MATHÉMATIQUES ET SCIENCES HUMAINES

Problèmes d'enseignement. Sur quelques points relatifs à l'enseignement élémentaire des mathématiques

Mathématiques et sciences humaines, tome 27 (1969), p. 29-57

<http://www.numdam.org/item?id=MSH_1969__27__29_0>

M.S.H. (7e année, no 27, 1969, p. 29-46)

PROBLÈMES D'ENSEIGNEMENT

SUR QUELQUES POINTS RELATIFS A L'ENSEIGNEMENT ÉLÉMENTAIRE DES MATHÉMATIQUES

par

M. BARBUT

Le vieux débat sur les mathématiques dites « modernes » a pris ces derniers temps, une acuité et une ampleur inégalée ; des déclarations fracassantes, et peut-être imprudentes du Ministère de l'Education Nationale, la floraison d'articles et d'ouvrages de vulgarisation destinés notamment aux parents d'élèves, et surtout l'anarchie qui semble caractériser la modernisation de l'enseignement des mathématiques, ont contribué à créer un état de l'opinion empreint d'un certain malaise.

Cet article ne veut pas entrer dans le fond du débat ; il vise simplement à l'éclairer, s'il se peut, en tentant de lever certaines confusions qui semblent à l'auteur être largement répandues.

De façon précise, il apparaît que pour le grand public, il y a opposition entre mathématiques « modernes » et mathématiques « de papa » ; et que pour ce même public, mathématiques « modernes » est quasiment synonyme de « Théorie des Ensembles », ou plus exactement, de ce que la plupart des articles et ouvrages auxquels il a été fait allusion plus haut, appellent ainsi.

Le but de cet article est donc d'abord de montrer ce qu'a, à certains égards, d'artificiel, l'opposition entre les deux mathématiques ; on veut attirer l'attention du public, et singulièrement des enseignants, sur les liens profonds et naturels qui unissent des aspects très élémentaires de la fameuse « théorie des ensembles » (il s'agit en fait de la combinatoire et de l'algèbre ensembliste élémentaires) à la géométrie, l'algèbre linéaire et l'arithmétique dans des aspects non moins élémentaires.

Il est ensuite nécessaire, à ce propos, de faire saisir que dans une modernisation bien comprise de l'enseignement des mathématiques, l'initiation au langage ensembliste ne constitue qu'un chapitre somme toute mineur.

La thèse de l'auteur est triple : ne pas cristalliser l'attention des jeunes esprits sur une représentation particulière (fut-elle « moderne ») d'une abstraction, ne pas créer d'opposition factice entre divers chapitres des mathématiques, éviter que la mise à jour de l'enseignement des mathématiques ne se fasse en passant à côté des acquisitions les plus fondamentales de celles-ci au cours des cent dernières années.

Il faut mettre en garde à l'égard d'un dogmatisme possible, des mathématiques « modernes », dogmatisme qui serait aussi néfaste que celui qu'on a reproché aux mathématiques « classiques » ; s'il y a divorce entre ces deux mathématiques, ce divorce doit moins porter sur ce qui est enseigné, que sur la manière de l'enseigner ; trop d'enseignants, hélas, l'ont mal compris (il faut dire à leur décharge que les responsables de la pédagogie ne les y ont guère aidés). On ne dit pas ici comment introduire et éche-

lonner les diverses notions abordées ; c'est l'affaire des pédagogues, et l'auteur n'est pas compétent. Certains lecteurs trouveront sans doute trop de longueurs sur des questions qu'ils connaissent bien (en particulier, tout le premier paragraphe peut être omis par ceux qui sont familiers de l'enseignement dans les Facultés des Lettres).

I. — DU CONCRET A L'ABSTRAIT : MODÈLE ÉLÉMENTAIRE DU STATISTICIEN.

Partons de la situation concrète suivante : le statisticien qui observe une population (de ménages, ou de machines, ou de communes, etc.) et s'intéresse à ceux des individus de cette population qui présentent tel caractère donné.

La population observée peut être assimilée à un ensemble *fini* E (fini, c'est-à-dire dont on puisse écrire la liste exhaustive de tous ses éléments : les individus de la population). Et le résultat d'une observation, c'est une *partie* de E, constituée des éléments de E qui présentent le caractère observé. Cette partie peut éventuellement être *vide* (on la note $\emptyset$) : aucun individu ne présente alors le caractère observé ; à l'opposé, elle peut être *pleine* (c'est E tout entier).

Mais l'on veut être en mesure ensuite d'*interpréter* les résultats d'une ou plusieurs observations sur une même population, de savoir quels sont ceux qui sont rares et ceux qui sont fréquents, ceux auxquels on pouvait s'attendre, et ceux qui doivent surprendre, de sorte qu'ils amèneront à se poser d'autres questions sur la population étudiée ; il est clair qu'il faut alors pouvoir d'une part, situer chaque résultat dans l'ensemble de tous les résultats qui étaient a priori possibles, d'autre part situer les uns par rapport aux autres plusieurs résultats d'observations : i.e., situer une partie donnée de E dans l'ensemble de *toutes les parties de E* (syn. : *sous-ensembles de E*), et plusieurs parties de E les unes par rapport aux autres.

Le programme est donc le suivant : étant donné un ensemble fini E,

1) Constituer l'ensemble de toutes ses parties, c'est-à-dire, *énumérer* et *dénombrer* toutes ses parties ;

2) *Organiser*, d'un point de vue qui soit signifiant pour le statisticien, cet ensemble des parties de E.

Bien entendu, il s'agit d'un programme minimum, loin d'épuiser toutes les techniques utiles de la statistique ; mais c'est sur la réalisation de ce programme que sont fondées toutes les techniques plus élaborées (en particulier, celles qui font appel au Calcul des Probabilités).

Ce programme très primitif, mais nécessaire à l'intelligence du métier de statisticien, nous aurions pu l'obtenir à partir de bien d'autres circonstances « concrètes » : réussite à des tests, en psychologie ; classement d'objets définis par des caractères, qu'ils peuvent présenter ou non, en archéologie, etc. De tels exemples peuvent être multipliés ; que le lecteur en trouve dans les domaines qui lui sont familiers.

Enumérer, dénombrer.

Rappelons le raisonnement classique : soit par exemple, un ensemble E ayant trois éléments, a, b et c ; on note : $E = \{ a, b, c \}$.

Ses parties se divisent en deux classes :

— celles qui contiennent c, celles dont c n'est pas élément ; ces dernières sont toutes les parties de l'ensemble $\{ a, b \}$, faciles à énumérer : $\emptyset$ (vide) ; $\{ a \}$ (a tout seul) ; $\{ b \}$ (b tout seul) ; $\{ a, b \}$ (plein) ;

— celles qui contiennent c : chacune se déduit de celles de cette liste en ajoutant c : $\{ c \}$; $\{ a, c \}$; $\{ b, c \}$; $\{ a, b, c \}$.

Donc 8 (= 2 × 4) parties en tout pour un ensemble ayant trois éléments ; et la loi fondamentale : *ajouter un* élément à un ensemble *double* le nombre de ses parties. Loi qui se résume dans la « formule » : si E a n éléments, il a 2^n parties (exemple : $8 = 2^3$).

En particulier, l'ensemble vide a une et une seule partie, lui-même : $2^0 = 1$.

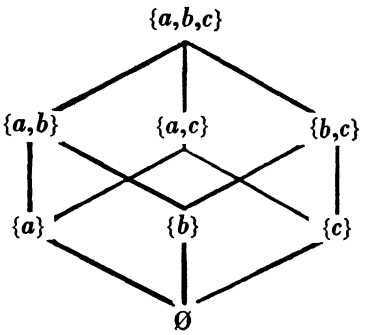
Remarquons que la *récurrence* porte ici non seulement sur un dénombrement, mais sur des listes ; et notons cette première (mais non dernière) rencontre avec le mécanisme *exponentiel :*

ajouter un	⟶	doubler
+		×
(addition)		(multiplication)

Organiser, structurer.

Qu'une organisation de l'ensemble des parties soit nécessaire découle du résultat précédent : si E a seulement 30 éléments, il admet 2^{30} (c'est-à-dire plus d'un milliard) parties. Un principe simple d'organisation : si A et B sont deux parties d'un même ensemble E, A est *incluse* dans B (on note : $A \subset B$) si tous les éléments de A sont éléments de B.

Cette relation d'inclusion, toujours significative pour le praticien, définit un *ordre* (partiel) sur les parties ; voici par exemple, l'organigramme (ordonné de bas en haut) de cet ordre pour les parties de $E = \{ a, b, c \}$.

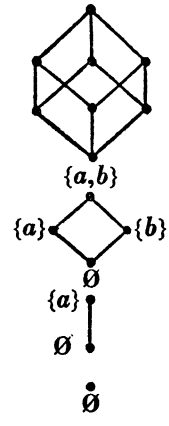


Ce diagramme peut également s'interpréter comme le diagramme des « voisinages » entre parties : deux parties sont *adjacentes* lorsqu'elles diffèrent le moins possible, c'est-à-dire d'un seul élément ; ainsi, dans un ensemble ayant n éléments, chaque partie est adjacente à n parties (obtenues soit en lui ôtant un élément, soit en lui ajoutant un élément). Plus généralement, la *distance* entre deux parties est ici le nombre d'éléments par lesquels elles diffèrent, et se lit sur le diagramme comme longueur commune (en nombre de traits) des chemins minimaux reliant entre eux les deux sommets du diagramme représentant ces deux parties.

Représenter : les parallélotopes.

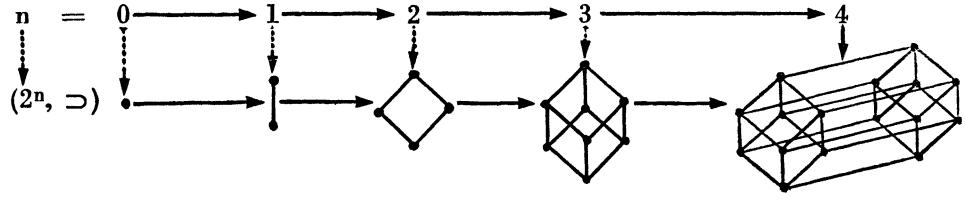
Le diagramme de l'ensemble des parties d'un ensemble ayant trois éléments comporte 8 sommets et 12 arêtes, et chaque sommet est incident à 3 arêtes exactement : ce diagramme est celui du *cube*, si on le « voit » dans l'espace de dimension 3. De même, si E a deux éléments, le diagramme comporte 4 ($= 2^2$) sommets (parties de E), 4 arêtes, et chaque sommet est incident à 2 arêtes : c'est un *carré*, de l'espace à deux dimensions. Si E a 1 seul élément, le diagramme est un segment ; si E a 0 éléments ($E = \emptyset$) le diagramme se réduit à un point.

