A propos du choix du partenaire en aikido : éléments d'algèbre et de combinatoire des appariements d'un ensemble fini

*About the choice of the partner in aikido: elements of algebra and combinatorics
of the pairing of a finite set*

Marc Barbut et Jean-Marie Duprez

Édition électronique

URL : <http://journals.openedition.org/msh/4083>

DOI : 10.4000/msh.4083

ISSN : 1950-6821

Éditeur

Centre d'analyse et de mathématique sociales de l'EHESS

Édition imprimée

Date de publication : 7 mai 2007

Pagination : 5-51

ISSN : 0987-6936

Référence électronique

Marc Barbut et Jean-Marie Duprez, « A propos du choix du partenaire en aikido : éléments d'algèbre et de combinatoire des appariements d'un ensemble fini », *Mathématiques et sciences humaines* [En ligne], 177 | Printemps 2007, mis en ligne le 16 mai 2007, consulté le 23 juillet 2020. URL : <http://journals.openedition.org/msh/4083> ; DOI : <https://doi.org/10.4000/msh.4083>

À PROPOS DU CHOIX DU PARTENAIRE EN AÏKIDO : ÉLÉMENTS D'ALGÈBRE ET DE COMBINATOIRE DES APPARIEMENTS D'UN ENSEMBLE FINI

Marc BARBUT¹, Jean-Marie DUPREZ²

RÉSUMÉ – Les règles de l'aïkido fournissent un cas particulier du problème général suivant : un ensemble fini E ($|E| = 2n$) d'individus est partitionné en k classes données (partition Π) ; chacun de ces individus doit choisir librement un partenaire parmi les autres ; on obtient ainsi une seconde partition de E en n paires ; c'est un appariement. La suite des effectifs des classes dans le croisement de Π et d'un appariement est appelé ici configuration.

L'article étudie l'ensemble des configurations possibles dans les deux cas les plus simples ($k = 2$ et $k = 3$). Pour le cas général ($k \geq 4$), quelques propriétés générales sont démontrées.

MOTS-CLÉS – Aïkido, Appariement, Configuration, Homogénéité, Probabilité, Treillis

SUMMARY – About the choice of the partner in aikido: elements of algebra and combinatorics of the pairing of a finite set

The rules of the game named aikido provide a particular occurrence of the following general problem: a finite set E ($|E| = 2n$) of persons is divided into k given classes (partition Π) ; each person is allowed to choose freely a partner among the others; the result is a second partition of E , into n 2-subsets; we name it pairing. The sequence of the sizes of the classes of the cut (infimum) of Π and a pairing (in the partitions lattice of E) is called here a configuration.

This paper studies the set of all possible configurations in the two elementary cases $k = 2$ and $k = 3$. For $k \geq 4$, some general properties are showed.

KEYWORDS – Aikido, Configuration, Homogeneity, Lattice, Pairing, Probability

1. INTRODUCTION

1.1. LES RÈGLES DE L'AÏKIDO

Ce travail est issu d'une interrogation sur la logique de la formation des paires dans un cadre sportif où les participants sont d'abord partenaires avant d'être adversaires[□] l'aïkido. Dans cet art martial de défense d'origine japonaise caractérisé par l'absence de compétition, chaque entraînement est constitué de plusieurs séquences pédagogiques ritualisées. Après une présentation par un enseignant de la technique qui doit être

---

¹ Centre d'analyse et de mathématiques sociales (CAMS), École des hautes études en sciences sociales (EHESS), 54 bd Raspail 75270 Paris cedex 06, mbarbut@ehess.fr

² Centre lillois d'études et de recherches sociologiques et économiques (CLERSE), Université des sciences et technologies de Lille, 59655 Villeneuve d'Ascq, jean-marie.duprez@univ.lille1.fr

étudiée, les participants se répartissent en paires pour reproduire au mieux ce qui a été présenté.

Dans le choix d'un partenaire, deux règles non écrites s'appliquent de fait□

- pour chaque séquence, chaque participant doit s'apparier à un autre qui sera son partenaire pendant toute la séquence□
- au début de chaque nouvelle séquence, chacun des participants doit changer de partenaire.

Si la première règle s'impose à tous comme une condition nécessaire, la seconde a un caractère moins impératif de sorte que tout se passe comme si, à chaque nouvelle séquence, chaque participant pouvait choisir n'importe quel partenaire, y compris celui avec lequel il vient de pratiquer.

Aussi, dans le texte qui suit, c'est, sauf au § 2, un aikido ne supposant pas la seconde règle que l'on étudiera.

1.2. LE PROBLÈME SOCIOLOGIQUE

Comme dans toute situation sociale, l'ensemble des participants peuvent toujours être répartis entre plusieurs catégories sociologiquement significatives□ en aikido, ce sera selon le sexe, l'âge, le grade, l'origine géographique, les relations préalables, etc.

Pour le sociologue la première question qui se pose est pratique□ il s'agit de savoir si les appariements observés peuvent être considérés comme réalisés «□ au hasard□ » ou en fonction de contraintes qu'il restera à expliciter. Existe-t-il une tendance à la formation de paires homogènes (on pratique de préférence avec un partenaire de même catégorie que soi) ou au contraire à la mixité□ (on pratique de préférence avec un partenaire de catégorie différente) ?

La réponse à la question précédente permet d'en aborder une seconde, théorique cette fois, qui porte sur les logiques sociales à la base des choix. Sont-elles réductibles à la recherche d'identités partagées ou relèvent-elles d'abord de contraintes relationnelles et structurelles□

Pour répondre à ces questions, une manière de procéder consiste à chercher à ordonner toutes les manières possibles de s'apparier des façons les plus homogènes aux plus hétérogènes, les appariements «□ au hasard□ » se situant dans une zone médiane. Il serait alors possible de situer les appariements effectivement observés par rapport à cette distribution des appariements possibles.

D'où l'intérêt d'une analyse combinatoire et probabiliste des possibilités d'appariements, qui permette ultérieurement une analyse statistique des observations empiriques.

1.3 LES PROBLÈMES MATHÉMATIQUES

Cette analyse de la combinatoire du jeu d'aikido pose au mathématicien bien des problèmes intéressants, dont seuls les plus simples sont résolus dans le texte de cet article.

Pour le reste, nous nous bornerons, dans le huitième et dernier paragraphe, à récapituler les conjectures et les problèmes ouverts (à notre connaissance) rencontrés au cours de ce travail.

Il est d'ailleurs possible que certains de ces problèmes soient résolus depuis longtemps. En effet, les questions posées à propos de l'aïkido nous paraissent largement dépasser le cadre étroit de ce jeu particulier. Et il nous semblerait invraisemblable que les mêmes problèmes d'appariements n'aient pas déjà été rencontrés dans l'industrie (problèmes de couplages de pièces), en recherche opérationnelle, ou encore par des auteurs de «*Créations mathématiques*», voire même, en sociologie, par les spécialistes des «*Réseaux sociaux*».

Nous sommes nous-mêmes très ignorants en ces matières. Mais quelques éminents combinatoristes auxquels nous en avons parlé nous ont affirmé n'avoir pas connaissance des problèmes que nous abordons ici à propos de l'aïkido.

Peut-être des lumières, partielles ou totales, viendront-elles de certains de nos lecteurs.

N.B. Les données empiriques sur lesquelles J.-M. Duprez a travaillé, et les traitements qui leur ont été appliqués font l'objet d'un article dans la *Revue française de sociologie*, 2007-2.

2. APPARIEMENTS

2.1. ENSEMBLE DES APPARIEMENTS POSSIBLES

Soit E un ensemble ayant un nombre pair $|E| = 2n$ d'éléments.

Un appariement de E est une partition de E en n classes de 2 éléments, ou paires.

Désignons par A_{2n} , ou encore a_n , le nombre des appariements distincts possibles pour E .

Un élément x donné de E peut s'apparier avec n'importe lequel des $(2n - 1)$ autres éléments, soit y . La paire $\{x, y\}$ étant constituée, il restera à appairer les $2n - 2 = 2(n - 1)$ éléments restants de E . D'où

$$a_n = A_{2n} = (2n - 1)A_{2n-2} = (2n - 1)a_{n-1} = (2n - 1)(2n - 3)a_{n-2}$$

Comme $a_1 = A_2 = 1$, il vient finalement

$$(2.1) \quad a_n = A_{2n} = 1.3.5... (2n - 3)(2n - 1) = \frac{2n!}{n! 2^n}$$

Lorsque l'ensemble E a un nombre impair $(2n - 1)$ d'éléments, l'un de ceux-ci ne peut s'apparier si l'on ajoute à E un élément fictif x_0 avec lequel l'élément qui reste seul est censé s'apparier, on est ramené au cas précédent³. D'où

$$(2.1)' \quad A_{2n-1} = A_{2n} = a_n$$

Remarque. En utilisant la formule de Stirling, on voit que pour n grand

$$a_n \sim \frac{2n!}{n! 2^n} \sim \sqrt{2} \quad \text{et} \quad \frac{a_n}{n!} \sim \frac{2^n}{\sqrt{\pi n}}$$

³ En pratique, dans l'aïkido, aucun joueur ne fait tapisserie : le joueur surnuméraire s'agresse à l'une des paires, qui devient ainsi un trio, ou au maître de jeu (l'enseignant).

La croissance de a_n est donc bien plus rapide que celle de n

N.B. Par convention, on posera $a_0 = A_0 = 1$.

2.2. APPARIEMENTS CONDITIONNELS

Selon la seconde règle de l'aïkido, à chaque changement de séquence, chaque joueur doit (en principe) changer de partenaire. (Il semble qu'en pratique cette règle soit peu respectée).

Appelons *appariements conditionnels* de E ceux qui peuvent suivre un appariement donné, lorsque la règle est respectée.

Désignons par $B_{2n} = b_n$ le nombre de ces appariements conditionnels éventuellement consécutifs à un appariement donné.

On a évidemment $B_2 = b_1 = 0$ $B_4 = b_2 = 2$

Pour $n \geq 3$ les nombres b_n satisfont à l'équation de récurrence

$$(2.2) \quad b_n = 2(n-1)(b_{n-1} + b_{n-2})$$

DÉMONSTRATION de (2.2). Soit x un élément de E , qui était apparié avec y . Il doit choisir son nouveau partenaire parmi les $2n-2 = 2(n-1)$ éléments différents de y et de lui-même.

Supposons qu'il s'apparie avec z , qui lui-même était apparié à t .

Alors, ou bien y s'apparie avec t . Il restera à appairer, en respectant la règle, les $2n-4 = 2(n-2)$ éléments autres que x, y, z et t il y a b_{n-2} tels appariements.

