiendrons la même expression :

$$\sigma_{12 \dots n} = |M|^{1/2} \quad (36)$$

même dans le cas où les variables ne sont pas indépendantes (*).

8. - VARIANCES COLLECTIVES CONDITIONNELLES

Quand nous aurons affaire à des distributions conditionnelles, nous ferons intervenir des variances collectives conditionnelles :

Par exemple nous tirons de (33) :

$$\sigma_{12.3} = \sigma_{123}$$

dans le cas particulier où $\sigma_3 = 1$. Remarquons qu'une variance conditionnelle est toujours plus petite qu'une variance non conditionnelle. Par exemple :

$$\begin{aligned} \frac{\sigma_{12.3}}{\sigma_{12}} &= \frac{\sigma_{12.3}}{\sigma_{12}} = \sqrt{\frac{1 - (r_{12}^2 + r_{23}^2 + r_{31}^2) + 2 r_{12} r_{23} r_{31}}{1 - r}} \\ &= \sqrt{\frac{1 - r_{13}^2 + r_{23}^2 - 2 r_{12} r_{23} r_{31}}{1 - r_{12}^2}} < 1 \end{aligned} \quad (37)$$

(*) Il s'agit d'indépendance marginale.

Ceci peut s'expliquer en disant que toute information nouvelle diminue la variance. Nous reviendrons plus longuement sur ce point quand nous étudierons les corrélations multiples et partielles.

9. - CALCUL DES VARIANCES COLLECTIVES A PARTIR DE LA MATRICE DES MOMENTS

La variance collective de n variables apparaît en dénominateur dans la distribution des variables ; qui est de la forme (en supposant les variables centrées) :

$$(2\pi)^{-n/2} \frac{1}{\sigma_{12\dots n}} \exp \left\{ -\frac{1}{2} \vec{x} \mathcal{C}^{-1} \vec{x} \right\} dx_1 \dots dx_n \quad (38)$$

Soit α un certain groupe de variables d'indices $i_1, i_2, \dots, i_n$. La variance collective σ_α^2 de ces variables est donnée par :

$$\sigma_\alpha = |M_\alpha|^{\frac{1}{2}}$$

La matrice M_α se déduisant de la matrice M des moments en n'y conservant que les lignes et les colonnes relatives aux indices α .

Comme la distribution conditionnelle des variables α connaissant les variables β est :

$$P_{\alpha.\beta} = \frac{P_{\alpha\beta}}{P_\beta}$$

nous voyons que la variance collective conditionnelle est donnée par :

$$\sigma_{\alpha.\beta} = \frac{\sigma_{\alpha\beta}}{\sigma_\beta} = \frac{|M_{\alpha\beta}|^{\frac{1}{2}}}{|M_\beta|^{\frac{1}{2}}}$$

$M_{\alpha\beta}$ représentant la matrice obtenue en conservant dans $M_{12\dots n}$ les lignes et colonnes relatives aux indices α et β .

Nous voyons ainsi que toutes les variances collectives s'expriment sous forme de déterminants ou de quotients de déterminants extraits de la matrice des moments du deuxième ordre.

10. - INDÉPENDANCES DE VARIABLES ET DE GROUPES DE VARIABLES

Supposons que les variables X_i ne soient pas centrées ; posant :

$$X'_i = X_i - \eta_i \quad X'_i = X_i - a_i$$

$$\mu_{ij} = \eta_i (X'_i - a_i) (X'_j - a_j)$$

nous reconnaissons en :

$$r_{ij} = \frac{\mu_{ij}}{\sigma_i \sigma_j} = \frac{\mu_{ij}}{\sqrt{\mu_{ii} \mu_{jj}}}$$

les coefficients de corrélation des X_i prises deux à deux. Dans le cas de la corrélation normale, la distribution est entièrement définie quand on se donne les a_i, σ_i, r_{ij} .

Si nous considérons les deux variables X_1 et X_2 , la condition $r_{12} = 0$ nécessaire pour que les variables soient indépendantes, est ici suffisante (de par la forme d'une distribution normale). De même, la condition nécessaire et suffisante pour que deux groupes de variables α et β soient indépendants (*) est que l'on ait $r_{ij} = 0$ pour tout couple d'indices i et j tels que i appartienne à α et j à β . Dans ces conditions en effet la matrice $M_{\alpha\beta}$ prend la forme :

$$M_{\alpha\beta} = \begin{bmatrix} M_\alpha & 0 \\ 0 & M_\beta \end{bmatrix}$$

et la loi de distribution prend la forme :

$$C^{\text{te}} x \frac{1}{\sigma_\alpha} \frac{1}{\sigma_\beta} \exp \left\{ -\frac{1}{2} \vec{x}_\alpha \mathcal{C}_\alpha^{-1} \vec{x}_\alpha \right\} \exp \left\{ -\frac{1}{2} \vec{x}_\beta \mathcal{C}_\beta^{-1} \vec{x}_\beta \right\} \quad (39)$$

Si nous considérons trois groupes de variables α, β, γ , il se peut que les groupes α et β soient indépendants quand les variables γ sont connues, ou inversement. Précisons cela dans le cas de trois variables X_1, X_2, X_3 . La condition nécessaire et suffisante pour que X_1 et X_2 soient indépendantes est, quand X_3 n'est pas connue :

$$r_{12} = 0 \quad (40)$$

et, quand X_3 est connue (distribution $p_{12.3}$) :

$$r_{12} = r_{13} r_{23} \quad (41)$$

Ce sont là deux conditions différentes. Dans le premier cas la loi de distribution est de la forme :

$$p_1(x_1) p_2(x_2) p_{3.21}(x_3, x_1, x_2) \quad (42)$$

dans le deuxième cas, elle est de la forme :

$$p_3(x_3) p_{1.3}(x_1, x_3) p_{2.3}(x_2, x_3) \quad (43)$$

La seconde forme est la plus facile à déceler ; il suffit en effet pour qu'elle se présente, que p_{123} puisse s'écrire :

$$p_{123} = f(x_1, x_3) g(x_2, x_3) \quad (44)$$

On dit alors que X_1 et X_2 ne sont liés que par l'intermédiaire de X_3 .