La série : point, segment, carré, cube, dans des espaces de dimension 0, 1, 2, 3, répond à la série 1, 2, 4, 8... des puissances successives de 2 rencontrées lors du dénombrement des parties. Ce n'est pas un hasard, et la considération de la *récurrence* fournit encore la clef. Supposons construit le diagramme des parties d'un ensemble E, et ajoutons un nouvel élément x à E ; nous obtenons un nouvel ensemble



qui contient E (et n'en diffère que par x) dont les parties se divisent en deux classes : celles qui ne contiennent pas x, et dont le diagramme est celui de P (E) [1] ; celles qui contiennent x, et dont chacune se déduit de celles qui ne contiennent pas x par adjonction de cet élément ; elles ont donc entre elles les mêmes relations d'inclusion que (P (E), ⊃), et chacun est « voisine » de la partie de E dont elle est issue par adjonction de x.

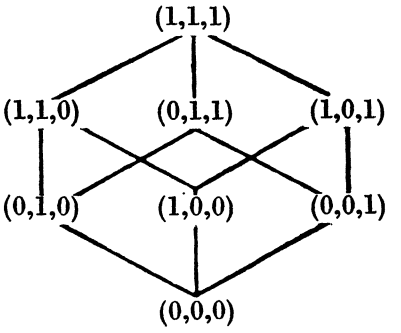
Par conséquent, le diagramme se *dédouble par translation*, le *vecteur* de translation signifiant « ajouter x » ; et la série se construit algorithmiquement à partir de son premier terme, le diagramme de P (Ø), qui se réduit à un point, espace de dimension 0 ; dédoublement du point : le segment, dimension 1 ; dédoublement du segment : le carré, dimension 2 ; dédoublement du carré : le cube, dimension 3 ; dédoublement du cube : l'hypercube, dimension 4 (dont la figure ci-dessous est une projection plane), etc. :



Autrement dit, ce diagramme se construit, si E a n éléments, en choisissant n *vecteurs de base* de l'espace de dimension n, chaque vecteur représentant un et un seul élément de E ; et en construisant la parallélotope obtenue en additionnant (au sens vectoriel) entre eux des vecteurs de base distincts : à la partie vide correspond le *vecteur nul O*, aux parties { a }, { b }, réduites à un seul élément les n vecteurs de base **a**, **b**... ; aux parties { a, b }, { a, b, c }, etc., les vecteurs *somme* (selon la vieille règle du parallélogramme) **a** + **b**, **a** + **b** + **c**, etc. Dans les diagrammes tout est significatif : les sommets signifient les parties, les arêtes signifient le voisinage (par inclusion), et chaque direction (pourvu que l'épure soit correctement dessinée) signifie l'adjonction d'un élément. Chaque partie de l'ensemble { a, b, c } ayant trois éléments, par exemple, peut être notée par un triplet (x_1, x_2, x_3) de *coordonnées* (toutes égales

1. On désigne par P(E) l'ensemble des parties d'un ensemble E.

à 0 ou à 1), celles du *vecteur* correspondant dans l'espace de dimension 3. Le *vecteur nul* (partie vide) a pour coordonnées (0, 0, 0) ; les vecteurs de base sont (1, 0, 0), (0, 1, 0), (0, 0, 1), etc. La valeur 0 de la coordonnée x_i s'interprète comme l'absence de l'élément i dans la partie considérée ; la valeur 1, comme la présence de cet élément.



Dénombrements.

Ainsi, étant partis d'un certain « concret » fort modeste : un ensemble fini, ses parties, ce avec quoi travaille, dans la portion la plus humble de son métier, le statisticien, nous rencontrons, précisément pour bien comprendre ce que fait le statisticien, des objets purement géométriques, les parallélotopes, et nous pressentons qu'est présente dans cette activité l'algèbre linéaire, celle des espaces vectoriels et des modules, qui domine l'essentiel des Mathématiques pures ou appliquées. Ce point de vue sera systématiquement exploité, et conduira à d'autres rencontres, qui ne sont pas fortuites, avec quelques autres grands thèmes des Mathématiques.

Mais il nous faut d'abord souligner que la seule considération des parallélotopes représentant l'organisation de $(\mathrm{P}(\mathrm{E}), \supset)$ permet de résoudre aisément quelques-uns des problèmes qui se posent d'emblée au statisticien, et singulièrement ceux qui ont trait aux dénombrements ; les dénombrements les plus classiques de l'analyse combinatoire, la trilogie un peu désuète : arrangements, permutations, combinaisons, se retrouve en considérant les chemins et les sommets du diagramme. Le nombre $n\,(n-1)\,(n-2)\,\ldots\,(n-p+1)$ « d'arrangements » p à p de n éléments est le nombre de chemin minimaux reliant, sur le diagramme, deux sommets à distance p ; en particulier, il y a $n\,!$ (factorielle de n) chemins entre deux sommets à distance n. Chacune des $n\,!$ permutations des n éléments de E se lit d'ailleurs en suivant les chemins minimaux reliant le sommet $\emptyset$ au sommet E, et en notant les directions prises successivement : la permutation $a\ c\ b$ de $\{a, b, c\}$ correspond au chemin, qui partant de $\emptyset$, passe successivement par les sommets représentant $\{a\}$, puis $\{a, c\}$, puis $\{a, b, c\}$. Quant au nombre $\dfrac{n\,!}{p\,!\,(n-p)\,!}$ des « combinaisons » p à p de n éléments, c'est le nombre des sommets à distance p d'un sommet donné, comme le montre un raisonnement immédiat sur les chemins (raisonnement du type « règle des bergers » : combien y a-t-il de moutons ? Comptez les pattes et divisez par 4). Si l'on compte les distances à partir du sommet figurant la partie vide, on voit que ce nombre des « combinaisons p à p » est également celui des parties ayant p éléments dans un ensemble qui en a n : dénombrement très significatif pour la statistique qui souvent s'intéresse plus au nombre d'individus (d'une population) qui possèdent tel caractère, qu'à l'identité de chacun de ces individus. La fréquence relative d'une partie ayant p éléments dans un ensemble de n éléments est, si l'on désigne par $\binom{n}{p}$ le nombre des « combinaisons p à p », $\dfrac{1}{2^n}\binom{n}{p}$. La notion de « rareté » ou de « fréquence » du résultat qu'on peut obtenir dans une observation commence ainsi à prendre un sens précis.

Enfin, les équations de récurrence liant les nombres considérés :

$$n! = n \times (n-1)! \qquad \text{et} \qquad \binom{n}{p} = \binom{n-1}{p} + \binom{n-1}{p-1},$$

résultent de façon non moins évidente de la récurrence (dédoublement — translation) liant les diagrammes correspondants.

D'autres jeux combinatoires sont possibles, tel celui qui consiste à reconnaître derrière chaque décomposition de 2^n en un produit : $2^n = 2^p \times 2^q$ $(p + q = n)$ une décomposition du parallélotope de dimension n en 2^p parallélotopes de dimension q (chacun a 2^q sommets) déduits les uns des autres par p translations. Décomposition que nous avons vue en détail dans le cas particulier : $2^n = 2 \times 2^{n-1}$, mais qui est générale, et qui d'ailleurs, débouche sur (ou peut être montrée à partir de) les *morphismes* (chercher le noyau d'un morphisme, prendre son quotient) des algèbres de Boole (ou des anneaux booléens) finies associées à l'ensemble des parties d'un ensemble fini ; algèbre dont il faut maintenant dire quelques mots.

Calculer : algèbres.

Il est en effet possible, et même utile, de munir l'ensemble P (E) des parties d'un ensemble fini E d'une *structure algébrique* (en fait, de plusieurs, on le verra ci-dessous), c'est-à-dire d'y définir des opérations, avec leurs règles, et des équations. On munit P (E) d'une structure d'*algèbre de Boole* et prenant pour opérations les deux opérations binaires d'intersection (symbolisée ∩) et union (symbolisée ∪) et l'opération unaire de complémentation. Comme il s'agit là de la tarte à la crème de ce que bien des auteurs de manuels appellent « mathématiques modernes », nous n'insisterons pas sur ces opérations, dont les définitions et les propriétés se trouvent partout. Nous nous contenterons, à leur propos, de quelques remarques :

1. Munir P (E) d'une structure algébrique, par exemple, celle d'algèbre de Boole, est *utile*, même pour l'humble pratique du statisticien. En effet, énumérer, dénombrer, organiser les parties de E nous ont permis de comprendre déjà bien des choses à leur propos, il n'en reste pas moins que ces parties sont en foule, et qu'il est nécessaire par conséquent de pousser plus loin l'automatisation de certaines démarches usuelles à leur propos (l'intersection, par exemple, correspond à la recherche très usuelle — tableaux croisés — des individus d'une population présentant deux caractères donnés), de *remplacer par du calcul* tout ce qui jusqu'ici, a fait appel à notre intuition (de l'espace en particulier). Calcul, c'est-à-dire jeu d'écritures, avec des règles bien définies, jeu d'équations : il n'est plus nécessaire de comprendre, si les signes et les règles d'assemblage de ceux-ci sont donnés ; à la limite, une machine remplacera l'homme. De la présentation combinatoire qui a été donnée ci-dessus de P (E) à l'algèbre de Boole des parties de E, la démarche est analogue à celle qui va de la « géométrie élémentaire » à la « géométrie analytique », qui remplace les raisonnements directs sur les points, droites, cercles, etc., par du calcul sur les équations (de la droite, du cercle, etc.).

Encore faut-il, pour qu'une telle réduction au calcul soit possible, s'assurer que les règles dégagées à partir de la signification (en termes de parties d'ensemble) permettent bien de caractériser celle-ci ; autrement dit, que tout « texte » écrit au moyen de signes arbitraires, mais en respectant les règles (ou axiomes de la structure) peut bien s'interpréter ensuite avec le sens voulu : toutes les parties d'un ensemble et les opérations considérées entre celles-ci. En ce qui concerne les algèbres de Boole finies, c'est bien le cas (théorème de représentation de Stone).

2. Introduire les opérations d'union, d'intersection, de complémentation, entre parties d'un ensemble n'est à aucun titre, faire de la « théorie des ensembles ». Tout au plus, est-ce emprunter au vocabulaire de celle-ci, dont l'objet principal est l'étude des problèmes soulevés par la considération

d'ensembles *infinis*. Dans tout ce qui a été développé jusqu'ici, seules les questions de récurrence ont quelque peu rapport avec la dite théorie. En introduisant ces opérations, on fait de l'algèbre, un point c'est tout.