Ou bien y s'apparie avec un élément autre que t il s'agit alors d'un appariement respectant la règle parmi les $2n-2 = 2(n-1)$ éléments autres que x et z il y a b_{n-1} tels appariements. D'où finalement (2.2)

$$b_n = 2(n-1)(b_{n-1} + b_{n-2})$$

N.B. Par convention, on pose $b_0 = B_0 = 1$.

2.3. PROPORTION DES APPARIEMENTS CONDITIONNELS DANS L'ENSEMBLE DES APPARIEMENTS POSSIBLES

Posons $r_n = \frac{B_{2n}}{A_{2n}} = \frac{b_n}{a_n}$ on a $r_1 = 0$ et $r_2 = \frac{2}{3}$

De (2.2) on déduit immédiatement, compte tenu de (2.1)

$$r_n = \frac{b_n}{a_n} = 2(n-1) \frac{b_{n-1}}{(2n-1)a_{n-1}} + \frac{b_{n-2}}{(2n-1)(2n-3)a_{n-2}}$$

Soit, pour $n > 2$, l'équation de récurrence

$$(2.3) \quad r_n = \frac{2(n-1)}{2n-1} r_{n-1} + \frac{r_{n-2}}{2n-3}$$

On va montrer que r_n est une suite monotone croissante comme cette suite est évidemment bornée supérieurement par 1, il en résulte que r_n a une limite r lorsque n augmente indéfiniment.

DÉMONSTRATION. Posons $u_n = \frac{r_n}{r_{n-1}}$.

De (2.3) on déduit l'équation de récurrence

$$(2.4) \quad \frac{r_n}{r_{n-1}} = u_n = \frac{2(n-1)}{2n-1} + \frac{1}{(2n-3)u_{n-1}}$$

Montrons que u_n satisfait, pour $n \geq 3$, à la double inégalité

$$(2.5) \quad 1 < u_n < 1 + \frac{1}{(2n-1)(2n-3)}$$

c'est-à-dire

$$(2.5)' \quad r_{n-1} < r_n < r_{n-1} + \frac{1}{(2n-1)(2n-3)}$$

Ce qui, à la fois, prouvera la monotonie de r_n et donnera une idée de la rapidité de la convergence vers la limite r .

Montrons que

$$(2.5)_1 \quad u_{m-1} > 1 \quad (2.5)_2 \quad u_m < 1 + \frac{1}{(2m-1)(2m-3)}$$

En effet, d'après (2.4), si $u_{m-1} \geq 1$, on a

$$u_m < \frac{2(m-1)}{2m-1} + \frac{1}{2m-3} = \frac{4(m-1)^2}{(2m-1)(2m-3)} = \frac{(2m-1)(2m-3)+1}{(2m-1)(2m-3)}$$

Réciproquement, montrons que

$$(2.5)_2 \quad u_m < 1 + \frac{1}{(2m-1)(2m-3)} \quad (2.5)_1 \quad u_{m+1} > 1$$

En effet, d'après (2.4), l'hypothèse (2.5)₂ implique

$$u_{m+1} > \frac{2m}{2m+1} + \frac{1}{(2m-1)(1 + \frac{1}{(2m-1)(2m-3)})} = v_m$$

Après réductions de la fraction rationnelle v_m , il vient, pour $m \geq 3$

$$v_m = 1 + \frac{m-2}{(2m+1)(m-1)^2} > 1$$

2.4. UN ENCADREMENT POUR LA LIMITE r DE LA PROPORTION r_n DES APPARIEMENTS CONDITIONNELS

Considérons la suite r'_n définie par

$$(2.6) \quad r'_n = r_n + \frac{1}{3n-1} = r_n \frac{3n}{3n-1}$$

On a évidemment, pour tout $n \geq 1$: $r'_n > r_n$.

Montrons que la suite r'_n est monotone décroissante pour $n \geq 4$. On a

$$\frac{r'_n}{r'_{n-1}} = u'_n = \frac{r_n}{r_{n-1}} \frac{1 + \frac{1}{3n-1}}{1 + \frac{1}{3n-4}} = u_n \frac{n(2n-4)}{(n-1)(3n-1)}$$

En majorant u_n selon (2.5)₂, on obtient

$$u'_n < 1 + \frac{1}{(2n-1)(2n-3)} \frac{n(3n-4)}{(n-1)(3n-1)}$$

Après réductions et simplifications, on a

$$u'_n < 1 + \frac{n-3}{4n(n-1)(n-\frac{4}{3})} < 1 \quad \text{si } n > 3$$

On a donc pour la limite commune r de la suite monotone croissante r_n et de la suite monotone décroissante r'_n l'encadrement

$$(2.7) \quad \forall n \geq 4 \quad r_n < r < r'_n = \frac{3n}{3n-1} r_n$$

En itérant (2.7) à partir de $n = 4$, on voit que r se situe aux alentours de 0,6⁴ pour $n = 24$, on a ainsi $0,6 < r < 0,61^4$.

CONCLUSION. Dès que le nombre de joueurs est de l'ordre de la cinquantaine ou plus (ce qui est un ordre de grandeur réaliste dans le cas de l'aïkido), les appariements conditionnels autorisés par la règle constituent 60 % environ de l'ensemble de tous les appariements.

⁴ Pour $n = 11$, la «fourchette» fournie par (2.7) est déjà l'intervalle $(0,592 < r < 0,611)$.

2.5. COMPORTEMENT ASYMPTOTIQUE DES APPARIEMENTS CONDITIONNELS

2.5.1. Une chaîne de Markov

Dans l'ensemble des $A_{2n}(=a_n)$ appariements, chaque appariement $\square$ peut être suivi de ses $B_{2n}(=b_n)$ appariements conditionnels. Si nous considérons toutes ces transitions comme équiprobables, chacune a une probabilité $1/b_n$ strictement positive de se réaliser il y a par contre une probabilité nulle de passer d'un appariement donné $\square$ à l'un des $A_{2n} \cap B_{2n}(=a_n \cap b_n)$ appariements ayant au moins une paire commune avec lui.

Nous pouvons ainsi considérer l'ensemble des a_n appariements comme l'ensemble des états d'une chaîne simple et homogène de Markov.

Cette chaîne est *régulière*. Cette expression signifie que, asymptotiquement, tout se passe comme s'il n'y avait pas de seconde règle de l'aïkido la totalité des $(a_n \cap 1)$ appariements autres que $\square$ peuvent lui être consécutifs à chaque changement de séquence.

Pour le démontrer, on s'appuie sur ce théorème classique de la théorie des chaînes simples et homogènes de Markov à un nombre fini d'états selon lequel si une puissance de la matrice de transition a tous ses termes strictement positifs, alors la chaîne est régulière.

Dans notre cas particulier, c'est le carré de la matrice qui respecte cette condition suffisante on peut passer d'un appariement quelconque $\square$ à n'importe quel autre $\square'$ en deux changements de partenaires au plus. La démonstration se fait en quatre points

1. On peut passer de $\square$ à $\square$ lui-même en 2 changements de partenaires dans le premier, on passe de $\square$ à l'un quelconque, soit $\square'$, des b_n appariements qui peuvent suivre $\square$ en respectant la 2^e règle (pas de paire commune) puis retourner à $\square$ par le changement inverse chacun reprend son partenaire primitif.
2. Pour $n \geq 4$, il existe au moins un appariement $\square''$ sans paire commune ni avec $\square$, ni avec $\square'$.

En effet, parmi les $(a_n \cap 2)$ appariements distincts de $\square$ et de $\square'$, il y en a $(a_n \cap b_n)$ qui ont au moins une paire commune avec $\square$, et autant qui ont au moins une paire commune avec $\square'$. On va montrer que

$$(2.8) \quad 2(a_n \cap b_n) < a_n \cap 2$$

Il y aura donc $(a_n \cap 2) - 2(a_n \cap b_n) > 0$ appariements sans paire commune ni avec $\square$, ni avec $\square'$ on choisira l'intermédiaire $\square''$ parmi ceux-ci.

Montrons (2.8). Cette relation équivaut à $a_n + 2 < 2b_n$

Soit encore

$$(2.9) \quad r_n = \frac{b_n}{a_n} > \frac{1}{2} + \frac{1}{a_n}$$

La suite des a_n est croissante (de façon plus que factorielle) et celle des r_n , on l'a vu (cf. § 2.3 *supra*) est monotone croissante. D'où

$$n \geq 4 \Rightarrow r_n \geq r_4 = \frac{b_4}{a_4} = \frac{60}{105} = \frac{1}{2} + \frac{15}{210} > \frac{1}{2} + \frac{1}{105} = \frac{1}{2} + \frac{1}{a_4} \geq \frac{1}{2} + \frac{1}{a_n}$$

Ce qui est (2.9).

3. Pour $n = 2$, E a quatre éléments, x, y, z et t . Les trois appariements possibles sont

$$xy / zt \quad xz / yt \quad xt / yz$$

Chacun d'entre eux n'a aucune paire commune avec les deux autres, et peut donc servir d'intermédiaire entre ces deux autres.

4. Pour $n = 3$, E a six éléments x, y, z, t, u, v . Il y a 15 appariements. Soient $\square$ et $\square'$ deux, distincts, parmi ces 15. $\square$ et $\square'$ ont au plus une paire en commun.

S'ils ont une paire commune, la paire $\{u, v\}$ par exemple, c'est que $\square$ et $\square'$ sont deux des 3 appariements

$$xy / zt / uv \quad xz / yt / uv \quad xt / yz / uv$$

Si, par exemple, $\square$ est le premier et $\square'$ le second de ces 3 appariements, alors l'appariement $\square''$

$$xt / yu / zv$$

n'a de paire commune ni avec $\square$, ni avec $\square'$.

S'ils n'ont aucune paire commune, et que $\square$ soit $xy / zt / uv$ les 8 appariements sans paire commune avec $\square$ sont, à une permutation près des lettres, de la forme $\square'$

$$xz / yu / tv$$

On pourra alors prendre pour intermédiaire entre $\square$ et $\square'$ l'appariement $\square''$

$$xt / yv / zu$$

Remarque. Comme le nombre de paires communes à deux appariements est l'un quelconque des nombres entiers $0, 1, 2, \dots, n$ et compte tenu de la convention $b_0 = 1$, on a

$$(2.10) \quad a_n = \sum_{i=0}^n \begin{bmatrix} n \\ i \end{bmatrix} b_i$$

3. TYPES DE PAIRES ET CONFIGURATIONS

3.1. TYPES DE PAIRES

Supposons maintenant l'ensemble E partitionné en k classes $E_1, E_2, \dots, E_i, \dots, E_k$, d'effectifs

$$m_i = |E_i| \quad i = 1, 2, \dots, k$$

On a ainsi $|E| = m = \sum_{i=1}^k m_i$

Dans les applications, et notamment dans le cas de l'aïkido, ces classes peuvent correspondre au sexe ($k = 2$), au grade ($k = 3$), aux classes d'âge ($k = 4$ ou plus), etc...