(*) Il s'agit d'indépendance marginale.

11. - CONDITIONS D'INDÉPENDANCE EXPRIMÉES PAR LES VARIANCES

Quand X_1 et X_2 sont indépendantes, la connaissance de X_3 ne doit pas modifier la variance de X_1 ; donc la condition (40) est équivalente à :

$$\sigma_{1.2} = \sigma_1 \quad (45)$$

C'est ce que l'on constate sans peine.

Si X_1 et X_2 sont indépendantes quand X_3 est connue, la connaissance de X_2 ne doit pas modifier la variance de X_1 , X_3 étant connue ; c'est-à-dire que :

$$\sigma_{1.23} = \sigma_{1.3} \quad (46)$$

Or (46) s'écrit :

$$1 - \frac{r_{31}^2 + r_{12}^2 - 2 r_{12} r_{23} r_{31}}{1 - r_{23}^2} = 1 - r_{13}^2$$

ce qui, toutes réductions faites, est bien équivalent à (41).

De même, la condition nécessaire et suffisante pour que le groupe X_1, X_2 soit indépendant de X_3 est que :

$$r_{13} = r_{23} = 0 \quad (47)$$

Or, si tel est le cas, la connaissance de X_3 ne doit pas modifier la variance collective de X_1 et X_2 :

$$\sigma_{12.3} = \sigma_{12} \quad (48)$$

Et nous reportant à (37), (48) conduit à :

$$r_{13}^2 + r_{23}^2 - 2 r_{12} r_{23} r_{31} = 0$$

ce qui équivaut bien à (47) en tenant compte de la condition $|r_{12}| < 1$.

12. - GÉNÉRALISATIONS DE LA NOTION DE COEFFICIENT DE CORRÉLATION

Nous voyons ainsi par ces confrontations de formules, l'importance fondamentale de la variance collective, grâce à laquelle on peut exprimer toutes les relations d'indépendance entre groupes de variables. Nous pouvons dès maintenant poser les principes de la théorie des corrélations multiples et partielles. Nous avons vu que le coefficient de corrélation ordinaire r_{12} est en relation avec la diminution de variance :

$$\frac{\sigma_{1.2}^2}{\sigma_1^2} = 1 - r_{12}^2 = \frac{\sigma_{12}^2}{\sigma_1^2 \sigma_2^2} \quad (49)$$

D'une manière générale, quand la connaissance des valeurs des variables d'un groupe β diminuera la variance collective d'un groupe α , nous

associerons à ces deux groupes un coefficient de corrélation généralisant la formule (49), et poserons :

$$\frac{\sigma^2_{\alpha,\beta}}{\sigma^2_{\alpha}} = 1 - R^2_{\alpha\beta} = \frac{|M_{\alpha\beta}|}{|M_{\alpha}| |M_{\beta}|} = \frac{|S_{\alpha\beta}|}{|S_{\alpha}| |S_{\beta}|} \quad (50)$$

$R_{\alpha\beta}$ sera un coefficient de corrélation multiple. Si la formule (50) est relative au cas où les variables d'un troisième groupe sont supposées connues, nous poserons :

$$\frac{\sigma^2_{\alpha,\beta\gamma}}{\sigma^2_{\alpha,\gamma}} = 1 - R^2_{\alpha,\beta(\gamma)} = \frac{|S_{\alpha\beta\gamma}| |S_{\gamma}|}{|S_{\alpha}| |S_{\beta}|} \quad (51)$$

et $R_{\alpha,\beta(\gamma)}$ sera coefficient de corrélation partielle. Les formules (50) et (51) grâce à la notion de variance collective montrent l'unité de conception qui existe entre les définitions des différents coefficients de corrélation.

De (50) et (51) on peut déduire :

$$1 - R^2_{\alpha,\beta(\gamma)} = \frac{1 - R^2_{\beta\gamma,\alpha}}{1 - R^2_{\gamma,\alpha}} = \frac{1 - R^2_{\alpha\gamma,\beta}}{1 - R^2_{\gamma,\beta}} = 1 - R^2_{\beta,\alpha(\gamma)} \quad (52)$$

TABLE DES MATIÈRES

ELEMENTS DE L'ALGÈBRE DE BOOLE	107
PRINCIPES D'ANALYSE MATRICIELLE	141
Chapitre premier. - Introduction algébrique	141
Chapitre II. - Notations vectorielles et matricielles	146
Chapitre III. - Réduction des matrices - rang	150
Chapitre IV. - Algèbre des matrices carrées	163
Chapitre V. - Matrices carrées particulières	186
Chapitre VI. - Des probabilités conditionnelles	190
Chapitre VII. - Notions sur la théorie de l'information	201
Chapitre VIII. - Matrices et corrélation	204

