Problèmes d'enseignement

Mathématiques et sciences humaines, tome 8 (1964), p. 27-37
<http://www.numdam.org/item?id=MSH_1964__8__27_0>

PROBLEMES D'ENSEIGNEMENT

A. BELTRAMONE

NOTE SUR LA DETERMINATION DES VALEURS TYPIQUES DE POSITION (1)

Les "quantiles" sont d'une utilisation courante en statistique descriptive et tous ceux qui ont l'occasion de s'en servir ne rencontrent aucune difficulté. Il n'en est plus de même quand on s'adresse à un public d'étudiants. C'est cet aspect qui nous intéresse ici et que nous développerons en tenant compte des réactions que nous avons pu observer dans trois promotions successives.

Cette particularité explique le plan que nous conviendrons de suivre. En bonne règle, il faudrait commencer par la définition des "quantiles". Or, nous serons conduit à une définition qui peut surprendre quand on la présente "ex abrupto". Cette définition viendra donc après l'exposé détaillé des réactions observées. Cependant, nous ne pouvons pas donner ces réactions sans avoir présenté, au préalable, quelques remarques sur la structure ordinale.

I – REMARQUES SUR LA STRUCTURE ORDINALE

L'information statistique recueillie dans le but d'étudier un phénomène donné se présente assez souvent sous la forme d'une collection d'observations ayant les caractéristiques d'un ensemble fini, dénombrable et totalement ordonné. Il arrive encore qu'en présence d'une collection d'observations dotée d'une structure plus riche que la structure ordinale, il soit suffisant, pour l'étude envisagée, de retenir seulement la structure ordinale.

Dans tous ces cas, la connaissance de la collection nous est apportée par une énumération exhaustive des éléments de la collection. Evidemment, nous choisissons l'énumération qui nous apporte la plus grande information et qui traduit donc l'une des propriétés essentielles de la collection: la possibilité d'être totalement ordonnée.

Il convient de s'arrêter un instant à cet ordre total.

La collection est formée d'éléments, d'entités, qui sont indivisibles et qui constituent autant de réalités distinctes. Désignons par E_i une collection d'éléments et par n_i le nombre d'éléments de cette collection; par e un élément quelconque. Il est possible que nous soyions dans l'obligation de grouper un certain

(1) Qu'il me soit permis d'exprimer ici toute ma gratitude à Monsieur Jacques Voranger, Chef de travaux à la Faculté de Droit et des Sciences Economiques d'Aix, Directeur du laboratoire de statistique et d'économétrie, pour tous les conseils qu'il dispense avec autant de bonne grâce que de compétence à tous ceux qui cherchent à progresser dans la dure voie de la statistique.

nombre d'éléments e, pourtant distincts, mais qui du point de vue du caractère étudié se ressemblent étrangement. Nous sommes souvent conduit à faire dans E_i une classification c'est-à-dire à établir une distinction sommaire entre les éléments. Soit un caractère donné, nous conviendrons de marquer les éléments d'un signe particulier pris dans la suite des signes I, II, III, ..., K. Un élément marqué I sera dit "l'un des plus petits", un élément marqué K sera dit "l'un des plus grands. Il y aura la classe "des plus petits éléments", toutes les classes intermédiaires II, III, etc..., enfin la classe "des plus grands éléments", K.

Plusieurs éléments peuvent ainsi porter la même marque et cette marque ne permet pas de différencier tous les éléments. Nous dirons que l'ensemble M des marques:

$$(1) \qquad M = \left\{ I,\ II,\ III,\ \ldots,\ K \right\}$$

détermine un préordre dans la collection E_i. Nous n'avons pas encore atteint l'ordre total. Pour celà, nous conviendrons de prendre les éléments dans l'ordre des marques et de répérer chaque élément par une deuxième marque qui sera celle du rang dans lequel arrive chaque élément dans un tirage exhautif. Autrement dit, la collection E_i se présentera sous la forme: $I_1, I_2, I_3, I_4, II_5, II_6, III_7, IV_8, IV_9, V_{10}$, par exemple, où les indices inférieurs indiquent l'ordre de tirage. La classe I contient 4 éléments différenciés par leur ordre de tirage, etc.... Connaissant le rang du tirage, 8 par exemple, nous avons immédiatement la marque du préordre, qui est IV. Par la suite, nous appellerons "valeur" de l'élément la marque du préordre. Donc connaissant le rang de l'élément dans l'ordre total, une simple correspondance nous permettra de trouver la "valeur" de cet élément. Quand nous parlerons du "rang" d'un élément, il s'agira toujours du rang dans l'ordre total. Autrement dit, nous nous sommes servi de l'ensemble N des entiers naturels pour indexer les éléments de la collection E_i.

Si nous posons maintenant l'existence de plusieurs collections E_i, telles que:

$$E_1,\ E_2,\ E_3,\ \ldots,\ E_m$$

portant sur un même caractère, mais différenciées, par exemple, soit par leur localisation dans le temps, soit par leur localisation dans l'espace, une collection quelconque E_i pourra toujours se mettre sous la forme:

$$(2) \qquad E_i = \left\{ e_{1_i},\ e_{2_i},\ e_{3_i},\ \ldots,\ e_{n_i} \right\}$$

Dans laquelle la suite 1, 2, 3, ...n, représente l'ordre total de la collection E_i. L'ensemble de l'information dont on dispose se présentera alors sous la forme du tableau suivant:

$$(3) \qquad \left(\begin{array}{l} E_1 = \left\{ e_{1_1},\ e_{2_1},\ e_{3_1},\ \ldots,\ e_{n_1}, \right\} \\ E_2 = \left\{ e_{1_2},\ e_{2_2},\ e_{3_2},\ \ldots,\ e_{n_2}, \right\} \\ E_3 = \left\{ e_{1_3},\ e_{2_3},\ e_{3_3},\ \ldots,\ e_{n_3}, \right\} \\ . \\ . \\ . \\ E_m = \left\{ e_{1_m},\ e_{2_m},\ e_{3_m},\ \ldots,\ e_{n_m}, \right\} \end{array} \right.$$

La difficulté (on se propose de comparer les E_i) consiste à maîtriser cette information et la difficulté est réelle quand les effectifs n_i sont grands et les collections E_i nombreuses. Mais le point essentiel est que nous avons m collections totalement ordonnées, et que connaissant le rang d'un élément dans une collection donnée nous pouvons en trouver immédiatement la valeur.

Quelques bases partielles de comparaison viennent immédiatement à l'esprit. Dans chaque collection totalement ordonnée on peut convenir de retenir le premier élément. On aurait alors une nouvelle collection formée par tous les éléments e_{1_i} du tableau ci-dessus. On pourrait aussi bien retenir tous les derniers éléments e_{n_i}.

Cette première sélection d'éléments, basée sur la place particulière que les éléments occupent dans chaque collection et qui est la même dans chacune, bien qu'elle soit très imparfaite puisque toute une information existe entre le premier et le dernier élément, dont nous ne tenons pas compte, nous suggère de retenir d'autres éléments particuliers ayant une place identique dans chaque collection. Nous pensons immédiatement à l'élément central, à la médiane, et à tous les autres quantiles usuels, quartiles, déciles, centiles.

Autrement dit, la structure ordinale conduit directement à la notion de quantiles qui apparaît comme une notion très simple. C'est à partir de ce moment là que les difficultés commencent et nous avons pu en déterminer deux de façon nette. Nous les exposerons maintenant.

II - LES DIFFICULTES RENCONTREES

Convenons de prendre comme exemple la détermination de la médiane. Il convient d'une part de faire la distinction entre l'opération qui permet de regrouper les éléments de la collection totalement ordonnée en deux sous-collections contenant chacune la moitié des éléments de la collection initiale et la médiane elle-même qui apparait comme la coupure qui réalise cette partition particulière.