Les $\frac{m(m-1)}{2}$ paires possibles d'éléments de E se répartissent en plusieurs *types*

- les paires de type *homogène*, constituées de deux éléments d'une même classe E_i ; il y a ainsi k types homogènes T_i , comprenant chacun $\frac{m_i(m_i-1)}{2}$ paires
- Les paires de type *mixte* (ou *hétérogène*), constituées d'éléments appartenant à deux classes E_i et E_j distinctes ; il y a ainsi $\frac{k(k-1)}{2}$ types mixtes T_{ij} , comprenant chacun $m_i m_j$ éléments.

On vérifie immédiatement que

$$(3.1) \quad \sum_{i=1}^k \sum_{i < j} m_i m_j + \sum_{i=1}^k \frac{m_i(m_i-1)}{2} = \frac{m(m-1)}{2}$$

3.2. PARITÉ DE m

Dans la suite, on supposera toujours que le nombre m d'éléments de E est pair

$$m = 2n$$

En effet, si m est impair

$$m = 2n - 1$$

on pourra, comme dans le paragraphe 2 *supra* adjoindre aux k classes $E_1, \dots, E_k$ une classe E_0 ayant pour seul élément l'élément fictif auquel, dans un appariement, l'élément de E qui reste seul sera censé s'apparier.

Les types *mixtes* T_{0i} sont au nombre de k , d'effectifs m_i respectivement ; on peut, si l'on veut, les appeler types *singuliers*⁵.

Compte tenu de $0! = 1$ et en posant par convention, dans les notations du § 2

$$(3.2) \quad A_0 = a_0 = 1$$

toutes les expressions figurant dans le présent paragraphe et notamment l'équation (3.1) ci-dessus et (3.4) ci-dessous, restent valides, appliquées à l'ensemble

$$E' = E \cup E_0$$

ayant le nombre pair $m' = m + 1 = 2n$ d'éléments.

⁵ Pour m impair, une autre solution est de constituer un trio. Nous n'étudierons pas ici cette variante.

3.3. CONFIGURATIONS

Soit $\square$ un appariement des éléments de E , avec, dans les mêmes notations qu'au § 3.1 ci-dessus

$$|E| = m = 2n = m_1 + m_2 + \dots + m_k = \sum_{i=1}^k m_i$$

Les n paires dont $\square$ est constitué se répartissent en

- t_i paires de type homogène T_i , pour $i = 1, 2, \dots, k$
- $t_{ij} (= t_{ji})$ paires de type mixte T_{ij} , pour $i \neq j$

On appelle *configuration* de l'appariement $\square$ la suite $t(\square)$ (de $k + \frac{k(k-1)}{2} = \frac{k(k+1)}{2}$ termes) :

$$t(\square) = (t_1, \dots, t_i, \dots, t_k, t_{12}, t_{13}, \dots, t_{ij}, \dots, t_{(k-1)k})$$

Comme les m_i éléments de la classe E_i appartiennent soit au type homogène correspondant, soit à l'un des types mixtes, on a évidemment

$$(3.3) \quad \forall i, \quad m_i = 2t_i + \sum_{j \neq i} t_{ij}$$

Ainsi, une suite t est une configuration si et seulement si elle vérifie le système (S_k) de k équations linéaires en nombres entiers ≥ 0

$$(S_k) \quad \begin{cases} m_1 = 2t_1 + \sum_{j=2}^k t_{1j} = 2t_1 + \sum_{j \neq 1} t_{1j} \\ m_i = 2t_i + \sum_{j \neq i} t_{ij} \\ m_k = 2t_k + \sum_{j \neq k} t_{kj} \end{cases}$$

Ce système est à $\frac{k(k-1)}{2}$ degrés de liberté c'est en effet la donnée des $\frac{k(k-1)}{2}$ nombres t_{ij} qui détermine les k nombres t_i .

Les nombres t_{ij} sont bien entendu soumis à la contrainte que, pour chaque i , l'entier positif ou nul $m_i - \sum_{j \neq i} t_{ij}$ doit être pair

$$(3.4) \quad \forall i, \quad m_i \geq \sum_{j \neq i} t_{ij} \quad \text{et} \quad \sum_{j \neq i} t_{ij} \equiv m_i \pmod{2}$$

En particulier

$$\forall (i, j) \quad t_{ij} \leq \min(m_i, m_j)$$

Pour énumérer toutes les configurations on peut considérer que chacune des suites de t_{ij} est un «hot» de longueur $\frac{k(k-1)}{2}$ écrit avec l'alphabet $\{0, 1, 2, \dots, \max_i (\min(m_i, m_j))\}$ et on écrira tous les «hots» satisfaisant au système (S_k) dans l'ordre lexicographique.

Par exemple, pour

$$m = 12 = 2 + n = 2 + 6 = m_1 + m_2 + m_3 + m_4 = 5 + 3 + 2 + 2$$

l'alphabet est $\{0 < 1 < 2 < 3\}$, et l'on a $t_{12} = 3$, les 5 autres t_{ij} étant au plus égaux à 2. Il y a 31 configurations en voici quelques-unes, dont la première et la dernière dans l'ordre lexicographique des «hots» ($t_{12}, t_{13}, t_{14}, t_{23}, t_{24}, t_{34}$)

t_{12}	t_{13}	t_{14}	t_{23}	t_{24}	t_{34}	t_1	t_2	t_3	t_4	P_r
0	0	1	0	1	0	2	1	1	0	$\frac{2}{231} \approx 0,86 \%$
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
1	1	1	0	0	1	1	1	0	0	$\frac{16}{231} \approx 6,92 \%$
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
2	1	0	1	0	0	1	0	0	1	$\frac{8}{231} \approx 3,46 \%$
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
3	2	0	0	0	0	0	0	0	1	$\frac{8}{3 \times 231} \approx 1,15 \%$

Tableau 3.1

Cette méthode d'énumération est très fastidieuse quand on l'applique «à la main». Elle a été programmée par Jean-Marie Duprez pour un traitement informatique, qui fournit rapidement la liste exhaustive.

Le nombre de configurations devient vite élevé. Par exemple, pour

$$m = 56 = 19 + 14 + 13 + 10$$

on obtient 25.143 configurations.

Nous ne connaissons pas de formule générale de dénombrement fournissant le nombre $C_{m,k}(m_1, m_2, \dots, m_k)$ de configurations correspondant à un même partage de m en une somme de k entiers pour $k \geq 4$. Un algorithme concernant le cas trichotomique ($k = 3$) est donné au § 5.3 infra.

Remarques.

1. Pour $m = 2n$, le nombre total de paires étant n , on a

$$(3.5) \quad n = \sum_{i=1}^k t_i + \sum_{i=1}^k \sum_{i < j} t_{ij}$$

2. Dans la liste par ordre lexicographique des configurations, les premières correspondent à une faible mixité, et les dernières à une forte mixité. Mais la première de la liste n'est pas nécessairement la plus homogène (la moins mixte), ni la dernière la plus mixte (la moins homogène).

Dans l'exemple numérique ci-dessus, on pourrait considérer que la moins mixte (la plus homogène) est la 9^e dans l'ordre alphabétique, à savoir

$$(1, 0, 0, 0, 0, 0 \mid 2, 1, 1, 1) \quad \text{qui n'a qu'une paire mixte}$$

De même, les plus mixtes (les moins homogènes) pourraient être la 25^e, la 26^e et la 30^e (sur 31), à savoir

$$(2, 1, 2, 1, 0, 0 \mid 0, 0, 0, 0)$$

$$(2, 2, 1, 0, 1, 0 \mid 0, 0, 0, 0)$$

$$(3, 1, 1, 0, 0, 1 \mid 0, 0, 0, 0)$$

Pour chacune, toutes les paires sont mixtes.

Les paragraphes 5, 6, et 7 seront pour l'essentiel consacrés à cette question d'un ordre «homogénéité croissante».

3. Le dernier exemple illustre le fait que, pour $k \geq 4$, à une même répartition $(t_1, t_2, t_3, \dots, t_k)$ des paires homogènes correspondent en général plusieurs répartitions des paires mixtes.
4. Lorsque $m = 2n - 1$ est impair, le système (S_k) comporte la $(k + 1)$ ^{ème} équation

$$(3.6) \quad |E_0| = m_0 = 1 = 0 + \sum_{i=1}^k t_{0i}$$

Cette équation signifie que $t_{0i} = 1$ si et seulement si l'élément non apparié de E appartient à la classe E_i .

Exemple. $k = 2, m = 2 \mid 4 \mid 1 = 4 + 3$

Le système (S_3) s'écrit

$$(s_3) \quad \begin{cases} 4 = 2t_1 + t_{01} + t_{12} \\ 3 = 2t_2 + t_{02} + t_{12} \\ 1 = t_{01} + t_{02} \end{cases}$$

L'énumération des configurations, ici immédiate, donne

t_{01}	t_{02}	t_{12}	t_1	t_2
0	1	0	2	1
0	1	2	1	0
1	0	1	1	1
1	0	3	0	0

Tableau 3.2

3.4. PROBABILITÉS DES CONFIGURATIONS

Soit $N(t_{12}, \dots, t_{ij}, \dots, t_1, \dots, t_i, \dots, t_k; m_1, m_2, \dots, m_i, \dots, m_k) = N(\mathbf{t}; m_1, \dots, m_k)$ le nombre d'appariements ayant la même configuration donnée $\mathbf{t}$, dans l'ensemble des $A_m = A_{2n} = a_n$ (dans les notations du § 2) appariements de E , avec $|E| = m = 2n$.

Ce nombre peut se calculer en trois étapes.