Pas plus que de la Théorie des Ensembles, d'ailleurs, on ne fait de la logique, comme certains naïfs le croient ; s'il est vrai que G. Boole créa, incomplètement (il ne sut jamais faire clairement la distinction entre ce que nous appelons maintenant *algèbre* de Boole et *anneau* booléen), les algèbres qui portent son nom à propos de recherches de logique, dire que l'on pratique cette science lorsqu'on fait de l'algèbre de Boole, c'est un peu comme dire que l'on pratique de la théorie de la musique, par exemple, lorsqu'on calcule sur les fractions.

3. La structure d'algèbre de Boole n'est pas la seule que l'on puisse introduire sur l'ensemble P (E) des parties d'un ensemble fini ; dans bien des cas, et en particulier parce que cela met à notre disposition toute la machinerie très élaborée de l'algèbre linéaire, il est plus avantageux de munir P (E) d'une structure d'anneau (l'anneau booléen) au moyen d'une « addition » (la différence symétrique) et d'une « multiplication » (l'intersection), ou d'espace vectoriel (de dimension n sur le corps de caractéristique 2), ou enfin d'*algèbre* (au sens technique du terme). Il y a là un aspect qui nous semble trop négligé dans les essais de « modernisation » de l'enseignement des Mathématiques.

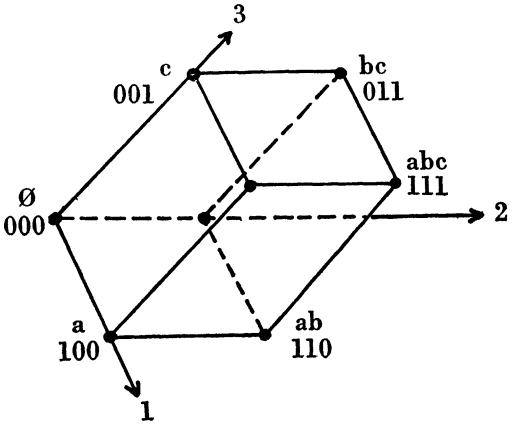
II. — PLUSIEURS REPRÉSENTATIONS D'UN MÊME OBJET ABSTRAIT.

Simplexes géométriques.

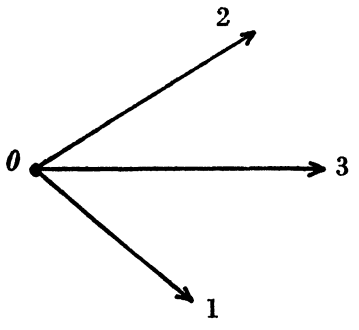
Les remarques précédentes ont pris un tour volontairement elliptique, dont le lecteur voudra bien nous excuser ; elles ne sont là que pour donner quelques indications relatives à la pédagogie de notre sujet, et chacune pourrait faire l'objet d'un long développement. Mais il nous semble plus intéressant de nous attacher à deux représentations de l'objet abstrait qui a été dégagé de la situation « concrète » initiale, représentations qui permettront d'en saisir mieux les liens avec les thèmes fondamentaux des Mathématiques de toujours.

Objet abstrait, disions-nous ; dès l'instant où nous avons représenté l'organisation de l'ensemble des parties d'un ensemble E fini, ordonnées par inclusion, par un parallélotope de dimension n, nous sommes en effet passés à une certaine abstraction : des mots distincts, avec des sens distincts, décrivaient des relations ayant la même combinatoire, les mêmes propriétés. Il y avait *représentation* de l'un : $(P(E), \supset)$, par l'autre : le parallélotope, et ses relations d'incidence.

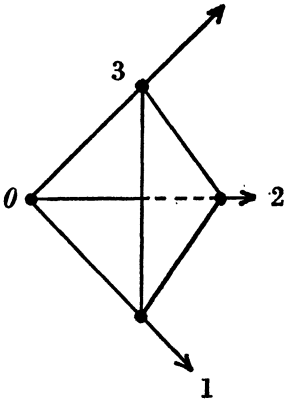
La prochaine représentation que nous allons étudier est géométrique. Nous savons que le diagramme de P ({ a, b, c }) par exemple, est représentable par le parallélotope de l'espace de dimension 3 (figure ci-dessous), rapporté à trois axes de coordonnées 1, 2 et 3.



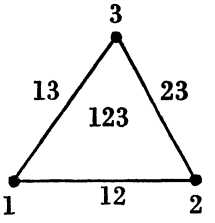
Considérons l'orthant positif de cet espace (fig. (a)), et coupons le par un plan (fig. (b)).



(a)



(b)

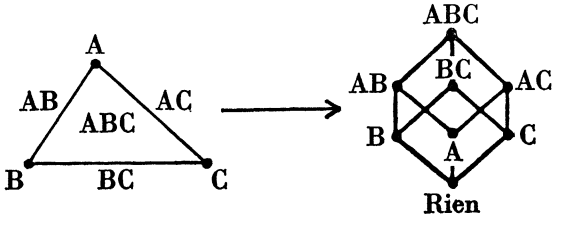


(c)

La section obtenue est un triangle, projection de l'orthant positif sur le plan sécant dans une perspective de sommet *O*, origine de l'espace (fig. (c)). Chaque axe a pour trace un sommet du triangle, désigné par le même symbole que l'axe dont il est projection, les côtés du triangle sont les projections des plans 12, 23 et 31 respectivement, et l'intérieur du triangle est la projection de l'intérieur de l'angle trièdre défini par les trois axes : on le nomme 123. Dans ces désignations, nous retrouvons, au changement de notations près, les 7 (= 8 — 1) parties non vides de $\{ a, b, c \}$; la partie vide, dont le point représentatif dans l'espace de dimension 3 a été pris comme centre de perspective, n'apparaît bien entendu pas.

Nous passons ainsi d'une représentation *vectorielle* (espace pointé, avec une origine *O*) à une représentation *affine :* il n'y a plus d'origine, tous les points jouent le même rôle.

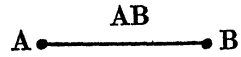
Exercice inverse : les triangles de la géométrie plane euclidienne peuvent être rectangles, isocèles, équilatéraux, voire quelconques ; mais si nous oublions les propriétés métriques, que reste-t-il ? Qu'y a-t-il de commun à tous les triangles, quels sont les invariants du triangle ? Un triangle, trois sommets,



trois côtés, et des relations d'incidence des uns par rapport aux autres : chaque côté est incident à deux sommets, chaque sommet est incident à deux côtés. Représentons par un diagramme ces relations d'incidence du triangle ABC ; le triangle ABC, incident aux trois côtés AB, BC, CA, etc. ; si l'on complète le diagramme par « rien », on retrouve le diagramme de P ($\{ a, b, c \}$) ; on peut d'ailleurs orienter

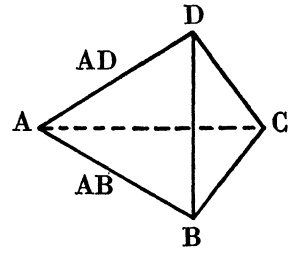
le diagramme (du plein au vide, du triangle ABC à rien) en associant à chaque objet son bord : le triangle ABC a pour bord les trois côtés AB, BC et CA, le côté AB a pour bord les sommets A et B, chaque sommet est sans bord.

De la même façon, P ({ a, b }) a pour représentation le segment AB ; (P ({ a }), $\supset$), le point A.

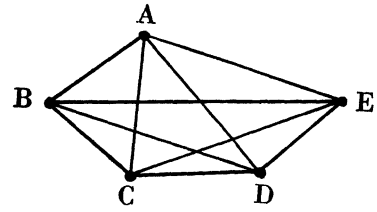


Pour (P ({ a, b, c, d }), $\supset$), sa représentation géométrique est le tétraèdre, et le mécanisme de récurrence pour passer d'une représentation à la suivante, est la transposition de celui qui nous a fait passer d'un diagramme d'inclusion au suivant.

Pour passer de trois à quatre, on ajoute un élément d : on ajoute un sommet D (non coplanaire aux trois précédents). Les parties de l'ensemble à quatre éléments se divisent en deux classes : celles qui ne contiennent pas d, celles qui le contiennent déduites des précédentes en ajoutant d : chaque sommet du triangle donnera une arête supplémentaire AD, BD, CD ; chaque côté, une face ABD, BCD, ACD ; et la face ABC, le tétraèdre ABCD.



Pour cinq, même mécanisme : on ajoute un sommet E, non dans le même espace de dimension trois que le tétraèdre précédent ; on obtient le « tétraèdre généralisé » de l'espace quatre, avec cinq sommets, 10 (= 5 + 5) arêtes, 10 (= 4 + 6) faces triangulaires, 5 (= 1 + 4) faces tétraédriques (de dimension 3).

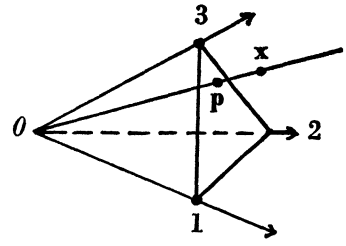


La suite des objets ainsi constitués, rien, point, segment, triangle, tétraèdre..., de dimensions respectives —1, 0, 1, 2, 3... sont les *simplexes* géométriques de la topologie.

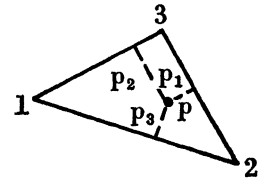
D'où l'habitude qui a été prise par certains d'appeler *simplexes* les objets *abstraits* dont nous avons vu trois représentations : l'ensemble des parties d'un ensemble fini ayant n éléments muni de l'ordre d'inclusion ; le parallélotope de dimension n, muni de l'ordre d'incidence ; le simplexe géométrique de dimension (n — 1), muni de la relation d'incidence.

Mesures et distributions.

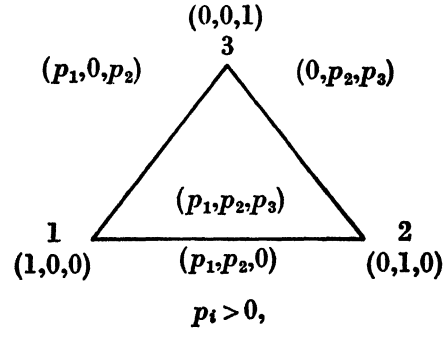
La représentation par les simplexes géométriques, permet notamment de représenter non seulement chaque ensemble fini, et l'organisation par inclusion de ses parties, mais aussi l'ensemble des mesures (de probabilité, par exemple) sur cet ensemble. Reprenons le cas de l'ensemble { *a*, *b*, *c* } à trois éléments, et l'espace de dimension trois rapporté à trois axes de coordonnées 1, 2 et 3. Chaque point **x** de cet espace est défini par le triplet (x_1, x_2, x_3) de ses coordonnées : celles-ci sont des nombres, par exemple réels.