Les meilleurs étudiants voient immédiatement les deux cas possibles déterminés par la parité de l'effectif n_i. On peut avoir:

(4) $\quad n_i = 2\,p$

ou

(5) $\quad n_i = 2\,p + 1$

Quand la relation (4) est vérifiée, il est bien possible de faire une partition de la collection initiale en deux sous-ensembles équipotents ayant chacun p éléments. On peut écrire:

$$(6)\quad \begin{cases} E'_i = \left\{ e_{1_i}, e_{2_i}, e_{3_i}, \ldots, e_{p_i} \right\} \\ \\ E''_i = \left\{ e_{(p+1)_i}, e_{(p+2)_i}, e_{(p+3)_i}, \ldots, e_{n_i} \right\} \end{cases} \quad \text{avec: } E'_i \cup E''_i = E_i$$

Et les étudiants sont embarrassés pour trouver la coupure!

Quand la relation (5) est vérifiée, il est difficile de faire admettre que l'élément central, cette fois bien évident, va, par moitié, dans chacun des deux sous-ensembles! La partition présentée est alors la suivante:

$$(7)\quad \begin{cases} E_i' = \{e_{1_i}, e_{2_i}, e_{3_i}, \ldots, e_{p_i}\} \\ E_i'' = \{e_{(p+1)_i}\} \\ E_i''' = \{e_{(p+2)_i}, e_{(p+3)_i}, e_{(p+4)_i}, \ldots, e_{n_i}\} \end{cases}$$

La partition réalisée est alors en contradiction avec la définition de l'opération qui conduit à la détermination de la médiane. On ne sait que faire de l'élément $e_{(p+1)_i}$ qui apparaît comme un élément en surnombre!

Par ailleurs, les meilleurs étudiants, ceux qui possèdent la notion de congruence, posent que les quantiles, en tant que coupures, ne peuvent exister dans les collections étudiées que si, et seulement si, les congruences suivantes sont vérifiées*:

(8) pour la médiane: $n_1 \equiv n_2 \equiv n_3 \equiv \cdots \equiv n_m = 1 \pmod{2}$

(9) pour les quartiles: $n_1 \equiv n_2 \equiv n_3 \equiv \cdots \equiv n_m = 3 \pmod{4}$

(10) pour les déciles: $n_1 \equiv n_2 \equiv n_3 \equiv \cdots \equiv n_m = 9 \pmod{10}$

(11) pour les centiles: $n_1 \equiv n_2 \equiv n_3 \equiv \cdots \equiv n_m = 99 \pmod{100}$.

En d'autres termes, pour les meilleurs étudiants, pour ceux qui se posent des questions, les termes de quantilage et de quantiles constituent un véritable paradoxe: quand les (q-1) coupures existent dans une collection, le partage en q sous-ensembles équipotents est impossible; quand le partage en q sous-ensembles équipotents est possible, les coupures disparaissent comme par enchantement.

Si l'on admet que ces réflexions ne manquent pas de pertinence, on est dans l'obligation soit de recourir à des conventions, soit d'imaginer une présentation qui supprime les difficultés. Les conventions risquent d'être oubliées et de conduire à des confusions. Nous avons opté pour la deuxième solution et nous craignons de n'avancer qu'une piètre banalité.

III - LA PRESENTATION ADOPTEE

On peut partir d'une base simple: la division euclidienne qui fait intervenir:

d'une part l'effectif n_i de la collection,

d'autre part, un nombre entier q choisi à l'avance et tel que la relation:

(12) $q < n_i$ soit vérifiée.

Toute collection E_i ayant une structure ordinale, une collection quelconque à étudier peut toujours être représentée par n_i points équidistants portés sur un

(*) $n \equiv p \pmod{q}$ signifiant que n et p diffèrent d'un multiple de q.

axe $\overrightarrow{OX}$. Sur cet axe, seul le segment $[1, n_i]$ est intéressant. De la figure (1), on passe à la figure (2). ($n_i = 9$).