1. Pour chaque i , choisir dans la classe E_i les $(k-1)$ parties disjointes deux à deux d'effectifs t_{ij} pour $i \neq j$, et la partie résiduelle d'effectif $2t_i$. Le nombre de façons de le faire est (formule *multinomiale*)

$$\frac{m_i!}{t_{1i}! \dots t_{ji}! \dots t_{ki}! (2t_i)!} = \frac{m_i!}{t_{ij}! (2t_i)!}$$

Donc, pour E tout entier, un nombre de choix égal à

$$\frac{\prod_{i=1}^k m_i!}{\prod_{i=1}^k t_{ij}! (2t_i)!} = \frac{\prod_{i=1}^k m_i!}{\prod_{i=1}^k t_{ij}! (2t_i)!}$$

2. Apparier entre eux, pour chaque i , les $2t_i$ éléments de E_i constituant des paires de type homogène. Dans les notations du § 2 *supra*, le nombre de ces appariements est

$$A_{2t_1} A_{2t_2} \dots A_{2t_k} = a_{t_1} a_{t_2} \dots a_{t_k} = \prod_{i=1}^k a_{t_i}$$

3. Apparier entre eux, pour chaque couple (i, j) , avec $i \neq j$, les t_{ij} éléments de E_i et les t_{ij} éléments de E_j constituant les t_{ij} paires de type mixte T_{ij} , on a $t_{ij}!$ tels appariements. Donc, pour E tout entier

$$\prod_{i < j} t_{ij}!$$

Comme, dans la première étape, chaque $t_{ij}!$ apparaît deux fois (une fois sous m_i et une fois sous m_j), on a, après simplification

$$(3.7) \quad N(\mathbf{t}; m_1, \dots, m_k) = \frac{\prod_{i=1}^k m_i!}{\prod_{i=1}^k t_{ij}! (2t_i)!} \prod_{i=1}^k a_{t_i}$$

Compte tenu de l'expression (2.1) des a_{t_i} , N s'écrit encore

$$(3.7)' \quad N = \prod_{i=1}^k m_i! \cdot 2^{t_i} \cdot \prod_i t_i! \cdot \prod_{i < j} t_{ij}!$$

En divisant N par le nombre total a_n des appariements, on obtient la *proportion* de ceux qui ont la même configuration.

Si l'on suppose que tous les appariements ont même probabilité de se réaliser, cette proportion peut être interprétée comme la *probabilité* de la configuration.

Pour $k = 4$ et $m = 2n = 12 = 5 + 3 + 2 + 2$ (exemple ci-dessus), ces probabilités sont, après simplifications

$$Pr(\dots t_{ij} ; t_i \dots ; 5, 3, 2, 2) = \frac{2^6}{231} (2^{\square t_i} \cdot t_1! t_2! t_3! t_4! \cdot t_{12}! t_{13}! t_{14}! t_{23}! t_{24}! t_{34}!)^{\square}$$

D'où la dernière colonne du Tableau 3.1 ci-dessus pour les 4 configurations qui y figurent.

3.5. CONFIGURATION MOYENNE

Pour $m = 2n = \sum_1^k m_i$ désignons par $\bar{t}_i$ la valeur moyenne des t_i et par $\bar{t}_{ij}$ celle des t_{ij} , pour $i \neq j$.

Compte tenu des dénombrements de types de paires faits au § 3.1 ci-dessus, on a

$$(3.8) \quad \begin{aligned} \bar{t}_i &= \frac{\frac{m_i(m_i - 1)}{2}}{\frac{m(m - 1)}{2}} n = \frac{m_i(m_i - 1)}{2(2n - 1)} \\ \bar{t}_{ij} &= \frac{\frac{m_i m_j}{m(m - 1)}}{\frac{m(m - 1)}{2}} n = \frac{m_i m_j}{2n - 1} \end{aligned}$$

La configuration $\bar{t}$, qui n'est pas en nombres entiers, sera appelée *configuration moyenne*.

On vérifie immédiatement qu'elle satisfait bien à la relation (3.5) ci-dessus, à savoir

$$\sum_i \bar{t}_i + \sum_i \sum_{j < i} \bar{t}_{ij} = n$$

Dans l'exemple $m = 12 = 2 \square 6 = 5 + 3 + 2 + 2$, on obtient ainsi le Tableau 3.3

$\bar{t}_{12}$	$\bar{t}_{13}$	$\bar{t}_{14}$	$\bar{t}_{23}$	$\bar{t}_{24}$	$\bar{t}_{34}$	$\bar{t}_1$	$\bar{t}_2$	$\bar{t}_3$	$\bar{t}_4$	Total
$\frac{15}{11}$	$\frac{10}{11}$	$\frac{10}{11}$	$\frac{6}{11}$	$\frac{6}{11}$	$\frac{4}{11}$	$\frac{10}{11}$	$\frac{3}{11}$	$\frac{1}{11}$	$\frac{1}{11}$	$\frac{66}{11} = 6$
1,36	0,91	0,91	0,545	0,545	0,364	0,91	0,273	0,091	0,091	5,999

Tableau 3.3

Une autre vérification immédiate est que, d'après les valeurs (3.8) □

$$(3.9) \quad \forall i, \quad m_i = 2\bar{t}_i + \sum_{j \neq i} \bar{t}_{ij}$$

La configuration moyenne est ainsi une solution (en nombres rationnels, en général non entiers) du système (S_k) . Le fait que cette configuration vérifie (3.5) en est d'ailleurs une conséquence immédiate.

4. LE CAS DICHOTOMIQUE

4.1. ÉCHELLE (ORDRE TOTAL) DES CONFIGURATIONS

Dans le cas où E est divisé en deux classes E_1 et E_2 , le système (S_2) est □

$$m_1 = 2t_1 + t_{12} = 2t_1 + t$$

$$m_2 = 2t_2 + t_{12} = 2t_2 + t$$

C'est la seule donnée de $t_{12} = t$ qui détermine chaque configuration.

Comme on suppose □ $m = 2n = m_1 + m_2$, m_1 et m_2 ont même parité que t .

Si m_1 et m_2 sont pairs, t peut prendre les valeurs □ $0, 2, \dots, \min(m_1, m_2)$ qui constituent une échelle de $\left\lfloor \frac{m_1}{2} + \frac{1}{2} \min(m_1, m_2) \right\rfloor$ échelons.

Si m_1 et m_2 sont impairs, t peut prendre les valeurs □ $1, 3, 5, \dots, \min(m_1, m_2)$ □ cette fois-ci, le nombre d'échelons est $\frac{1 + \min(m_1, m_2)}{2}$.

Exemples

- Pour $m = 60 = 38 + 22$, il y a 12 configurations totalement ordonnées par l'échelle des valeurs de t , celle des 12 nombres pairs de 0 à 22.
- Pour $m = 20 = 11 + 9$, il y a 5 configurations totalement ordonnées sur l'échelle □

$$1 < 3 < 5 < 7 < 9$$

des valeurs possibles de t .

L'ordre ainsi obtenu pour les configurations est un ordre de mixité croissante et d'homogénéité décroissante.

Les tableaux T.4.1 et T.4.2, § 4.3 ci-après, détaillent cet ordre total des configurations pour ces deux exemples.

4.2. ÉCARTS À LA MOYENNE

Les configurations constituent un ordre total, ou échelle. Mais on peut également les repérer par rapport à la configuration moyenne □

$$\bar{t} = \frac{m_1 m_2}{m \square 1} \quad \bar{t}_1 = \frac{m_1 (m_1 \square 1)}{2(m \square 1)} \quad \bar{t}_2 = \frac{m_2 (m_2 \square 1)}{2(m \square 1)}$$

Dans ce cas, on parlera de mixité négative pour $t < \bar{t}$, et de mixité positive pour $t > \bar{t}$. Quant aux écarts absolus, tous sont proportionnels à $|t - \bar{t}|$, puisque le système (S_2) implique

$$t_1 - \bar{t}_1 = \frac{m_1 - t}{2} - \frac{m_1 - \bar{t}}{2} = \frac{\bar{t} - t}{2}, \quad t_2 - \bar{t}_2 = \frac{\bar{t} - t}{2}$$

4.3. PROBABILITÉS DES CONFIGURATIONS

Le calcul de la probabilité de chaque configuration peut se faire soit par raisonnement direct (particulièrement simple en l'occurrence), soit en utilisant les formules générales (3.7) ou (3.7)'.

Comme chaque configuration est déterminée par la seule valeur de t , on peut noter $P(t; m_1, m_2)$ sa probabilité, et l'on a

$$P(t; m_1, m_2) = \frac{1}{a_n} \binom{m_1}{t} \binom{m_2}{t} t! a_{t_1} a_{t_2}$$

Soit, après quelques simplifications

$$(4.1) \quad P(t; m_1, m_2) = \frac{1}{2^n} \frac{m_1! m_2!}{1.3.5 \dots (m-1)} \frac{2^t}{t! t_1! t_2!}$$

De l'équation (4.1), et de $\sum_t P(t; m_1, m_2) = 1$, on peut déduire

$$(4.2) \quad \sum_t \frac{2^t}{t! t_1! t_2!} = \frac{2n!}{n! m_1! m_2!} \quad (m = 2n = m_1 + m_2)$$

la somme étant prise sur la suite des valeurs possibles de t .

Si (4.2) n'est pas évidente, elle permet par contre de retrouver aisément que

$$(4.3) \quad \bar{t} = \sum_t t P(t; m_1, m_2) = \frac{m_1 m_2}{m-1}$$

Le calcul est laissé en exercice au lecteur.

La proportion des paires de type mixte, rapportée à la valeur maximale m_1 (si $m_1 \leq m_2$), a donc pour moyenne

$$\frac{\bar{t}}{m_1} = \frac{m_2}{m-1}$$

C'est-à-dire que cette proportion moyenne est, pour m grand, à peu près égale à la proportion de la classe la plus nombreuse dans la population totale : c'est conforme au «Bon sens».

Quant à la proportion des paires de type mixte dans le nombre total $n = \frac{m}{2}$ de paires, c'est

$$\frac{\bar{t}}{n} = 2 \frac{m_1 m_2}{m(m-1)}$$

Si m augmente indéfiniment de telle sorte que $\frac{m_1}{m}$ et $\frac{m_2}{m}$ aient des limites, p_1 et p_2 respectivement, alors

$$\lim_{m \rightarrow \infty} \frac{\bar{t}}{m_1} = p_2 \quad \text{et} \quad \lim_{m \rightarrow \infty} \frac{\bar{t}}{n} = 2p_1 p_2$$

Un calcul analogue à celui qui a été laissé en exercice au lecteur pour établir l'égalité (4.3) permet d'obtenir le *moment factoriel* M_2 d'ordre 2

$$(4.4) \quad M_2 = \sum_t t(t-1) P(t; m_1, m_2) = \frac{m_1 m_2 (m_1 - 1)(m_2 - 1)}{(m - 1)(m - 3)}$$

$$\text{comme} \quad M_2 = \left(\sum t^2 P(t; m_1, m_2) \right) - \bar{t}$$

On en déduit l'expression de la *variance* de t

$$(4.5) \quad V(t; m_1, m_2) = 2 \frac{m_1 m_2 (m_1 - 1)(m_2 - 1)}{(m - 1)^2 (m - 3)}$$

On a donc, pour la variance de la proportion t/m_1 de paires de type mixte

$$(4.6) \quad V\left(\frac{t}{m_1}\right) = \frac{2}{m_1} \cdot \frac{m_2}{m - 3} \cdot \frac{m_1 - 1}{m - 1} \cdot \frac{m_2 - 1}{m - 1}$$

Pour $m \geq 4$, on a $\frac{m_2}{m - 3} \leq \frac{m}{m - 3} \leq 4$. On en déduit

$$(4.7) \quad V\left(\frac{t}{m_1}\right) < \frac{2}{m_1} \cdot 4 \cdot 1 \cdot 1 = \frac{8}{m_1}$$

Cette variance tend donc vers zéro lorsque m_1 augmente indéfiniment.