Coupons l'espace par le plan d'équation $x_1 + x_2 + x_3 = 1$; lorsque l'on projette, dans la perspective de centre O, l'espace sur ce plan, chaque point **x** a pour perspective un point **p** de coordonnées (p_1, p_2, p_3) et telles que $p_1 + p_2 + p_3 = 1$. Si les trois coordonnées de **x** sont positives ou nulles (i.e., si **x** appartient à l'orthant positif), p_1, p_2 et p_3 sont également des nombres positifs. Le point **p** peut donc être interprété comme une distribution sur l'ensemble { 1, 2, 3 } ou, par changement de notations, { *a*, *b*, *c* [1]. Les nombres p_1, p_2 et p_3 sont d'ailleurs respectivement proportionnels aux distances euclidiennes de **p** aux côtés 23, 13 et 12 du triangle, distances dont la somme est constante; c'est d'ailleurs le principe sur lequel reposent les papiers millimétrés triangulaires, dont la statistique descriptive fait grand usage pour figurer les distributions en trois classes.



En particulier, les trois distributions (mesures) concentrées sur un seul élément ont pour image le sommet correspondant ; celles dont le *support* (i.e., la partie de E dont les éléments *i* ont une mesure p_i non nulle) a deux éléments, ont pour image un point de côté correspondant du triangle ; celle dont le support est l'ensemble E tout entier ont pour image un point intérieur au triangle.

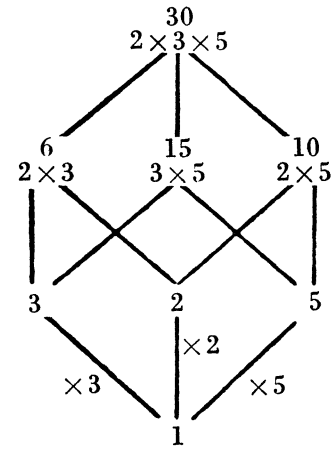


Ainsi, tant qu'on ne considérait que l'architecture combinatoire du triangle : sommets, côtés, relations d'incidence, on avait une représentation de l'ensemble des parties de $E = \{ a, b, c \}$; dès l'instant où ce triangle est considéré comme plongé dans le plan euclidien, on a en outre une représentation de l'ensemble des distributions sur E : à chaque point correspond une distribution, à chaque distribution un point. Et la situation du point sur le triangle définit une partie, le support de la distribution.

Nous avons parlé de distributions, car nous avons en tête le « concret » dont nous sommes partis au début de cet article : la statistique descriptive. Dans d'autres contextes, on interprétera les points comme des *distributions de probabilité* (en statistique inductive, en particulier), ou l'on parlera de « coordonnées barycentriques » (en mécanique, par exemple).

Arithmétique.

Une autre représentation classique des simplexes se trouve en arithmétique : soit par exemple, le nombre entier $30 = 2 \times 3 \times 5$; représentons par un diagramme l'ensemble des diviseurs entiers naturels de 30, les traits indiquant la divisibilité. Nous obtenons, de façon évidente, le diagramme du



simplexe — 3. Ici, les directions des arêtes ne signifient plus, comme dans le cas ensembliste, « ajouter l'élément x », mais « multiplier par le nombre premier x ». La généralisation est immédiate : une représentation du simplexe — n s'obtient en choisissant n nombres premiers distincts (et différents de 1) $p_1, p_2, \ldots, p_n$, et en considérant tous les diviseurs de $p_1 \times p_2 \times \ldots \times p_n$. La « traduction » de la représentation ensembliste à la représentation arithmétique se fait selon le tableau :

∅	1
⊂ (inclusion) ⊃	(divisibilité) (multiple de)
∪ union	p.p.c.m.
∩ intersection	p.g.c.d.
éléments	nombres premiers

Ainsi, nous disposons maintenant de quatre « langages » : le langage ensembliste, le langage vectoriel (parallélotopes), le langage géométrique (simplexes géométriques), le langage arithmétique, quatre « sémantiques », pour réaliser une même abstraction (une même « syntaxe »), le simplexe — n. Les signes et les sens changent, mais la combinatoire des signes, lorsque les « traductions » sont correctement faites, est la même. D'autres représentations du même objet sont d'ailleurs possibles, en logique formelle notamment : faute d'être compétent, nous n'en parlerons pas.

La mise en évidence de ces *isomorphismes*, montre ce qu'il y a de commun à plusieurs types d'activités apparemment fort éloignées les unes des autres.

Mais elle souligne surtout pour nous ce qu'a d'assez puérile l'opposition entre « mathématiques qualitatives » et « mathématiques quantitatives », mathématiques « modernes » et mathématiques « de papa ». Le point de vue adopté ici montre au contraire que ce que l'on a à dire sur les ensembles finis (le « qualitatif ») est en fait constitué de parties importantes de ce que l'on a toujours dit, dans les programmes « sclérosés », en arithmétique et en géométrie, ou de ce que l'on devrait dire, dans des programmes « rénovés », en algèbre linéaire.

Seul est nouveau « l'éclairage » porté sur les objets étudiés ; ce qui est fondamental, c'est la construction des isomorphismes ; et à cet égard, il convient de fortement mettre en garde les pédagogues. A trop insister, et fixer l'intelligence des enfants, sur une seule représentation d'un même être mathématique (la représentation ensembliste, dans l'enseignement « moderne »), on ne fait pas une bonne pédagogie ; dans l'esprit de l'élève moyen, il risque de ne subsister que la seule représentation concrète que l'on a privilégiée ; au mieux ne pourra-t-il s'élever à l'abstraction qu'au prix d'efforts sans commune mesure avec leur objet.

Bien entendu, il n'est pas dans notre propos ici de plaider pour que l'on enseigne effectivement aux enfants (de quel âge, dans quel contexte ?) les quelques points abordés dans cet article ; ni a fortiori de proposer une méthode d'enseignement de ces points ; notre métier n'est pas la pédagogie des jeunes. Notre but n'était que de faire saisir l'unité profonde de quelques chapitres des mathématiques, dont certains sont réputés « anciens » et d'autres « modernes ».

Nous espérons avoir été assez explicite sur le passage de l'ensembliste au vectoriel et au géométrique ; le passage à l'arithmétique, qui n'a rien de mystérieux, mérite quelques remarques supplémentaires ; ce qu'il met en jeu, en effet, est un autre mécanisme fondamental, celui de la fonction exponentielle.

III. — DU VECTORIEL AU MULTIPLICATIF : L'EXPONENTIELLE.

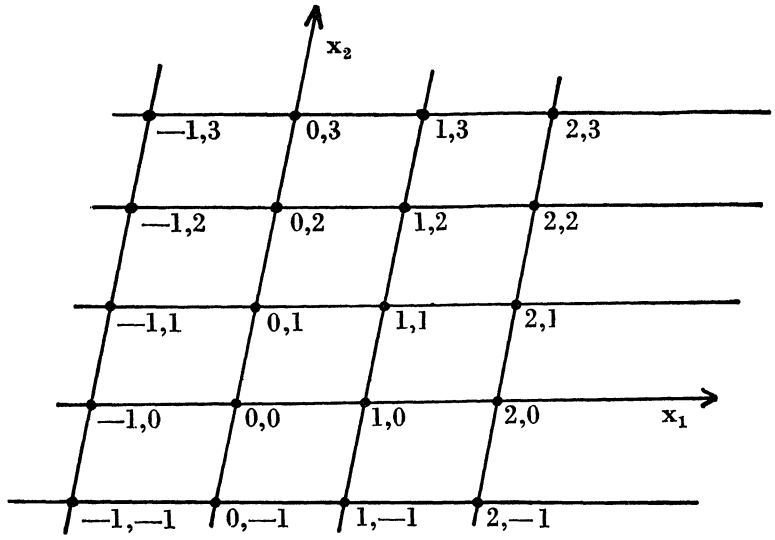
Abéliens, modules, vectoriels en treillis.

Le parallélotope de dimension n, représentatif de l'ensemble des parties d'un ensemble fini ordonnées par inclusion, est constitué des vecteurs $(x_1, x_2, \ldots, x_n)$ dont toutes les coordonnées sont égales soit à 0, soit à 1.

Considérons plus généralement l'ensemble de tous les vecteurs $(x_1, \ldots, x_n)$ dont les coordonnées sont des entiers positifs ou négatifs, c'est-à-dire appartiennent à l'ensemble $\mathbf{Z} = \{ \ldots, -2, -1, 0, 1, 2, 3, \ldots \}$; l'ensemble de ces vecteurs est noté $\mathbf{Z}^n$, et figuré ci-après pour le cas $n = 2$.

$\mathbf{Z}$ est totalement ordonné, dans l'ordre « naturel » :

$$\ldots < -2 < -1 < 0 < 1 < 2 < 3 < \ldots$$



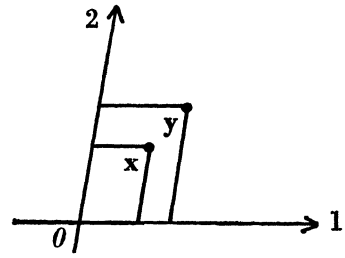
Il est d'autre part, muni de l'addition $+$, dont les propriétés sont bien connues.

Addition et ordre sont d'autre part *compatibles* en ce sens que si deux nombres sont dans un certain ordre, ajouter un même troisième nombre quelconque à chacun d'entre eux conserve l'ordre : l'ordre est invariant par translation.

Z, muni de son ordre et de son addition, soit $(\mathbf{Z}, <, +)$ constitue un groupe abélien totalement ordonné.

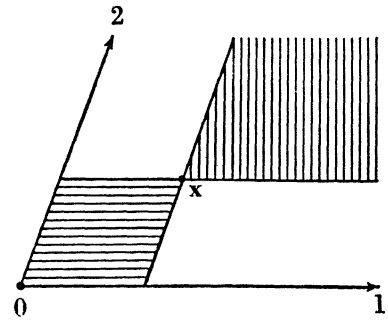
La structure de $(\mathbf{Z}, <, +)$ permet de définir sur $\mathbf{Z}^n$, dont les éléments $\mathbf{x} = (x_1, \ldots, x_n)$ sont des suites de n nombres, une *structure produit :* un ordre, une addition.

a) *Un ordre.* $\mathbf{x} = (x_1, \ldots, x_n)$ et $\mathbf{y} = (y_1, \ldots, y_n)$ étant deux éléments $\mathbf{Z}^n$, nous dirons que x est inférieur ou égal à y ($\mathbf{x} \leqq \mathbf{y}$) si et seulement si chaque coordonnée de x est au plus égale à la coordonnée correspondante de y. Dans le cas où $n = 2$, cela signifie que x et y sont disposés comme l'indique la figure.