Fig. 1

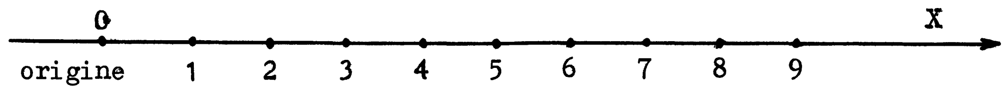


Fig. 2

Nous nous demandons alors quels sont les points de l'axe OX qui permettent de partager le segment $[1, n_i]$ en q segments égaux. Telle est la première étape qui permet de préciser $(q-1)$ positions sur l'axe OX. Ces positions étant précisées, il faudra leur trouver une "valeur".

Nous envisagerons successivement ces deux points.

1. La détermination de la position sur l'axe :

Les points cherchés sont immédiatement déterminés par leur abscisse mesurée sur l'axe. Quand nous voulons diviser le segment $[1, n_i]$ en q segments égaux, nous porterons $(q-1)$ points $t_1, t_2, t_3, \ldots, t_{q-1}$ dont les abscisses sont respectivement:

$$(13) \quad \left(\begin{array}{l} a_1 = \dfrac{1 + n_i}{q} \\ a_2 = 2 . \dfrac{1 + n_i}{q} \\ a_3 = 3 . \dfrac{1 + n_i}{q} \\ . \\ . \\ . \\ a_{k-1} = q-1 . \dfrac{1 + n_i}{q} \end{array} \right.$$

Nous avons ainsi un ensemble T_q de points qui constitue un ensemble de référence adapté à chaque collection. Par construction, l'ensemble des abscisses correspondantes vérifie la relation:

$$(14) \quad A_q \subset Q^+ \quad \text{(Ensemble des fractions positives)}$$

Les étudiants ne sont plus tentés de voir dans les quantiles des coupures qui existent nécessairement dans les éléments de la collection. Les coupures forment un ensemble bien distinct de toute collection. Ce n'est que dans les cas particuliers des congruences précédemment mentionnées qu'une relation d'inclusion plus restrictive que la précédente est vérifiée, à savoir:

(15) $A_q \subset N$ (N: ensemble des entiers

La catégorie ordinale reste bien à la base de cette construction, mais au lieu de prendre cette catégorie dans l'ensemble des entiers naturels, nous la prenons directement dans l'ensemble plus riche des rationnels positifs.

2. Détermination de la valeur :

Toute abscisse a_j, avec $j \in \{1, 2, 3, \ldots, q-1,\}$ est nécessairement ou confondue avec l'une des abscisses des n_i points représentatifs des éléments de E_i, ou comprise entre deux abscisses entières consécutives. Les deux cas sont à envisager.

a. L'abscisse a_j est confondue avec une abscisse entière:

La détermination de la valeur de l'élément construit est immédiate: c'est celle de l'élément de E_i qui a cette abscisse entière. Nous avons vu qu'une simple correspondance permettait de passer du rang, de l'abscisse entière, à la valeur il n'y a donc aucune difficulté dans ce cas.

b. L'abscisse a_j est comprise entre deux abscisses entières consécutives:

nous nous référons aux valeurs qui caractérisent ces deux éléments consécutifs de E_i. Deux cas sont possible:

- les deux éléments consécutifs portent la même valeur: il n'y a aucune raison de donner à l'élément construit une valeur distincte de cette valeur commune à deux éléments consécutifs de E_i. La détermination de la valeur est encore immédiate.
- les deux éléments consécutifs portent deux valeurs consécutives: dans ce cas, nous dirons que la valeur de l'élément construit est comprise entre les deux valeurs consécutives.

Supposons, par exemple, que le point médian, d'abscisse $\frac{1 + n_i}{2}$ soit entre la pième et le (p+1)ème élément de E_i, que le p ième élément porte la marque III, le p+1 ème la marque IV, nous dirons que la valeur de la médiane est comprise entre III et IV, bornes exclues.