De même, pour $m \geq 4$

$$(4.8) \quad V\left(\frac{t}{n}\right) = \frac{1}{n} \cdot \frac{m_1}{m} \cdot \frac{m_2}{m} \cdot \frac{m_1 - 1}{m - 1} \cdot \frac{m_2 - 1}{m - 1} \cdot \frac{m}{m - 3} < \frac{4}{n} = \frac{8}{m}$$

Ces propriétés amènent à conjecturer que, sous des conditions assez larges, la distribution de probabilité de la variable «*entrée réduite*»

$$z = \frac{t - \bar{t}}{(V(t))^{1/2}}$$

converge vers la distribution de Laplace-Gauss, lorsque m_1 et m_2 augmentent indéfiniment.

Reste à le démontrer, et à déterminer à partir de quels effectifs m_1 et m_2 l'approximation par la loi normale est acceptable.

À partir de la probabilité de chaque configuration, on peut calculer les *probabilités cumulées* (cf. Tableaux 4.1 et 4.2) ceci fournit une échelle des configurations graduée non en nombre d'échelons (valeurs de t), mais en probabilité de dépasser chaque échelon.

$$m = 2n = 20 = m_1 + m_2 = 11 + 9$$

n°	t	t_1	t_2	Probabilité	Pr cumulée
1	1	5	4	$7 \square 3^2 K_0 = 63K_0 \square 1,5 \%$	1,5 %
2	3	4	3	$2^3 \square 3 \square 5 \square 7K_0 = 840K_0 \square 20 \%$	21,5 %
3	5	3	2	$2^5 \square 3^2 \square 7K_0 = 2016K_0 \square 48 \%$	69, %
4	7	2	1	$2^7 \square 3^2 K_0 = 1152K_0 \square 27,5 \%$	97 %
5	9	1	0	$2^7 K_0 = 128K_0 \square 3 \%$	100 %
Moyenne	$\frac{99}{19}$	$\frac{55}{19}$	$\frac{36}{19}$	$K_0 = (19 \square 17 \square 13)^{\square 1} = \frac{1}{4199}$	
$\square$	5,21	2,895	1,895		

Tableau 4.1

$$m = 2n = 60 = m_1 + m_2 = 38 + 22$$

n°	t	t_1	t_2	Probabilité	$\square$	Pr cumulée
1	0	19	11	K_0	$4.10^{\square 9}$	0
2	2	18	10	$2 \square 9 \square 11K_0 = 198K_0$	$0,8.10^{\square 6}$	
3	4	17	9	$2 \square 3 \square 5 \square 11 \square 19K_0 = 6270K_0$	$2,4.10^{\square 5}$	
4	6	16	8	$2^4 \square 3^2 \square 11 \square 17 \square 19K_0 = 511.632K_0$	$1,97.10^{\square 3}$	0,2 %
5	8	15	7	$4.677.778K_0$	1,85 %	2,05 %
6	10	14	6	$21.829.632K_0$	8,425 %	10,5 %
7	12	13	5	$55.566.336K_0$	21,445 %	32 %
8	14	12	4	$79.380.480K_0$	30,63 %	62,6 %
9	16	11	3	$63.504.384K_0$	24,51 %	87 %
10	18	10	2	$27.394.048K_0 = 2^{17} \square 11 \square 19K_0$	10,57 %	97,66 %
11	20	9	1	$5.767.168K_0 = 2^{19} \square 11K_0$	2,225 %	99,225 %
12	22	8	0	$2^{20} \square \frac{3}{7}K_0$	0,173 %	100 %
Moyenne	$\frac{836}{59}$	$\frac{703}{59}$	$\frac{231}{59}$	$K_0 = (41 \square 43 \square 47 \square 53 \square 59)^{\square 1} \square 4.10^{\square 9}$		
$\square$	14,17	11,915	3,915			

Tableau 4.2

On peut également déterminer les *intervalles de confiance* autour de la configuration moyenne.

Ainsi, pour l'exemple du Tableau 4.2, l'intervalle de confiance « 5 % » (et même, en l'occurrence, à 4 %) est

$$10 \leq t \leq 18 \quad (t \in \{10, 12, 14, 16, 18\})$$

soit, pour la proportion $t/22$, l'intervalle $[5/11, 9/11]$.

De même, l'intervalle « 1 % » est

$$8 \leq t \leq 20 \quad (t \in \{0, 2, 4, 6, 22\})$$

4.4. LE CAS OÙ L'EFFECTIF TOTAL EST IMPAIR

Si $m = 2n - 1 = m_1 + m_2$ est impair, le système (S_2) est remplacé, comme on l'a dit au § 3.2 ci-dessus, par

$$S'_2 = \begin{cases} m_1 = 2t_1 + t + t_{01} \\ m_2 = 2t_2 + t + t_{02} \\ = t_{01} + t_{02} \end{cases}$$

D'où l'on déduit

$$(4.3) \quad n = t_1 + t_2 + t + 1$$

Chaque configuration est déterminée par le triplet (t_{01}, t_{02}, t) . Compte tenu de la convention $a_0 = 1$, et de (4.3) ci-dessus, sa probabilité est, en appliquant (3.7)' du § 3.3

$$P(t_{01}, t_{02}, t) = \frac{1}{2^{n-1}} \frac{m_1! m_2!}{1.3.5... (2n-1)} \frac{2^t}{t! t_1! t_2!}$$

D'autre part, l'énumération par ordre lexicographique de toutes les configurations placera en tête, lorsque m_1 est pair, toutes les valeurs paires de t , et ensuite toutes les valeurs impaires ; c'est l'inverse si m_1 est impair.

Il faudra donc, si l'on s'intéresse au degré de mixité, rétablir l'ordre naturel des valeurs de t .

Exemple. $m = 2n - 1 = 11 = 2 \times 6 - 1 = 6 + 5$

Enumération lexicographique

t_{01}	t_{02}	t	t_1	t_2
0	1	0	3	2
0	1	2	2	1
0	1	4	1	0
1	0	1	2	2
1	0	3	1	1
1	0	5	0	0

Tableau 4.3

Rangées selon les valeurs croissantes de t , les 6 configurations et leurs probabilités sont

n°	t	t_1	t_2	Probabilité	Pr. cumulée
1	0	3	2	$5 K_0 \square 2,16 \%$	2,16 %
2	1	2	2	$30 K_0 \square 13 \%$	15,16 %
3	2	2	1	$60 K_0 \square 26 \%$	41,16 %
4	3	1	1	$80 K_0 \square 34,63 \%$	75,8 %
5	4	1	0	$40 K_0 \square 17,32 \%$	93, 1 %
6	5	0	0	$16 K_0 \square 6,92 \%$	100 %
Moyenne	$\frac{30}{11}$	$\frac{15}{11}$	$\frac{10}{11}$	$K_0 = (3 \square 7 \square 11)^{-1} = (231)^{-1}$	
$\square$	2,73	1,36	0,91		

Tableau 4.4

5. LE CAS TRICHOTOMIQUE

5.1. PRÉLIMINAIRES

Dans ce paragraphe, les notions introduites seront illustrées sur deux exemples.

Dans le premier, les effectifs des trois classes E_1 , E_2 et E_3 ont des parités différentes. On a

$$m = 2n = 14 = 2 \square 7 = m_1 + m_2 + m_3 = 7 + 4 + 3$$

La liste, dans l'ordre lexicographique des triplets (t_{12}, t_{13}, t_{23}) , des 13 configurations est celle du Tableau 5.1 ci-dessous.

$$m = 2n = 14 = m_1 + m_2 + m_3 = 7 + 4 + 3$$

n°	t_{12}	t_{13}	t_{23}	t_1	t_2	t_3	Probabilité	Pr. %
1	0	1	0	3	2	1	K	0,7
2	0	1	2	3	1	0	$4K$	2,8
3	0	3	0	2	2	0	$2K$	1,4
4	1	0	1	3	1	1	$4K$	2,8
5	1	0	3	3	0	0	$8/3K$	1,865
6	1	2	1	2	1	0	$24K$	16,8
7	2	1	0	2	1	1	$12K$	8,4
8	2	1	2	2	0	0	$24K$	16,8
9	2	3	0	1	1	0	$16K$	11,2
10	3	0	1	2	0	1	$8K$	5,6
11	3	2	1	1	0	0	$32K$	22,4
12	4	1	0	1	0	1	$8K$	5,6
13	4	3	0	0	0	0	$16/3K$	3,75
Moyenne	$\frac{28}{13}$	$\frac{21}{13}$	$\frac{12}{13}$	$\frac{21}{13}$	$\frac{6}{13}$	$\frac{3}{13}$	$K = (13 \square 11)^{\square} = \frac{1}{143}$	
	2,15	1,615	0,92	1,615	0,46	0,23		

Tableau 5.1

Dans le second, les effectifs des trois classes ont même parité. C'est□

$$m = 2n = 18 = 2 \square 9 = m_1 + m_2 + m_3 = 8 + 6 + 4$$

Il y a 27 configurations, données par le Tableau 5.2, toujours dans l'ordre lexicographique pour les triplets de paires de type mixte.

$$m = 2n = 18 = m_1 + m_2 + m_3 = 8 + 6 + 4$$

n°	t_{12}	t_{13}	t_{23}	t_1	t_2	t_3	Score	Probabilité	Pr. %
a	0	0	0	4	3	2	0	K	0,014
b	0	0	2	4	2	1	2	$12K$	0,165
c	0	0	4	4	1	0	4	$8K$	0,11
d	0	2	0	3	3	1	2	$16K$	0,22
e	0	2	2	3	2	0	4	$96K$	1,3
f	0	4	0	2	3	0	4	$16K$	0,22
g	1	1	1	3	2	1	3	$192K$	2,63
h	1	1	3	3	1	0	5	$256K$	3,5
i	1	3	1	2	2	0	5	$384K$	5,26
j	2	0	0	3	2	2	2	$24K$	0,33
k	2	0	2	3	1	1	4	$192K$	2,63
l	2	0	4	3	0	0	6	$64K$	0,88
m	2	2	0	2	2	1	4	$288K$	3,95
n	2	2	2	2	1	0	6	$1152K$	15,8
o	2	4	0	1	2	0	6	$192K$	2,63
p	3	1	1	2	1	1	5	$768K$	10,53
q	3	1	3	2	0	0	7	$512K$	7
r	3	3	1	1	1	0	7	$1024K$	14
s	4	0	0	2	1	2	4	$48K$	0,66
t	4	0	2	2	0	1	6	$192K$	2,63
u	4	2	0	1	1	1	6	$384K$	5,26
v	4	2	2	1	0	0	8	$768K$	10,53
w	4	4	0	0	1	0	8	$128K$	1,76
x	5	1	1	1	0	1	7	$1536/5K$	4,21
y	5	3	1	0	0	0	9	$1024/5K$	2,81
z	6	0	0	1	0	2	6	$64/5K$	0,176
□	6	2	0	0	0	1	8	$256/5K$	0,7
moyenne	$\frac{48}{17}$ 2,82	$\frac{32}{17}$ 1,88	$\frac{24}{17}$ 1,41	$\frac{28}{17}$ 1,65	$\frac{15}{17}$ 0,88	$\frac{6}{17}$ 0,353	Total 9	$K = (3 \square 11 \square 13 \square 17) \square^1 = \frac{1}{7293}$	