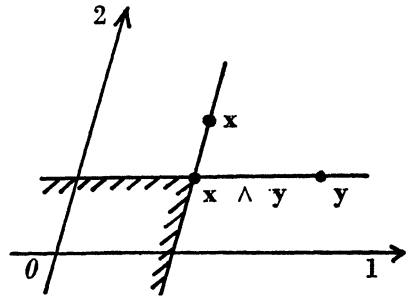
La relation ainsi définie est bien un ordre au sens mathématique.

C'est un *ordre partiel :* chaque élément x a des *minorants* (zone hachurée $\equiv$), des *majorants* (zone hachurée ||||), et des éléments qui ne lui sont pas comparables (zones en blanc).



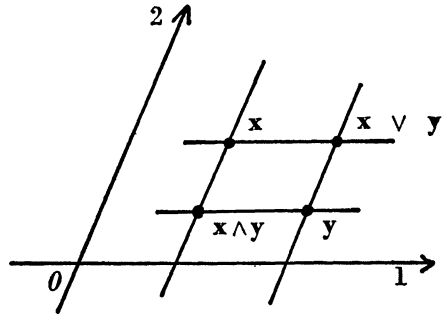
Enfin, la restriction de cet ordre aux seuls vecteurs de coordonnées égales à 0 ou à 1, est l'ordre d'inclusion entre les parties d'ensemble correspondantes : cet ordre généralise bien l'ordre partiel sur le parallélotope.

Les opérations ensemblistes d'union et d'intersection se généralisent aussi : l'intersection $\cap$ de deux parties X et Y d'un ensemble E, c'est la plus grande, dans l'ordre d'inclusion, de toutes les parties incluses à la fois dans X et dans Y. De même, si nous nous donnons dans $\mathbf{Z}^n$ deux vecteurs $\mathbf{x} = (x_1, x_n)$ et $\mathbf{y} = (y_1, y_n)$ il y a des vecteurs inférieurs à la fois à x et à y : ce sont ceux dont chaque coordonnée est au plus égale à la plus petite des deux coordonnées correspondantes de x et de y : parmi ces vecteurs minorants communs de x et de y, il y en a un qui est plus grand que tous les autres (un maximum) ; chacune de ses coordonnées est égale au minimum (à la plus petite) des deux coordonnées correspondantes de x et de y. Ce plus grand des minorants communs, à deux éléments x et y se note souvent $\mathbf{x} \wedge \mathbf{y}$, et s'appelle l'*infinum* de x et y (si en particulier $\mathbf{x} \leqslant \mathbf{y}$, on a : $\mathbf{x} \wedge \mathbf{y} = \mathbf{x}$).



Dualement, deux éléments x et y possèdent un *supremum :* le plus petit de leurs majorants communs, noté $\mathbf{x} \vee \mathbf{y}$.

Au moyen de l'ordre produit ainsi défini sur $\mathbf{Z}^n$, $(\mathbf{Z}^n, \geqslant)$ est muni de ce qu'on appelle une structure de *treillis* (ordre tel que tout couple d'éléments possède un suprenum et un infinum), et même de treillis *distributif :* chacune des opérations $\wedge$ et $\vee$ est distributive par rapport à l'autre, comme il est facile de le vérifier à partir de leurs définitions.



Cet ordre partiel doit être considéré dans toutes les questions « pratiques » où interviennent des classements et des choix. L'évocation d'un seul exemple suffira à le faire comprendre : des candidats à un concours sont soumis à plusieurs épreuves ; chaque candidat x est ainsi repéré par la suite $(x_1, x_2, \ldots, x_n)$ de ses notes, ou de ses rangs, à chacune des épreuves.

Dire que $\mathbf{x} \geqslant \mathbf{y}$ au sens qui a été défini (ordre produit : $x_1 \geqslant y_1$, $x_2 \geqslant y_2$, ..., $x_n \geqslant y_n$), c'est dire que x est meilleur que y dans toutes les épreuves : il y a unanimité du jury pour classer x avant y. Si x et y ne sont pas comparables, $\mathbf{x} \vee \mathbf{y}$ et $\mathbf{x} \wedge \mathbf{y}$ sont les deux candidats qui sont, ou seraient, parmi ceux qui sont comparables à x et à y, les plus proches de x et de y.

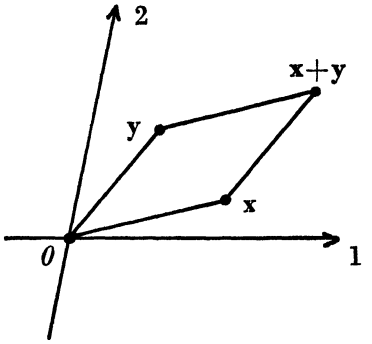
b) *Une addition.* Deux vecteurs $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ et $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ peuvent s'ajouter selon la règle :

$$\mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n).$$

Il s'agit bien d'une *addition*, en ce sens que l'opération ainsi définie possède les mêmes propriétés combinatoires que l'addition usuelle des nombres : le vecteur $O = (0, 0, \ldots, 0)$ joue le rôle d'élément neutre, cette addition est commutative et associative, chaque vecteur $\mathbf{z}$ a un opposé :

$$-\mathbf{x} = (-x_1, -x_2, -x_n).$$

Dans le cas de la dimension 2 ($n = 2$), ou 3, on vérifiera que cette addition a pour image dans le plan, ou l'espace, l'addition des vecteurs selon la vieille « règle du parallélogramme ».



On remarquera d'ailleurs que cette addition est définie, mutatis mutandis, comme l'ont été supra les opérations supremum ($\vee$) et infimum ($\wedge$) : coordonnée par coordonnée.

Muni de l'addition, l'ensemble $\mathbf{Z}^n$ des suites $(x_1, x_2, —, x_n)$ de n nombres entiers constitue un groupe abélien : $(\mathbf{Z}^n, +)$.

c) Ordre et addition sont reliés entre eux, les deux structures de treillis et de groupe abélien sur $\mathbf{Z}^n$ sont ce qu'on appelle *compatibles* l'une avec l'autre : si deux vecteurs sont comparables, et dans un certain ordre l'un par rapport à l'autre, ajouter un même troisième vecteur quelconque à chacun ne change pas leur ordre : $(\mathbf{Z}^n, +, \geqslant)$ constitue ainsi un groupe abélien ordonné (en treillis), ou module ordonné.

Lorsque l'on passe, par interpolation des nombres entiers, aux rationnels et aux réels, on obtient de même les espaces vectoriels ordonnés $\mathbf{Q}^n$ et $\mathbf{R}^n$.

Une autre façon, plus symétrique, d'exprimer la comptabilité des deux structures sur $\mathbf{Z}^n$ s'obtient de la manière suivante : considérons deux nombres a et b ; l'un des deux est le plus grand, et l'autre le plus petit, donc :

$$\max(a, b) + \min(a, b) = a + b.$$

Par conséquent, pour deux suites de nombres (deux vecteurs) :

$$\mathbf{x} = (x_1, x_2, \ldots, x_n) \qquad \text{et} \qquad \mathbf{y} = (y_1, y_2, \ldots, y_n),$$

composées comme il a été indiqué coordonnée par coordonnée, il vient :

$$\mathbf{x} \vee \mathbf{y} + \mathbf{x} \wedge \mathbf{y} = \mathbf{x} + \mathbf{y}.$$

Cette relation, qui met en jeu à la fois les deux opérations $\vee$ et $\wedge$ de treillis et l'opération $+$ de groupe abélien, fournit une clef pour comprendre les questions de *mesure* : appelons mesure (sur $\mathbf{Z}^n, +, \geqslant$) toute application m de cet abélien dans l'ensemble des nombres réels (dans le groupe abélien ordonné des réels) qui transporte la relation en question ; quels que soient $\mathbf{x}$ et $\mathbf{y}$, on a :

$$m(\mathbf{x} \vee \mathbf{y}) + m(\mathbf{x} \wedge \mathbf{y}) = m(\mathbf{x}) + m(\mathbf{y}).$$

Usuellement, on impose à l'application m des conditions supplémentaires telles que :

$$m(O) = 0 \qquad \text{et} \qquad \mathbf{x} \geqslant \mathbf{y} \Rightarrow m(\mathbf{x}) \geqslant m(\mathbf{y}).$$

On montre alors que :

$$m(\mathbf{x}) = m_1(x_1) + m_2(x_2) + \ldots + m_n(x_n)$$

où chaque fonction $m_i(x_i)$ est monotone de $\mathbf{Z}$ dans $\mathbf{R}$, et satisfait à $m_i(0) = 0$.

Si l'on choisit en particulier $m_i(x_i) = \lambda_i x_i$, avec $\lambda_i \geqq 0$, on obtient les mesures *linéaires*, qui sont en outre des morphismes pour la structure d'Abélien : $m(\mathbf{x}) + m(\mathbf{y}) = m(\mathbf{x} + \mathbf{y})$.

Dans le cas des parties d'un ensemble $\mathrm{E} = \{1, 2, \ldots, n\}$, ces conditions deviennent :

$$\forall\, \mathrm{X}, \forall\, \mathrm{Y} \subset \mathrm{E},\ m(\mathrm{X} \cup \mathrm{Y}) + m(\mathrm{X} \cap \mathrm{Y}) = m(\mathrm{X}) + m(\mathrm{Y})$$

$$m(\varnothing) = 0\,; \qquad m(\mathrm{X}) \geqslant 0.$$

Comme chaque partie, X est définie par un vecteur $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ de $\mathbf{Z}^n$ dont toutes les coordonnées valent 0 ou 1 ($x_i = 1 \Leftrightarrow i \in \mathrm{X}$), on a alors, en appliquant le résultat général qu'on vient de voir :

$$m(\mathrm{X}) = m(\mathbf{x}) = \sum_{i \in \mathrm{X}} p_i$$

où :

$$p_i = m_i(1) \geqslant 0 \quad (m_i(0) = 0).$$

On retrouve bien que la mesure m (X) de chaque partie X est la somme des masses p_i attachées à chacun de ses éléments. Si on impose en outre à la mesure m sur E d'être *normée*, i.e. que :

$$m(\mathrm{E}) = 1$$

alors :

$$\mathbf{p} = (p_1, p_2, \ldots, p_n)$$

satisfait en outre à :

$$p_1 + p_2 + \ldots + p_n = 1$$

$$(m(\mathrm{E}) = {}^{p}{}_1 + {}^{p}{}_2 + \ldots + {}^{p}{}_n),$$

$\mathbf{p}$ est une distribution de probabilité et m est une mesure de probabilité.