Il reste à donner maintenant une définition des quantiles.

IV – DEFINITION DES "QUANTILES"

Etant donné une information statistique se présentant sous la forme d'une collection d'observations ayant les caractéristiques d'un ensemble fini, dénombrable et totalement ordonné de n éléments, on appelle "quantiles" les $(q - 1)$ éléments a_j d'un ensemble de référence défini et totalement ordonné par la relation:

$$(16) \qquad a_j = j \cdot \frac{1 + n}{q} \qquad \text{avec :} \quad \begin{array}{l} q, \text{ entier quelconque inférieur à } n. \\ j = 1, 2, 3, \ldots, q-1 \end{array}$$

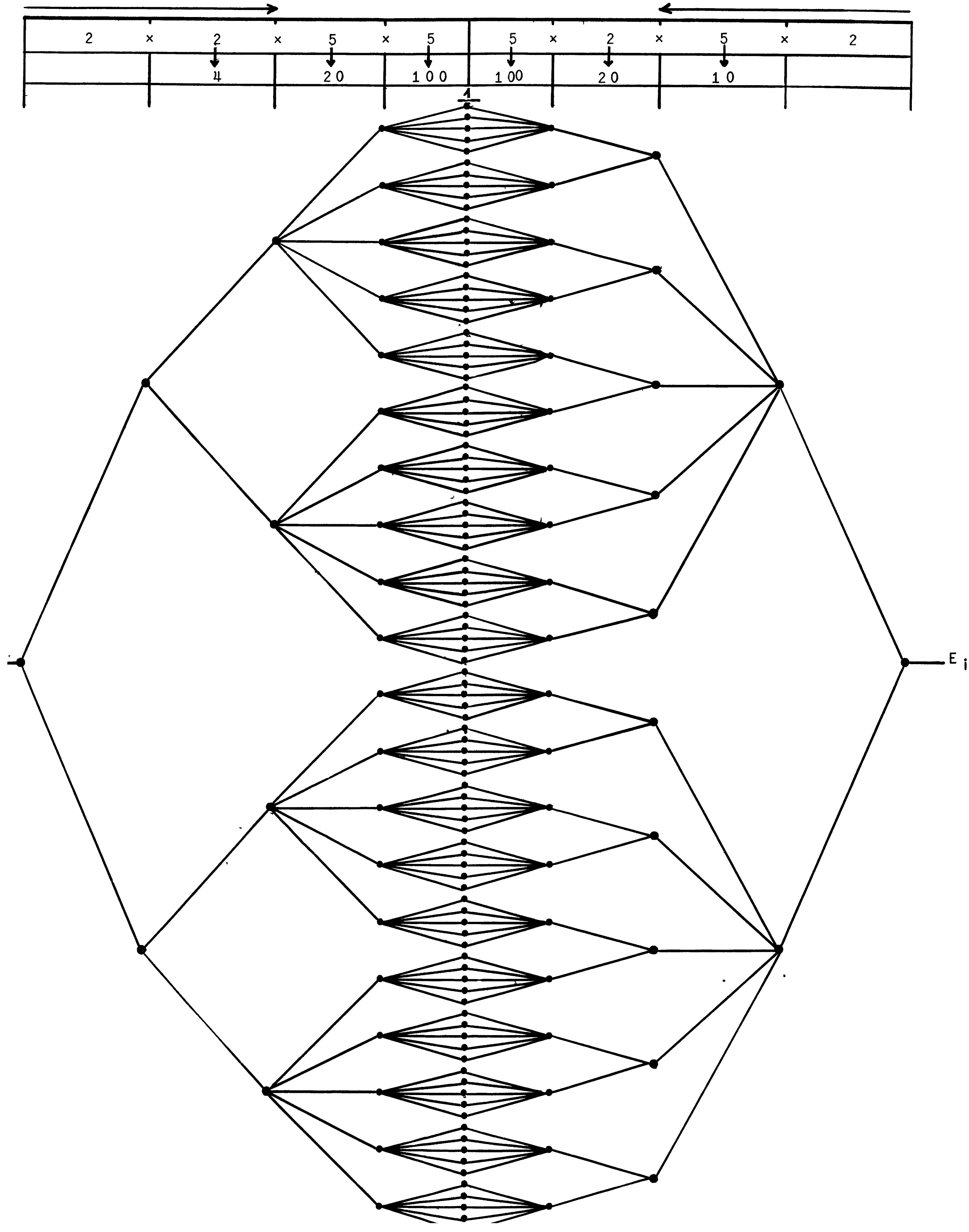
En donnant à q les valeurs 2, 4, 10, 100, on obtient les quantiles traditionnels. On peut remarquer que les quantiles traditionnels ne font intervenir que la division par deux et par cinq, et que, de plus, la médiane reste la base de toute construction.