Tableau 5.2

5.1.1. Patrons de parité

Lorsque le nombre k de classes est égal à 3, le système (S_3) du § 3.3 auquel doivent satisfaire les configurations s'écrit, pour $m = 2n = m_1 + m_2 + m_3$ □

$$(S_3) \begin{cases} m_1 = 2t_1 + t_{12} + t_{13} \\ m_2 = 2t_2 + t_{12} + t_{23} \\ m_3 = 2t_3 + t_{13} + t_{23} \end{cases}$$

Les t_i et les t_{ij} sont des entiers positifs ou nuls. Les trois nombres t_{12} , t_{13} et t_{23} de paires de type mixte ont des parités dépendant de celles de m_1 , m_2 , m_3 .

Si m_1 , m_2 , et m_3 sont tous pairs, les t_{ij} sont soit pairs tous les trois, soit impairs tous les trois. On dira que leur *patron de parité* est soit *PPP*, soit *III* (cf. exemple du Tableau 5.2).

Si l'un des m_i est pair, et les deux autres impairs, il y a de même deux patrons de parité possibles. Par exemple, pour m_1 et m_3 impairs et m_2 pair (exemple du Tableau 5.1), les patrons de parité sont *PIP* et *IPI*.

Les deux autres cas possibles se déduisent de celui-ci par permutation des indices.

Remarque. Cette notion de patron de parité de la suite des $\frac{k(k-1)}{2}$ nombre t_{ij} ($j > i$) est toujours définie. On y reviendra au § 8 ci-après.

5.1.2. Inversibilité

Une particularité du cas trichotomique est que le système (S_3) est *inversible* et c'est le seul cas où il en est ainsi.

Les formules d'inversion s'écrivent ($m = 2n$)

$$(S'_3) = \begin{cases} t_{12} = n + t_3 - (t_1 + t_2) - m_3 \\ t_{13} = n + t_2 - (t_1 + t_3) - m_2 \\ t_{23} = n + t_1 - (t_2 + t_3) - m_1 \end{cases}$$

Chaque configuration pourra être désignée soit par son *triplet de mixité* $t_m = (t_{12}, t_{13}, t_{23})$, soit par son *triplet d'homogénéité* $t_h = (t_1, t_2, t_3)$.

5.2. ORDRE PARTIEL DE MIXITÉ

5.2.1. Définition des ordres Mix et Hom

Il y a ainsi *bijection* entre l'ensemble des triplets (t_{12}, t_{13}, t_{23}) de mixité et les triplets (t_1, t_2, t_3) d'homogénéité.

Mais, comme on l'a souligné plus haut (§ 3.2), l'ordre lexicographique dans lequel les configurations ont été énumérées ne correspond pas à un ordre de mixité croissante, ni à un ordre d'homogénéité décroissante.

En outre, comme on le constate en examinant les Tableaux T.5.1 et T.5.2, l'ordre lexicographique pour les triplets (t_{12}, t_{13}, t_{23}) de paires de type mixte n'est pas le dual (l'inverse) de l'ordre lexicographique pour les triplets (t_1, t_2, t_3) de paires de type homogène.

Or, pouvoir *comparer* les configurations dans un ordre de mixité croissante, ou d'homogénéité décroissante, comme ce fut possible dans le cas dichotomique (cf. § 4), c'est là l'une des principales demandes du sociologue.

Ces comparaisons sont possibles, à condition de ne considérer que les *ordres partiels*, dans le produit direct de trois ordres totaux de type $\square = \{0 < 1 < 2 < \dots < h < \dots\}$, défini par

$$(5.1) \quad x = (x_1, x_2, x_3) \square y = (y_1, y_2, y_3) \quad \text{ssi} \quad \begin{array}{l} \square x_1 \square y_1 \\ \square x_2 \square y_2 \\ \square x_3 \square y_3 \end{array}$$

C'est ainsi que l'on a, dans le premier exemple (T.5.1), colonnes mixité, pour les configurations n° 6, 11 et 5

$$121 < 321, \quad \text{mais} \quad 121 \text{ non comparable à } 103$$

De même, côté homogénéité, pour les mêmes trois configurations n° 6, 11 et 5

$$210 > 100, \quad \text{mais} \quad 300 \text{ non comparable à } 210$$

Appelons *Mix* l'ordre partiel ainsi défini sur les triplets (t_{12}, t_{13}, t_{23}) de mixité, et *Hom* son homologue pour les triplets (t_1, t_2, t_3) d'homogénéité.

Mix et *Hom* sont en bijection définie par les systèmes (S_3) et (S'_3) respectivement désignons par f la fonction

$$f: \text{Mix} \rightarrow \text{Hom}$$

définie par (S_3) . Il est clair qu'elle est monotone décroissante si t_{12}, t_{13} et t_{23} augmentent tous les trois, t_1, t_2, t_3 ne peuvent que diminuer, puisque chacune des sommes $(2t_i + \sum_{j>i} t_{ij})$ est constante et égale à m_i .

Par contre, l'inverse f^{-1} , définie par (S'_3) de la fonction f de *Hom* vers *Mix*, n'est pas monotone. Par exemple, pour T.5.1, la configuration n° 4 majore la configuration n° 7 dans *Hom* $311 > 211$ mais dans *Mix* ces deux configurations, 101 et 210 respectivement, ne sont pas comparables.

5.2.2. Une décomposition en deux inf-demi-treillis

L'ordre *Mix* n'est pas quelconque tous les triplets de même patron de parité constituent un inf-demi-treillis

- l'infimum de deux triplets $t_m = (t_{12}, t_{13}, t_{23})$ et $u_m = (u_{12}, u_{13}, u_{23})$ est celui du produit direct

$$v_{ij} = \min(t_{ij}, u_{ij})$$

$$(5.2) \quad t_m \square u_m = v_m = (v_{12}, v_{13}, v_{23})$$

avec $\square (i, j), \quad v_{ij} = \min(t_{ij}, u_{ij})$.

En effet, le minimum de deux entiers de même parité a leur parité commune. Donc v_m a même patron de parité que t_m et u_m . Ainsi, les différences $(m_i \sqcap \sqcap_{j \neq i} v_{ij})$ sont toutes paires, comme le sont, par hypothèse, $(t_m$ et u_m correspondent à deux configurations satisfaisant au système (S_3)), les différences $(m_i \sqcap \sqcap_{j \neq i} t_{ij})$ et $(m_i \sqcap \sqcap_{j \neq i} u_{ij})$.

Or, d'après la définition des v_{ij} , on a

$$\forall i, (m_i \sqcap \sqcap_{j \neq i} v_{ij}) \geq \max((m_i \sqcap \sqcap_{j \neq i} t_{ij}), (m_i \sqcap \sqcap_{j \neq i} u_{ij})) \geq 0$$

Donc il y a trois entiers positifs ou nuls v_1, v_2 et v_3 tels que

$$\forall i, m_i = 2v_i + \sqcap_{j \neq i} v_{ij}$$

La configuration infimum satisfait bien au système (S_3) .

- Quant au *supremum* de deux triplets t_m et u_m de même patron de parité

- soit il n'existe pas
- soit il est défini, et est celui du produit direct

$$(5.3) \quad t_m \sqcup u_m = w_m = (w_{12}, w_{13}, w_{23}) \quad \text{avec} \quad w_{ij} = \max(t_{ij}, u_{ij})$$

En effet, comme dans le cas de l'infimum, les différences $(m_i \sqcap \sqcap_{j \neq i} w_{ij})$ sont toutes paires.

Si elles sont toutes positives, on a 3 nombres w_i tels que

$$2w_i = m_i \sqcap \sqcap_{j \neq i} w_{ij} \geq 0$$

Le système (S_3) est satisfait.

Par contre, si au moins l'une de ces différences est négative, il est impossible de satisfaire au système (S_3) , et t_m et u_m n'ont pas de supremum parmi les configurations ayant même patron de parité.

Tout ceci est bien visible sur les Figures F.5.1 et F.5.2. On y remarquera en outre que

- deux configurations de même patron de parité peuvent avoir plusieurs majorants communs minimaux non comparables et ayant leur patron de parité commun
- deux configurations non comparables de patrons de parité différents peuvent avoir une configuration infimum. Pour l'exemple du Tableau 5.1 (Figure 5.1.), il en est ainsi des configurations n° 9 (230) et n° 11 (321), qui ont pour infimum la configuration n° 7 (210).

Remarque. La démonstration qui vient d'être faite, qui ne supposait rien quant au nombre k des classes, montre que la propriété mise en évidence de l'ordre partiel *Mix* est générale. Quel que soit le nombre k de classes, l'ensemble des configurations de même patron de parité est un inf-demi-treillis.

Figure 5.1

Mais le nombre des patrons de parité peut augmenter très rapidement□ dans l'exemple du § 3.3 ($k = 4$), il y a 8 patrons de parité.

N.B. 8, c'est 2^3 , et 3, c'est $\frac{(k-1)(k-2)}{2}$. On y reviendra au § 8 *infra*.

5.3. LE TREILLIS D'HOMOGENÉITÉ

5.3.1. Infimum de deux configurations

L'ordre partiel *Hom* est un treillis. Considérons deux configurations de triplets d'homogénéité non comparables□

$$\underline{t}_h = (t_1, t_2, t_3) \quad \underline{u}_h = (u_1, u_2, u_3)$$

avec□ $t_1 > u_1$, $t_2 \geq u_2$, $t_3 < u_3$

A une permutation près des indices, ceci est le cas le plus général.

$\underline{t}_h$ et $\underline{u}_h$ ont un *infimum* $\underline{v}_h = (v_1, v_2, v_3)$ qui est celui du produit direct□

$$\forall i, v_i = \min(u_i, t_i)$$

i.e.□ $v_1 = u_1$ $v_2 = u_2$ $v_3 = t_3$

Pour le prouver, il faut montrer que l'on peut déterminer v_{12} , v_{13} , v_{23} entiers positifs ou nuls tels que (S_3) soit satisfait.