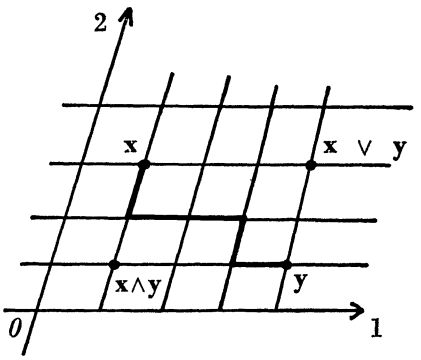
Autre retour en arrière : les distances dont il a été question plus haut (§ *II*) entre parties d'un ensemble ne sont également que la particularisation au parallélotope des distances définies dans $\mathbf{Z}^n$ à partir d'une mesure ; on pose :

$$d(\mathbf{x}, \mathbf{y}) = m(\mathbf{x} \vee \mathbf{y}) - m(\mathbf{x} \wedge \mathbf{y}).$$

Si la mesure m est strictement monotone :

$$(\mathbf{x} < \mathbf{y} \Rightarrow m(\mathbf{x}) < m(\mathbf{y}))$$

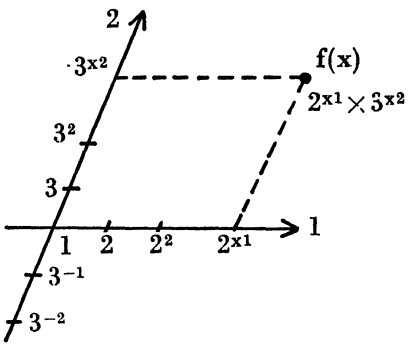
d est une distance qui s'interprète comme la longueur commune des chemins maximaux de longueur minimum du point x au point y dans le « quadrillage » de l'espace défini par les points à coordonnées entières.

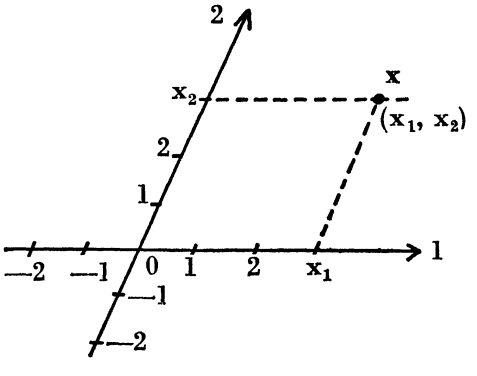


L'exponentielle.

Nous venons de plonger le parallélotope de dimension n, dont les 2^n sommets représentent les parties d'un ensemble E ayant n éléments, dans le module réticulé $(\mathbf{Z}^n, +, \geqslant)$ de dimension n ; or, nous avons vu plus haut (§ *II*) que l'ensemble des parties d'un ensemble fini est représentable par l'ensemble des diviseurs d'un nombre entier naturel de la forme $p_1 \times p_2 \ldots \times p_n$ où $p_1, p_2, \ldots, p_n$ sont des nombres premiers distincts. Ceci nous conduit à rechercher l'analogue de $(\mathbf{Z}^n, +, \geqslant)$ dans le langage de la divisibilité des nombres.

La réponse est simple : plaçons nous par exemple dans le cas où $n = 2$, et à chaque vecteur $\mathbf{x} = (x_1, x_2)$ de $\mathbf{Z}^2$, faisons correspondre le nombre (entier si x_1 et x_2 sont positifs ou nuls, rationnels sinon) $f(\mathbf{x}) = 2^{x1} \times 3^{x2}$.





Il est clair que :

a′) La fonction f est monotone : si $\mathbf{x} \leqslant \mathbf{y}$ en ce sens que $x_1 \leqslant y_1$ et $x_2 \leqslant y_2$, alors $f(\mathbf{x})$ divise $f(\mathbf{y})$.

b′) Cette fonction transforme les sommes vectorielles en produits de nombres :

$$f(\mathbf{x} + \mathbf{y}) = 2^{x_1+y_1} \times 3^{x_2+y_2} = (2^{x_1} \times 3^{x_2}) \times (2^{y_1} \times 3^{y_2}) = f(\mathbf{x}) \times f(\mathbf{y}).$$

f est un isomorphisme pour l'ordre, et entre addition et multiplication, entre $\mathbf{Z}^2$ et l'ensemble des nombres rationnels de la forme $2^{x_1}\,3^{x_2}$ (x_1 et x_2 entiers) ; elle conserve l'ordre, et transforme les sommes en produits : c'est une *fonction exponentielle.*

La transformation par f de l'ordre partiel de $\mathbf{Z}^2$, en ordre par divisibilité a pour conséquence que :

$$f(\mathbf{x} \vee \mathbf{y}) = \text{ppcm}\,(f\mathbf{x}, f\mathbf{y}),$$
$$f(\mathbf{x} \wedge \mathbf{y}) = \text{pgcd}\,(f\mathbf{x}, f\mathbf{y}).$$

En particulier, la relation fondamentale :

$$\mathbf{x} \vee \mathbf{y} + \mathbf{x} \wedge \mathbf{y} = \mathbf{x} + \mathbf{y}$$

a pour image la propriété classique :

$$\mathrm{X} \times \mathrm{Y} = \text{ppcm}\,(\mathrm{X}, \mathrm{Y}) \times \text{pgcd}\,(\mathrm{X}, \mathrm{Y}).$$

Cet isomorphisme rend intelligible la possibilité de passage du langage vectoriel, ou géométrique (simplexe géométrique) au langage arithmétique pour la description de notre petit objet d'étude. Or cet isomorphisme n'est autre que cet antique ingrédient de tous les cours de mathématiques qu'est la fonction exponentielle ; tout repose sur la vieille correspondance entre « progressions arithmétiques » (les x_i) et « progressions géométriques » (les $p_i{}^{xi}$). Le passage de l'additif vectoriel au multiplicatif est d'ailleurs largement exploité sous la forme graphique : le même « papier millimétré » peut servir, selon le code utilisé sur les graduations équidistantes, de papier « à échelles arithmétiques » (code additif, x_i) ou de papier « à échelles logarithmiques » (code multiplicatif, 2^{xi}).

Parties d'ensembles finis, diviseurs d'entiers naturels, parallélotopes, simplexes géométriques, quatre représentations d'une même abstraction, chacune avec ses prolongements vers des chapitres principaux des mathématiques ; à l'intérieur de chaque classe représentative, les enchaînements entre objets de la classe et, pour chaque couple de représentations, le moyen de passer l'une à l'autre (les isomorphismes).

Aucune de ces représentations n'est à privilégier d'emblée ; leur mise en parallèle constitue au contraire, pensons-nous, un bon moyen d'introduire, en s'appuyant sur des exemples simples, aux notions de catégorie et de foncteur : or le langage des catégories apparaîtra bientôt aux pédagogues comme aussi nécessaire, parce qu'il fournit des « mots outils » aux mathématiques, que ne l'est apparu naguère le langage ensembliste.

M.S.H. (7e année, no 27, 1969, p. 47-53)

PROBLÈMES D'ENSEIGNEMENT

APPLICATIONS PRATIQUES DES LOIS DE PROBABILITÉ (4)

par

B. LECLERC

LOI BINOMIALE NÉGATIVE *ÉCONOMIE*

A. S. C. EHRENBERG, « The pattern of consumers purchases », *Applied statistics*, Vol. 8, p. 26-41, 1959.

La distribution du nombre d'unités d'une denrée à emballage standard (pain, céréales pour déjeuner, nourriture pour chiens, café, savonnettes, etc.) achetées par ménage et par unité de temps, s'ajuste bien à la loi binomiale négative. Parmi les modèles mathématiques conduisant à cette loi, l'auteur retient celui qui suppose poissonnienne la distribution des achats par unité de temps d'un consommateur, et de type γ (ou du χ^2) la distribution des moyennes des achats par unité de temps des consommateurs (voir fiche SICHEL, *M.S.H.*, no 25). La forme de cette dernière distribution rend plausible l'hypothèse.

Modèle.

La probabilité P_r d'observer l'achat de r unités du produit donné par un ménage en une unité de temps est donnée par :

$$P_r = \left(1 + \frac{m}{k}\right)^{-k} \frac{\Gamma(k + r - 1)}{\Gamma(r)\,\Gamma(k - 1)} \left(\frac{m}{k + m}\right)^r$$

$r = 0, 1, 2, \ldots$

m est la moyenne théorique et k un paramètre positif.

Estimation.

m est estimé par la moyenne empirique. C'est l'estimateur du maximum de vraisemblance, sans biais.

Par contre, l'estimation de k par le maximum de vraisemblance est malaisée.

Par ailleurs, la méthode des moments — identification de la variance empirique avec la variance théorique $m\left(1 + \frac{m}{k}\right)$ — fournit un estimateur peu efficace de k.

L'auteur préconise l'identification de $P_0 = \left(\frac{1 + m}{k}\right)^{-k}$ avec la donnée empirique correspondante, d'où une équation se résolvant par itérations. L'efficacité de cet estimateur est toujours bonne.

Test de convenance du modèle.

Le test du χ^2 peut être utilisé. Il est plus rapide de comparer la variance empirique à la variance théorique calculée à partir des paramètres estimés. Anscome et Evans ont donné une formule pour la variance de la différence de ces quantités.

Application.

Un exemple numérique pour une denrée non précisée, est fourni par l'auteur (échantillons portant sur 2 000 ménages et 26 semaines). L'ajustement semble très bon. Un graphique résume les résultats du test ci-dessus pour l'ensemble des distributions ajustées (150 environ). Les deux valeurs à comparer sont toujours proches.

L'auteur s'intéresse aussi à l'ajustement obtenu pour des sous-groupes de la population (classes d'âge, ou habitat géographique). L'ajustement à la loi binomiale négative reste bon en pratique, ce que l'auteur juge être en contradiction avec le modèle mathématique sous-jacent. L'article se termine sur une étude de l'estimation de la variance de la moyenne empirique.

RÉFÉRENCES BIBLIOGRAPHIQUES.

R. D. WADSWORTH, « The experience of a user of a consumer panel », *Applied statistics*, 1, p. 169, 1959.

F. J. ANSCOMBE, « Sampling theory of the negative binomial distribution », *Biometrika*, 37, p. 308, 1950.

D. A. EVANS, « Experimental evidence concerning contagious distributions in ecology », *Biometrika*, 40, p. 186, 1953.

F. J. ANSCOMBE, « The transformation of Poisson, binomial and negative binomial data », *Biometrika*, 35, p. 246, 1948.

LOI LOG-NORMALE
LOI EXPONENTIELLE — LOI GAMMA *MÉDECINE*

John W. BOAG, « Maximum Likelihood estimates of the proportion of patients cured by cancer therapy », *Journal of the Royal Statistical Society*, Vol. II, p. 15-44, 1949.