En effet, en partant de la médiane, on obtient les quartiles en refaisant la même opération de partage sur chacun des deux segments définis par la médiane. Pour avoir les déciles $(q = 10 = 2 \times 5)$, il suffit de diviser en cinq segments égaux les deux segments définis par la médiane. Pour avoir les centiles $(q = 100 = 2 \times 5 \times 2 \times 5)$ on partira des déciles et, sur chacun des dix segments égaux, on refera les mêmes opérations de partage appliquées au segment initial $[1, n_i]$

On peut se demander si la tradition ne nous prive pas de l'usage de certains quantiles impairs qui permettraient, d'une part d'éliminer des comparaisons les extrémités qui contiennent, le plus souvent, des valeurs anormales, d'autre part d'avoir une classe centrale et pas seulement une valeur centrale. Les comparaisons entre les collections auraient une base plus étendue dans une partie de chaque collection qui est particulièrement représentative.

Nous avons essayé d'étudier la structure des "quantiles" traditionnels, il s'agit d'une structure arborescente (cf. fig. 3).

Figure N° 3: Deux façons d'arriver aux centiles en partant de la médiane et en passant soit par les quartiles, soit par les déciles.



G. Th. GUILBAUD

UNE DISTRIBUTION SINGULIERE

Kléber :	0,19
Etienne :	0,94
Bernard :	0,55
David :	0,39
Charles :	0,77
André :	0,06
Gaston :	0,64
Félix :	0,28
Ignace :	0,87
Jules :	0,47
Henri :	0,13
Louis :	0,70
Marcel :	0,33
Noël :	0,82

Si ces nombres avaient été obtenus comme résultats d'une suite d'observations statistiques - par exemple pourcentage de réussites à une épreuve répétée - il y aurait de quoi s'ébahir. Pourquoi ?

Prenez la peine de porter ces quatorze points sur le segment $(0 \leqslant x \leqslant 1)$. La répartition vous semblera grosso modo assez uniforme; mais elle l'est même beaucoup, en ce sens que, si nous divisons le segment en quatorze parties égales, il y a un point et un seul dans chaque partie. Mais ce n'est pas tout: si vous ôtez le dernier point (Noël) vous obtenez une série de treize qui a une régularité analogue (à savoir un point dans chacun des treize morceaux égaux). Et vous pouvez continuer: ôtez les deux derniers (Marcel et Noël) et les douze points restant sont encore également répartis (au sens précisé) - de même encore si vous ôtez les trois ou quatre ou cinq.... derniers. L'ordre de présentation joue un rôle; mais vous pouvez recopier la liste en suivant cette fois l'ordre alphabétique et vous pourrez faire les mêmes constatations. Reste à savoir comment obtenir de telles régularités. Vous pouvez vous y amuser. Mais d'après ce qu'on dit dans certains milieux polonais et généralement bien informés, il n'est pas possible de fabriquer de telles listes beaucoup plus longues que celles qui est présentée ici. Peut-être pourrez-vous obtenir une liste de dix-sept nombres: ce sera un record.