Or les formules d'inversion (S'_3) s'écrivent ici□

$$v_{12} = n + t_3 - (u_1 + u_2) - m_3 > n + t_3 - (t_1 + t_2) - m_3 = t_{12} \geq 0$$

$$v_3 = n + u_2 - (t_3 + u_1) - m_2 > n + u_2 - (u_3 + u_1) - m_2 = u_{13} \geq 0$$

$$v_{23} = n + u_1 - (u_2 + t_3) - m_1 > n + u_1 - (u_2 + u_3) - m_3 = u_{23} \geq 0$$

Ce qui démontre la propriété.

Figure 5.2. L'ordre partiel *Mix* et ses deux inf-demi-treillis
 $m = 2n = 18 = 8 + 6 + 4 = m_1 + m_2 + m_3$

5.3.2. Supremum de deux configurations

Quant au *supremum* de $\underline{t}_h$ et $\underline{u}_h$, il est éventuellement strictement supérieur à celui, $\underline{w}_h$ du produit direct.

On a pour celui-ci□

$$w_1 = t_1 \quad w_2 = t_2 \quad w_3 = u_3$$

Les formules d'inversion (S_3') donnent□

$$w_{12} = n + u_3 \square (t_1 + t_2) \square m_3$$

$$w_{13} = n + t_2 \square (t_1 + u_3) \square m_2$$

$$w_{23} = n + t_1 \square (t_2 + u_3) \square m_1$$

Figure 5.3. Le treillis *Hom* d'homogénéité et sa distribution de probabilité cumulée

Si les trois nombres des seconds membres sont positifs ou nuls, $\underline{w}_h = (w_1, w_2, w_3)$ est le supremum de $\underline{t}_h$ et $\underline{u}_h$. S'il n'en est pas ainsi, au plus un des trois nombres w_{ij} est négatif.

On a en effet $\square$

$$(5.4) \quad \begin{cases} w_{12} + w_{13} = 2n - 2t_1 - (m_2 + m_3) = m_1 - 2t_1 = t_{12} + t_{13} \geq 0 \\ w_{12} + w_{23} = 2n - 2t_2 - (m_1 + m_3) = m_2 - 2t_2 = t_{12} + t_{23} \geq 0 \\ w_{13} + w_{23} = 2n - 2u_3 - (m_1 + m_2) = m_3 - 2u_3 = u_{12} + u_{13} \geq 0 \end{cases}$$

Si deux des w_{ij} étaient négatifs, l'une au moins de ces trois sommes serait négative.

Supposons, par exemple, w_{23} négatif. Les deux premières inégalités de (5.4) ci-dessus prouvent que

$$w_{12} \geq -w_{23} = |w_{23}| \quad \text{et} \quad w_{13} \geq -w_{23} = |w_{23}|$$

Considérons alors la configuration $\underline{z}_h$ définie par

$$(5.5) \quad \begin{cases} z_1 = w_1 + |w_{23}| = t_1 - w_{23} & z_2 = w_2 = t_2 & z_3 = w_3 = u_3 \\ z_{12} = w_{12} + w_{23} \geq 0 & z_{13} = w_{13} + w_{23} \geq 0 & z_{23} = 0 \end{cases}$$

On vérifie immédiatement qu'elle satisfait au système (S₃)

$$\begin{cases} 2z_1 + z_{12} + z_{13} = 2t_1 - 2w_{23} + w_{12} + w_{13} + w_{23} = 2t_1 + t_{12} + t_{23} = m_1 \\ 2z_2 + z_{12} + z_{23} = 2t_2 + w_{12} + w_{23} = 2t_2 + t_{12} + t_{23} = m_2 \\ 2z_3 + z_{13} + z_{23} = 2u_3 + w_{13} + w_{23} = 2u_3 + u_{12} + u_{23} = m_3 \end{cases}$$

Elle majore $\underline{w}_h$ et il est clair que parmi les majorantes de $\underline{w}_h$ dans l'ordre partiel *Hom*, elle est la plus petite de celles qui satisfont à (S₃).

En effet, toute diminution de z_1 en $z_1' = z_1 - h$ ($h > 0$) rendrait le z_{23}' correspondant négatif.

Les Figures F.5.1 et F.5.2 représentent le treillis *Hom* pour les deux exemples.

Remarques.

- Ces démonstrations reposent essentiellement sur le fait que, dans le cas trichotomique, le système (S) est inversible.

Elle n'est donc pas généralisable à un nombre $k \geq 4$ de classes.

- Lorsqu'on a une distribution de probabilité sur un ordre partiel fini, les probabilités cumulées se calculent de la façon habituelle par sommation. Si l'on désigne par p_x la probabilité de chaque élément x , la probabilité cumulée P_y en y est la somme

$$(5.6) \quad P_y = \sum_{x \leq y} p_x$$

Ces probabilités cumulées sont indiquées pour le treillis *Hom* de la Figure 5.3.

Par exemple, il y a 75 % de chances pour qu'une configuration ait un triplet (t_1, t_2, t_3) d'homogénéité inférieur ou égal, dans l'ordre latticiel, à 211. Mais cette probabilité n'est plus que de 21 % pour le triplet 300, et par contre monte jusqu'à 98 % pour le triplet 321.

5.4. DÉNOMBREMENT DES CONFIGURATIONS

Notations. Dans ce paragraphe, on notera

- $x \vee y$ le plus grand de deux nombres x et y , et $x \wedge y$ le plus petit
- $[x]$ la partie entière d'un nombre positif x , et $\lceil x \rceil$ l'entier immédiatement supérieur ou égal à x . On a ainsi
  - $[x] = \lceil x \rceil = x$ ssi x est entier
  - $\lceil x \rceil = [x] + 1$ sinon.

Méthode de dénombrement. On va construire un algorithme de dénombrement des configurations t , définies en homogénéité par le triplet (t_1, t_2, t_3) , en comptant, pour chaque entier s , le nombre $C(s; m_1, m_2, m_3)$ de configurations de score s ($s \geq 0$), c'est-à-dire telles que

$$t_1 + t_2 + t_3 = s$$

si $2n = m = m_1 + m_2 + m_3$, on a évidemment $0 \leq s \leq n$.

Le résultat essentiel est que, quel que soit s , $C(s; m_1, m_2, m_3)$ (noté plus brièvement $C(s)$ dans la suite) est un *nombre triangulaire* de forme

$$C(s) = \frac{(h+1)(h+2)}{2}$$

où h est une fonction bien définie de s et des m_i .

Triangle des configurations de score s . Du système (S_3) , § 5.1.1, résultent les inégalités

$$(5.7.1) \quad \forall i \quad m_i \geq 2t_i \geq 0, \quad \text{i.e.} \quad \left\lfloor \frac{m_i}{2} \right\rfloor \geq t_i \geq 0$$

Des formules d'inversion (S'_3) , §5.1.2, et de $t_{ij} \geq 0$ pour $i \neq j$, résultent trois inégalités telles que

$$t_{12} = n + t_3 - (t_1 + t_2) - m_3 = n + 2t_3 - (t_1 + t_2 + t_3) - m_3 \geq 0$$

Soit, compte tenu de $t_1 + t_2 + t_3 = s$,

et en posant $n_i = n - m_i$,

$$\forall i \quad 2t_i \geq s - n_i, \quad \text{i.e.} \quad t_i \geq \left\lceil \frac{s - n_i}{2} \right\rceil$$

$$\text{Posons } \underline{t}_i(s) = \left\lceil \frac{s - n_i}{2} \right\rceil$$

Comme $t_i \geq 0$, on a

$$(5.7.2) \quad \forall i \quad t_i \geq \underline{t}_i(s)$$

Mais, lorsque trois nombres positifs ont une somme constante s , chacun d'entre eux est au plus égal à la différence entre s et la somme des valeurs minimum des deux autres. Ainsi

$$(5.7.3) \quad s \sqcup \square_2 \sqcup \square_3 \geq t_1, \quad s \sqcup \square_1 \sqcup \square_3 \geq t_2, \quad s \sqcup \square_1 \sqcup \square_2 \geq t_3$$

Posons

$$(5.8) \quad \begin{aligned} \square_1(s) &= \frac{m_1}{2} \square (s \sqcup \square_2 \sqcup \square_3) \\ \square_2(s) &= \frac{m_2}{2} \square (s \sqcup \square_1 \sqcup \square_3) \\ \square_3(s) &= \frac{m_3}{2} \square (s \sqcup \square_1 \sqcup \square_2) \end{aligned}$$

On a finalement, les $\square_i$ étant entiers par définition

$$(5.9) \quad \square_i \square_i(s) \geq t_i \geq \square_i(s)$$

Si l'on prouve que, quel que soit s ($0 \leq s \leq n$)

$$\square_1(s) = s \sqcup \square_2 \sqcup \square_3, \quad \square_2(s) = (s \sqcup \square_1 \sqcup \square_3), \quad \square_3(s) = s \sqcup \square_1 \sqcup \square_2$$

Alors, l'ensemble des triplets (t_1, t_2, t_3) de score s est représenté par l'ensemble des points à coordonnées entières (coordonnées triangulaires) du triangle de sommets

$$I_1 = (\square_1, \square_2, \square_3) \quad I_2 = (\square_1, \square_2, \square_3) \quad I_3 = (\square_1, \square_2, \square_3)$$

La hauteur de ce triangle est

$$h = \square_1 \sqcup \square_1 = \square_2 \sqcup \square_2 = \square_3 \sqcup \square_3$$

Le nombre $C(s)$ sera donc (cf. Figure 5.4)

Figure 5.4

$$C(s) = (h+1) + h + \cdots + 3 + 2 + 1 = \frac{(h+1)(h+2)}{2}$$

La démonstration peut se faire en plusieurs étapes.

Supposons, pour fixer les idées, que $m_3 \leq m_2 \leq m_1$. Il en résulte que, par définition

$$n_1 \leq n_2 \leq n_3 \leq n \quad (\text{puisque } n_i = n - m_i)$$

D'autre part $m_1 + m_2 + m_3 = m = 2n$

donc $2m_2 < m = 2n$,

D'où $n_2 = n - m_2 > 0$

D'autre part $n_1 + n_2 = m_3 \geq 0$

donc $n_2 \geq -n_1$

Si $n_1 < 0$, alors $\lfloor_1(s) \geq 0$ ssi $s \geq s_0 = \lfloor n_1 = m_1 \rfloor n$

Il y a ainsi deux cas

Cas 1. $0 \leq n_1$ (i.e. $m_1 \leq n$)

Cas 2. $n_1 < 0$ (i.e. $m_1 > n$)

Dans le second cas, les expressions trouvées dans le premier cas gardent leur forme algébrique mais elles ne sont applicables qu'au-delà du seuil s_0 .