ÉTUDE DES TEMPS DE SURVIE.

L'auteur a étudié la distribution des temps de survie de malades de cancer, dans le but de déterminer la proportion de guéris après traitement.

Modèle.

Tout d'abord sont comparés l'ajustement à la loi log-normale et celui à la loi exponentielle. Pour la première l'auteur précise la forme de la loi. Si $x = \text{Log } t$ est le logarithme du temps de survie, la distribution de x a pour densité :

$$F(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

pour tout x réel, μ et $\sigma > 0$ sont des paramètres.

Estimation.

La méthode d'estimation des paramètres n'est pas donnée. L'auteur précise que les deux distributions théoriques ajustées (log-normale et exponentielle) ont même médiane.

Test du χ^2 sans seuil précisé.

On donne la probabilité P pour une variable du χ^2 d'être supérieure à la quantité test.

Application.

Application à quatre distributions suivant la localisation du mal fournies par des hôpitaux (temps de survie après les soins de malades décédés du cancer pour lequel ils ont été soignés). Pour trois d'entre elles l'ajustement à la loi log-normale est bon, (P entre 0,24 et 0,32), mais non celui à la loi exponentielle ($P = 10^{-5}$). Pour la quatrième, le résultat est sensiblement différent : l'ajustement exponentiel est suffisant ($P = 0{,}13$), mais non celui à la loi log-normale.

L'auteur fournit les données pour ces quatre distributions et donne des tracés de droites de Henry, ainsi que pour quatre autres dont trois (données de Greenwood) correspondent à des malades non soignés. Pour deux d'entre elles, l'ajustement log-normal ne convient pas aux plus courtes durées de vie, mais s'accorde avec les plus longues.

D'autres modèles ont été testés, avec les densités :

$$F(t) = \alpha^2 t\, e^{-\alpha t}$$
$$F(t) = 2\,\alpha\, t\, e^{-\alpha t^2}$$
$$F(t) = \frac{\alpha^3 t^2}{2} e^{-\alpha t}$$

Le premier a été le meilleur, mais ne s'ajuste pas à toutes les distributions. Notons que c'est, ainsi que le troisième, une loi de type gamma.

II. — ÉTUDE DE LA GUÉRISON.

Modèle.

On a une proportion c de patients guéris. Une proportion $(1 - c)$ de patients non guéris dont les durées de survie sont supposées distribuées suivant la loi log-normale, de paramètr s μ et σ.

Estimation.

L'auteur présente les calculs auxquels conduit la méthode du maximum de vraisemblance dans trois cas.

A. — Estimation du triplet (c, μ, σ).
B. — Estimation du couple (c, μ), σ étant supposé connu.
C. — Estimation de c, μ et σ étant supposés connus.

Application de ces calculs à quelques distributions.

Pas de test dans cette partie de l'article. Celui-ci est suivi d'une discussion.

RÉFÉRENCES BIBLIOGRAPHIQUES.

Bliss & Stevens, « The calculation of the mortality curve », *Ann. of Appl. Biology*, 24, p. 815, 1937.

J. W. Boag, « The presentation and analysis of the results of radiotherapy », *Brit. Journ. Rad.*, 21, p. 128-189, 1948.

D. J. Finney, *Probit Analysis*, Cambridge University Press, p. 210, 1947.

R. A. Fisher, « On the mathematical foundations of theorical statistics », *Phil. Transactions Royal Soc.*, A 222, p. 309, 1922.

J. H. Gaddum, « Log-normal distributions », *Nature*, 156, p. 463, 1945.

F. Garwood, « The application of maximal likelihood to dosage mortality curves », *Biometrika*, 32, p. 46, 1940.

Greenwood, « The natural duration of cancer », *Reports on public health and medical subjects*, n° 33, H.M.S.O., 1926.

LOI NORMALE — *INDUSTRIE*
LOI DE WEIBULL — *DURÉE DE VIE*

A. H. Zaludova, « Problèmes de durée de vie. Applications à l'industrie automobile », *Revue de Statistique appliquée*, vol. XIII, n° 4, p. 75, 1965.

L'auteur fait un exposé général sur les problèmes de durée de vie, avec référence particulière à l'industrie automobile : les données présentes dans l'article proviennent d'un parc de taxis de Prague (voitures de tourisme Skoda).

Modèles.

Parmi quatorze modèles présentés, deux sont retenus pour la loi des durées de vie de pièces d'automobiles soumises à l'usure mécanique.

1) Modèle normal, comme forme asymptotique de la distribution gamma. Pour celle-ci, la densité de probabilité de la durée de vie est :

$$\frac{\lambda^B}{\Gamma(B)} x^{B-1} e^{-\lambda x} \qquad x \geqq 0$$

où $\lambda > 0$, $B > 1$ sont des paramètres. Γ est la fonction gamma. Pour la distribution normale, la densité est, pour tout x réel :

$$\frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$$

où m est la moyenne et σ l'écart-type.

Ce modèle est le plus approprié avant la première révision générale.

2) Après celle-ci, la nature des phénomènes d'usure se modifie un peu (augmentation des jeux entre les pièces), et le modèle préconisé par l'auteur est celui de Weibull, pour lequel la densité de probabilité des temps de survie est :

$$\frac{B x^{B-1}}{\delta^B} e^{\left(\frac{x}{\delta}\right)^B} \qquad x \geqq 0$$

où $\delta > 0$ et $B > 1$ sont des paramètres.

Estimation.

Seule l'estimation des paramètres de Weibull pour des populations et échantillons tronqués (les plus grandes durées de vie n'ont pu être recensées) est étudiée ici. La méthode est celle du maximum de vraisemblance, aboutissant à des équations compliquées.

Pas de *test.*

Deux graphiques sont donnés pour deux pièces particulières. Un tableau permet la comparaison des valeurs empiriques et théoriques pour une troisième, dans le cas de la distribution de Weibull tronquée.

RÉFÉRENCES BIBLIOGRAPHIQUES. Dix-neuf titres, dont :

W. MARSHALL, «A bibliography on life-testing and related topics», *Biometrika*, p. 521-543, 1958.
B. EPSTEIN, « Recents developments in life-testing», *Bull. Inst. Int. de Stat.*, 33e session, 1961.
A. KAUFFMANN, *Modèles et méthodes de la recherche opérationnelle*, Dunod, Paris, 1959.
W. WEIBULL, « The phenomenon of rupture in solids», *Ingen. Velensk. Akad. Handl.*, 2, p. 1939, 1953.

LOIS DE YULE, BOREL et FISHER
LOI DE ZETA — *PSYCHOLOGIE*

Frank A. HAIGHT, « Some Statistical problems in connection with word association data», *Journal of Mathematical Psychology*, Vol. 3, p. 217-233, 1966.

Les données sur les tests d'association de mots, telles celles de Horvath (1963) sont difficiles à analyser statistiquement : manque de modèle convainquant, très longues branches infinies, etc.

Une série d'autres phénomènes présente la même caractéristique fondamentale : ordonnancement, selon les rangs (approche de Zipf) ou selon les fréquences (approche de Yule).

Exemples :

1. — Mots dans un livre donné.
2. — Scientifiques en fonction du nombre d'articles qu'ils ont publiés.
3. — Automobiles par marque ou par couleur.
4. — Individus en fonction de leurs revenus.
5. — Villes en fonction de leur taille.
6. — Genres biologiques en fonction du nombre d'espèces qu'ils contiennent,

etc. (voir Simon, 1955).

Modèles.

1) Distribution de Yule, de paramètre η :

$$P_n = \eta\, \mathrm{B}(n, \eta + 1) \qquad \text{où} \qquad \mathrm{B}(x, y) = \int_0^1 t^{x-1} (1 - t)^{y-1}\, dt.$$

Proposée par Yule (1924), travaillant sur l'exemple (6).

2) Distribution de Borel (1942), de paramètre β :

$$P_n = \frac{n^{n-2}}{(n-1)!}\, e^{-\beta n} \beta^{n-1}$$

proposée par Borel travaillant sur un modèle de files d'attente, et tabulée par Haight et Breuer (1960).

3) Distribution de la série logarithmique de Fisher, de paramètre φ

$$P_n = \frac{-\varphi^n}{n \operatorname{Log}(1-\varphi)}$$

4) L'auteur propose également, à partir des travaux de Zipf, la distribution Zéta de paramètre σ

$$P_n = \frac{1}{(2n-1)^{1/6}} + \frac{1}{(2n+1)^{1/6}}$$

P_n représentant partout ci-dessus la probabilité de l'entier n positif ou nul.

Estimation.

Toutes ces lois ont un paramètre unique. L'auteur propose trois méthodes d'estimation (identification de valeurs empiriques et théoriques) :

a) par la moyenne m,

b) par P_1,

c) par $q_N = \sum_{j=N+1}^{\infty} P_j$ pour un N choisi.

Il fournit les résultats suivants :

1) Loi de Yule : $m = \frac{\eta}{\eta - 1}$ $\quad p_1 = \frac{\eta}{\eta + 1}$ $\quad q_{N-1} = \eta\, B(N, \eta)$.

2) Loi de Borel : $m = \frac{1}{1-\beta}$ et pour les méthodes *b*) et *c*) se référer aux tables.

3) Loi de Fisher : $m = \frac{-\varphi}{(1-\varphi)\operatorname{Log}(1-\varphi)}$

Les trois méthodes demandent l'utilisation de tables qui ont été préparées par Haight et Reichenbach (1965) et seront publiées séparément.

4) Loi Zéta : $m = (1 - \frac{1}{2^{1/\sigma}})\ \zeta\left(\frac{1}{\sigma}\right)$

où ζ est la fonction ζ de Riemann.

Seule la méthode *a*) est proposée pour cette loi-ci.

L'auteur signale que dans les cas 2) et 3) la méthode *a*) est celle du maximum de vraisemblance.

Test de Kolmogorov-Smirnov.

Cependant l'auteur précise qu'il utilise la quantité — test uniquement pour comparer les ajustements entre eux car, à en juger par les commentaires de Smirnov (1948), Massey (1951) et Siegel (1956), les niveaux de signification de la statistique de Smirnov n'ont pas été établis dans le cas d'une distribution discrète.

Application.

Aux données de Horvath : résultat du test d'association de mots de Kent-Rosanof, pour les quatre mots Table, Music, Blossom, High.