J. GARDELLE

A PROPOS DES "CADENCES"

Considérons un alphabet à deux symboles: a, b. On dira que le mot

a a b a a b b a (1)

comporte une cadence d'ordre 3, en ce sens qu'on y trouve un même symbole (a, en l'occürence) à trois places équidistantes, la 2ème, la 5ème et la 8ème:

a <u>a</u> b a <u>a</u> b b <u>a</u>

. a . . a . . a

On dira que la cadence est de "pas" 2, car il existe deux symboles entre les éléments successifs de la cadence.

Cet autre mot, de huit lettres également,

a b b a a b b a (2)

ne comporte aucune cadence d'ordre 3, mais on constatera sans peine qu'on ne saurait lui adjoindre une lettre sans faire apparaître une telle cadence.

Il n'existe pas, dans l'alphabet à deux symboles, de mot de neuf lettres, qui ne comporte aucune cadence d'ordre 3.

On cherchera les deux mots de huit lettres, autres que (2), qui ne comportent aucune cadence d'ordre 3.

Dans l'alphabet à trois symboles a, b, c, le mot

a b a bb c b a a c a c c b c b b a b a c c a a c c (3)

ne comporte aucune cadense d'ordre 3. On vérifiera qu'on ne peut adjoindre à ce mot aucune lettre, sans y faire apparaître au moins une cadence d'ordre 3, par exemple, adjoindre c fait apparaître deux cadences d'ordre 3, l'une de pas 0, l'autre de pas 5.

Enfin, le mot

a b c c b c b a a b b a a b c b c c b a a c c a c a (4)

comporte également 26 lettres, et est dépourvu de cadence d'ordre 3. Mais le mot (3) est doué d'une propriété qui nous paraît intéressante et que ne possède pas le mot (4). Laquelle ? (pour la découvrir, on amputera le mot (3) de ses lettres extrêmes a et c).

On s'en expliquera ultérieurement.

STAGES DU CENTRE DE MATHEMATIQUE SOCIALE

Du lundi 22 Mars au samedi 27 Mars aura lieu un stage de perfectionnement, destiné aux chercheurs en Sciences Humaines, sur les Processus.

Principaux thèmes traités:
- Algèbre linéaire et processus
- Processus de Markov
- Processus de Poisson
- Algorithmes
- Processus en génétique, en démographie et en économie
- Processus d'apprentissage.

A l'automne 1965 sera organisé un stage sur les Algèbres de Boole.

- - - - - - - - - -

CHANTIERS MATHEMATIQUES

Programme des émissions pour le 2ème trimestre

Lundi à 17 h. 55

11-1-65	Dimension d'un espace vectoriel
18-1-65	Application linéaire
25-1-65	Factorisation d'une application linéaire
1-2-65	Bords
8-2-65	Espace des applications linéaires
15-2-65	Anneau des endomorphismes d'un espace linéaire
22-2-65	Automorphismes
1-3-65	Matrices
8-3-65	Le linéaire en physique 1
15-3-65	Valeurs propres
22-3-65	Le linéaire en physique 2
29-3-65	Un exercice de calcul matriciel
5-4-65	Espace affine

Vendredi à 17 h.55

8-1-65	Apprentissage des structures
15-1-65	Groupes en géométrie
22-1-65	Algèbre et géométrie
29-1-65	Congruences et Idéaux
5-2-65	Idéaux et p.g.c.d.
12-2-65	Morphismes 1
19-2-65	Morphismes 2
5-3-65	Génération de structures
12-3-65	Ordre et types d'ordre
19-3-65	L'ordre du continu
26-3-65	Mathématiques sagittales
2-4-65	Naissance des notations modernes

Des documents d'accompagnement, par fascicules trimestriels, sont en vente à l'Institut Pédagogique National, 29, rue d'Ulm, Paris 5ème.