Un tableau donne les quantités — test pour tous les ajustements proposés. Les ajustements de Yule sont dans l'ensemble meilleurs que ceux de Borel, eux-mêmes supérieurs à ceux de Fisher. L'ajustement Zéta semble d'autant meilleur, que la moyenne est grande (test facile). Il l'emporte nettement dans deux cas.

RÉFÉRENCES BIBLIOGRAPHIQUES

F. A. HAIGHT, M. A. BREUER et H. REICHENBACH, « The Borel-Tanner Distribution », *Biometrika*, 1960, 47, 143-50. « Fisher's distribution », *Research Report* n° 38, Institute of Transportation and Traffic Engineering, Univ. of Calif., Los Angeles, 1965.

E. BOREL, « Sur l'emploi du Théorème de Bernoulli pour faciliter le calcul d'une infinité de coefficients. Application au problème de l'attente à un guichet », *Compte rendus hebdom. des séances de l'Acad. des Sc.*, 1942, 214, 452-456.

W. HORVATH, « A stochastic Model for word association tests », *Psych. Rev.*, 1963, 70, 361-64.

F. J. MASSEY, « The Kolmogorov-Smirnov Test of goodness of fit », *Journal of the Am. Stat. Soc.*, 1951, 46, 68-70.

N. A. RUSSEL and J. J. JENKINS, « The complete Minnesota norms for responses to 100 words for the Ken-Rosanof word association test », *Techn. Report* n° 2, 1954, Univ. of Minnesota, Department of Psychology.

H. A. SIMON, « On a class of skew distribution function », *Biometrika*, 1955, 42, 425-40.

SIEGEL, *Non parametric statistics for the behavioral sciences*, New York, Mc Graw-Hill, 1956.

SMIRNOV, « Table for estimating goodness of fit », *AMS*, 1948, 19, 278-81.

G. U. YULE, « A mathematical theory of evolution, based on the conclusions of Dr J. C. Willis, F.R.S. », *Phil. Trans. Royal Soc. of London*, Series B, 1925, 213, 21-87.

M.S.H. (7e année, no 27, 1969, p. 55-57)

L'EMPLOI DES CALCULATEURS EN ARCHÉOLOGIE

PROBLÈMES MATHÉMATIQUES ET SÉMIOLOGIQUES

Organisé par le Centre d'Analyse Documentaire pour l'Archéologie, le Colloque International du C.N.R.S. qui s'est tenu à Marseille du 7 au 12 avril 1969 a réuni, sur le thème qui donne son titre à cette note, une soixantaine de participants venant des principaux pays des deux Amériques et d'Europe (U.R.S.S. exceptée, les savants soviétiques invités n'étant pas en mesure de se joindre au Colloque). L'objectif de la réunion n'était pas de procéder à une revue de l'ensemble des applications des calculateurs à l'archéologie ; il s'agissait plutôt, à partir des expériences poursuivies durant ces dernières années, de tenter de dégager les grandes lignes d'une théorie archéologique en cours d'élaboration et plus particulièrement les incidences sémiologiques et mathématiques d'une construction de cet ordre.

Par problèmes sémiologiques ou linguistiques on entend d'abord ceux qui se posent dans la représentation symbolique (analyse descriptive) des documents de l'archéologie : monuments, objets, textes, informations stratigraphiques, etc. Mais ils interviennent aussi dans la phase proprement cognitive (analyse structurale) par le biais des modèles élaborés par les linguistes pour les besoins propres de leur discipline et dont il s'agit d'examiner s'ils sont ou non utilisables par l'archéologie. C'est naturellement à ce niveau, dans le traitement systématique des données résultant de l'analyse descriptive, que les méthodes mathématiques et leur prolongement informatique apparaissent comme les moyens susceptibles de donner aux classifications et aux sériations la rigueur propre à toute démarche scientifique.

Cependant, l'utilisation d'outils formels, sémiologiques ou mathématiques, ne saurait être tenus pour satisfaisante si elle ne visait à leur intégration dans l'ensemble de la démarche intellectuelle de l'archéologue, ce qui suppose bien entendu la modification de cette démarche mais aussi la critique et l'adaptation des formalismes en fonction des conditions spécifiques de l'archéologie.

Sous des formes et à des degrés divers, ces trois thèmes fondamentaux se retrouvent dans la quasi-totalité des communications présentées à Marseille. Certaines d'entre elles, en premier lieu, se sont attachées à définir en termes de cohérence le cadre général des problèmes de méthode, qu'il s'agisse des rapports entre les différents types de données (données intrinsèques vs. données extrinsèques) ou de l'interaction des procédures formelles et des démarches empiriques dans la résolution d'un problème. (Deshayes, Moberg, Soudsky). En même temps qu'ils illustraient les possibilités d'emploi des calculateurs, les exposés consacrés à l'analyse sémiologique montraient la diversité des interventions de cette discipline, qu'il s'agisse des applications documentaires (Rogers) ou des problèmes que pose la transcription de textes rédigés dans une langue imparfaitement connue (Heyler et ass.) ; en même temps étaient soulignées les limites mathématiques et logiques d'une telle analyse, intrinsèquement (Jaulin) ou par suite de l'empirisme de ses prolongements classificatoires (Borillo). On aimerait pouvoir écrire que la discussion sur la fécondité des modèles linguistiques et sur les conditions de leur validité pour l'Archéologie (Hymes) trouvait par ailleurs (Kovalevskaya) un premier exemple d'application. En fait,

alors que le premier traitait essentiellement des modèles linguistiques au sens des grammaires transformationnelles ou génératives, la deuxième fondait ses résultats sur les théories probabilistes du langage.

Pour les méthodes mathématiques, deux exposés généraux permettaient de faire le point des orientations principales, statistiques (MILLIER et TOMASSONE) et classificatoires (SPARCK-JONES). S'il fallait ordonner les communications dont il n'a pas encore été fait état, nous ajouterions à ce premier critère celui de l'attention prioritaire selon qu'elle est portée à l'application d'un outil mathématique pour la résolution d'un problème archéologique particulier (COWGILL, ELISSEEFF, SHER) ou au contraire à la définition de méthodes mathématiques originales, relevant de l'analyse factorielle (IHM, LINGOES) ou de la taxinomie numérique (BORDAZ et BORDAZ). L'examen des propriétés purement mathématiques des classifications a fait par ailleurs l'objet de certains autres travaux (LERMAN, RÉGNIER, de la VEGA) tandis qu'en sens inverse les possibilités d'utilisation heuristique du calculateur par simulation du raisonnement archéologique étaient également explorées (DORAN).

Mais ce compte rendu ne serait pas fidèle s'il se bornait à découper l'ensemble des communications en des groupes étanches voués à l'étude de difficultés parcellaires. Nombreux ont été les exposés consacrés à la définition et à l'élucidation des problèmes d'intégration des procédures formelles, entre elles d'abord (propriétés des systèmes sémiologiques et conditions de validité des algorithmes), dans l'ensemble du raisonnement archéologique d'autre part (formalismes et interprétation). Ces préoccupations se sont affirmées rapidement au fil des discussions, jusqu'à prendre une place prépondérante dans l'exposé conclusif (J.-C. GARDIN) et la discussion finale.

Longtemps étrangère aux bouleversements méthodologiques qui ont affecté les sciences de l'homme, l'archéologie témoigne aujourd'hui de sa maturité par la nature des questions qu'elle se pose.

LISTE DES COMMUNICATIONS.

BORDAZ, V. et J. (Computation Center, University of Montréal), A computer-assisted pattern recognition method of classification and seriation applied to archaeological material.

BORILLO, M. (Centre d'Analyse Documentaire pour l'Archéologie, Marseille), La vérification des hypothèses en archéologie. Deux pas vers une méthode.

COWGILL, G. (Department of Anthropology, Brandeis University), Some Sampling and reliability problems in Archaeology.

DESHAYES, J. (Institut d'Archéologie, Université de Paris), Points de vue subjectifs sur la construction d'une typologie.

DORAN, J. (Department of Machine Intelligence, University of Edinburgh), Archeological reasonnig and machine reasoning.

ELISSEEFF, V. (Musée Cernuschi, Paris), Données de classement fournies par les scalogrammes privilégiés.

HEYLER, A.; LESLANT, J.; MARETTI, E.; ZARRI, G. P. (É.P.H.É. VI^{e} section et Centro di Cibernetica e di Attivita Linguistiche, Universita degli Studi Milano), Problèmes relatifs à l'enregistrement et au traitement de documents épigraphiques rédigés dans une langue très imparfaitement connue, le méroïtique.

HYMES, D. (Department of Linguistics, University of Pennsylvania), Linguistic models in Archaeology.

IHM, P. (Institut für Mediziniche-Biologische Statistik, Marburg Universität), Distance et similitude en taximétrie.

JAULIN, B. (Centre de Calcul, Maison des Sciences de l'Homme, Paris), Mesure de la ressemblance en anarchéologie.

KOVALEVSKAYA, V. B. (Institut d'Archéologie, Moscou), Recherches sur les systèmes sémiologiques en archéologie par les méthodes de la théorie de l'information.

LERMAN, C. (Centre de Calcul, Maison des Sciences de l'Homme, Paris), H-Classificabilité.

LINGOES, J. (Computing Center, University of Michigan), A general non-parametric model for representing objects and attributes in a joint metric space.

MILLIER, C.; TOMASSONE, R. E. (Institut National de la Recherche Agronomique), Méthodes d'ordination et de classification : leur efficacité et leurs limites.

MOBERG, C. A. (Université de Göteborg), Au-delà des classifications. Remarques pragmatiques sur quelques problèmes théoriques dans l'archéologie de la Scandinavie Occidentale aux environs de 1300 ans av. J.-C.

RÉGNIER, S. (Centre de Calcul, Maison des Sciences de l'Homme, Paris), Non fécondité du modèle statistique général de la classification automatique.

ROGERS, D. J.; ESTABROOK, G. F. (Department of Biology, University of Colorado), Theoretical and practical consideration on data structuring for a computerized information retrieval system.

SHER, J. A. (Institut d'Archéologie, Moscou), Un algorithme de classification typologique.

SOUDSKY, B. (Institut d'Archéologie, Université de Prague), Le problème des propriétés dans les ensembles archéologiques.

SPARCK-JONES, K. (The University Mathematical Lab. Cambridge), The evaluation of archaeological classification.

VEGA, W. F., DE LA (Centre d'Analyse Documentaire pour l'Archéologie, Marseille), Quelques propriétés des hiérarchies de classifications.

Les actes du Colloque seront publiés dans le courant de 1970 par les soins du C.N.R.S.

M. Borillo.