Supposons maintenant $0 \leq n_1$

• Si $0 \leq s \leq n_1$

On a $\lfloor_1 = \lfloor_2 = \lfloor_3 = 0$

par ailleurs $m_3 + 2m_1 \geq m_3 + m_2 + m_1 = m = 2n$

d'où $m_3 \geq 2(n - m_1) = 2n_1 \geq 2s$

d'où $\frac{m_3}{2} \geq s$ et, comme s est entier, $\lfloor \frac{m_3}{2} \rfloor \geq s$

Soit, d'après la définition (5.8) des $\lfloor_i$, et compte tenu de $m_1 \geq m_2 \geq m_3$

$$\forall i, \lfloor_i(s) = s$$

Le triangle $I_1 I_2 I_3$ de base est celui de sommets

$$I_1 = (s, 0, 0) \quad I_2 = (0, s, 0) \quad I_3 = (0, 0, s)$$

On a $h(s) = s$ et $C(s) = \frac{(s+1)(s+2)}{2}$

• Si $n_1 < s \leq n_2$

On a : $\lfloor_2 = \lfloor_3 = 0$

Quant à $\lfloor_1(s)$, si $(s - n_1)$ est *pair*, c'est $\frac{s - n_1}{2}$ si $(s - n_1)$ est *impair*, $(s - 1 - n_1)$ est pair et l'on a

$$\lfloor \frac{s - n_1}{2} \rfloor = \frac{s - 1 - n_1}{2}, \quad \text{d'où} \quad \lfloor_1(s) = \lfloor \frac{s - n_1}{2} \rfloor + 1 = \frac{s - n_1 + 1}{2}$$

D'autre part, on a la suite d'inégalités

$$2s \leq 2n_2 = 2(n - m_2) = m - 2m_2 = m_1 - (m_2 - m_3) \leq m_1$$

D'où, comme s est entier $\square$ $s \square \frac{m_1}{2}$

et $\square_1(s) = s$

Par ailleurs $\square$

$$2(s \square \square_1) = 2s \square 2\square_1 = \begin{cases} 2s \square (s \square n_1) = s + n_1 \\ 2s \square (s \square n_1 + 1) = s + n_1 \square 1 \end{cases}$$

selon la parité de $s - n_1$.

On a maintenant la suite d'inégalités $\square$

$$s + n_1 \square 1 < s + n_1 \square n_2 + n_1 = m \square m_1 \square m_2 = m_3$$

C'est-à-dire $\square$ $2(s \square \square_1) \square m_3$

et comme $(s \square \square_1)$ est entier $\square$

$$s \square \square_1 \square \frac{m_3}{2} \square \frac{m_2}{2}$$

Finalement, d'après la définition (5.8) $\square$

$$\square_2(s) = \square_3(s) = s \square \square_1$$

Le triangle $I_1 I_2 I_3$ de base a pour sommets $\square$

$$I_1 = (s, 0, 0) \quad I_2 = (\square_1, s \square \square_1, 0) \quad I_3 = (\square_1, 0, s \square \square_1)$$

Sa hauteur h est, selon la parité de $s - n_1$ $\square$

$$h = s \square \square = \begin{cases} \frac{s + n_1}{2} \\ \text{ou} \\ \frac{s + n_1 \square 1}{2} \end{cases}$$

Remarque. On a vu que si $n_1 < 0$, $s_0 = \square n_1 \square n_2$. Donc l'expression ci-dessus de h vaut dès que $s \geq s_0$.

- Si $n_2 < s \square n_3$, on a $\square_3 = 0$ et, selon la parité de $s - n_1$ et $s - n_2$ $\square$

$$2\square_1 = \begin{cases} s \square n_1 \\ \text{ou} \\ s - n_1 + 1 \end{cases} \quad 2\square_2 = \begin{cases} s \square n_2 \\ \text{ou} \\ s - n_2 + 1 \end{cases}$$

d'où, pour $2(\square_1 + \square_2)$, les valeurs possibles $\square$

$$2s \square (n_1 + n_2) = 2s \square m_3 \square \quad 2s \square m_3 + 1 \square \quad 2s \square m_3 + 2$$

d'où $\square$

$$2(s - n_1 - n_2) = m_3, \text{ ou } (m_3 - 1), \text{ ou } (m_3 - 2)$$

$$\text{Donc } \square_3(s) = \frac{m_3}{2} - (s - n_1 - n_2) = s - n_1 - n_2$$

Par ailleurs

$$2(s - n_1) = \begin{cases} s + n_1 \\ \text{ou} \\ s + n_1 - 1 \end{cases} \quad \text{donc} \quad 2(s - n_1) - n_3 + n_1 = m_2$$

$$\text{Comme } (s - n_1) \text{ est entier } s - n_1 = \frac{m_2}{2}.$$

Et, d'après la définition (5.8)

$$\square_2(s) = s - n_1$$

On a de même

$$2(s - n_2) = \begin{cases} s + n_2 \\ \text{ou} \\ s + n_2 - 1 \end{cases} \quad \square \quad n_3 + n_2 = m_1$$

$$\text{Et } \square_1(s) = s - n_2.$$

Le triangle $I_1 I_2 I_3$ de base est

$$I_1 = (s - n_2, n_2, 0) \quad I_2 = (n_1, s - n_1, 0) \quad I_3 = (n_1, n_2, s - n_1 - n_2)$$

Sa hauteur est $h(s)$

$$h(s) = \frac{m_3}{2} \quad \text{si } (s - n_1) \text{ et } (s - n_2) \text{ sont tous deux pairs}$$

$$h(s) = \frac{m_3 - 1}{2} \quad \text{si l'un est pair et l'autre impair}$$

$$h(s) = \frac{m_3}{2} - 1 \quad \text{si tous deux sont impairs.}$$

- Si $n_3 < s - n$.

La discussion est analogue, avec quatre cas possibles selon que les $2n_i$, qui sont tous les trois strictement positifs, sont

- (a) tous pairs. On a alors $\square \square i, 2(s - \square \square_j) = m_i$

- (b) deux pairs et un impair. On a alors $2(s - \square \square_j) = \begin{cases} m_i \\ \text{ou} \\ m_i - 1 \end{cases}$

(c) un pair et deux impairs. On a alors $2(s \prod_{j \neq i} m_j) = \begin{cases} m_i - 1 \\ m_i + 2 \end{cases}$ ou

(d) tous impairs. On a alors $2(s \prod_{j \neq i} m_j) = m_i - 2$

donc, dans tous les cas

$$m_i(s) = \frac{m_i}{2} (s \prod_{j \neq i} m_j) = s \prod_{j \neq i} m_j$$

Le triangle $I_1 I_2 I_3$ est inclus dans le triangle de hauteur s , et l'on a

$$I_1 = (m_1, m_2, m_3) \quad I_2 = (m_1, m_2, m_3) \quad I_3 = (m_1, m_2, m_3)$$

La hauteur est $h(s)$ avec $2h = 2s \prod (m_1 + m_2 + m_3)$.

Dans les 4 cas de parités ci-dessus, on a respectivement

$$(a) \quad h = \frac{n \prod s}{2}$$

$$(b) \quad h = \frac{n \prod s + 1}{2}$$

$$(c) \quad h = \frac{n \prod s}{2} - 1$$

$$(d) \quad h = \frac{n \prod s + 3}{2}$$

Remarque. Certaines des valeurs $h(s)$ ainsi définies peuvent éventuellement être strictement négatives. C'est le cas notamment pour $s = 8$, dans l'exemple $18 = m = m_1 + m_2 + m_3 = 2n = 8 + 6 + 4$.

Dans cet exemple, on a $n_1 = 1, n_2 = 3, n_3 = 5$. D'où le tableau 5.3.

Le lecteur pourra se reporter au Tableau 5.2 et à la Figure 5.3.

In fine, on obtient le nombre total des configurations par sommation

$$C(m_1, m_2, m_3) = \sum_{s=0}^n C(s; m_1, m_2, m_3)$$

soit, 27 dans l'exemple.

s	$h(s)$	$C(s)$
0	$s = 0$	1
1	$s = 1$	3
2	$\frac{s + n_1 - 1}{2} = 1$	3
3	$\frac{s + n_1}{2} = 2$	6
4	$\frac{\begin{smallmatrix} \square & m_3 & \square \\ \square & 2 & \square \end{smallmatrix}}{2} = 1$	3
5	$\frac{\begin{smallmatrix} \square & m_3 & \square \\ \square & 2 & \square \end{smallmatrix}}{2} = 2$	6
6	$\frac{n - s - 3}{2} = 0$	1
7	$\frac{n - s}{2} = 1$	3
8	$\frac{n - s - 3}{2} = -1$	0
9	$\frac{n - s}{2} = 0$	1

Tableau 5.3

6. LE TREILLIS D'HOMOGENÉITÉ POUR PLUS DE 3 CLASSES

Supposons l'ensemble E partitionné en k classes, ou catégories, avec $k \geq 4$, et soient $m_1, \dots, m_i, \dots, m_k$ les effectifs de chacune des classes. On a

$$|E| = m = \sum_{i=1}^k m_i$$

Rappelons que m est supposé *pair*.

Une configuration (§ 3.3) est une suite de $\frac{k(k+1)}{2}$ entiers positifs ou nuls

$$\underline{t} = (t_1, \dots, t_i, \dots, t_k, t_{12}, \dots, t_{ij}, \dots, t_{k-1,k}) \quad \text{avec} \quad t_{ij} = t_{ji}$$

satisfaisant au système linéaire (S_k)

$$(S_k) \quad m_i = 2t_i + \sum_{j \neq i} t_{ij} \quad (i = 1, 2, \dots, k)$$

Les suites, dites *suites d'homogénéité*

$$\underline{t} = (t_1, t_2, \dots, t_k)$$

de nombres de paires de type homogène sont ordonnées partiellement dans l'ordre produit direct de $\square^k$ de k ordres totaux $\square = \{0 < 1 < 2 \dots < n < \dots\}$ par

$$\underline{t} \square \underline{u} \square \square i, \quad t_i \square u_i$$

Désignons par $Hom(m_1, m_2, \dots, m_k)$ l'ordre partiel qu'elles constituent.

6.1. ON PEUT TOUJOURS SUPPOSER PAIRS TOUS LES EFFECTIFS DE CLASSE

Supposons que, dans la suite des nombres $m_1, \dots, m_k$, les h premiers soient impairs, et les autres pairs. Le nombre h lui-même est nécessairement pair.

Le *majorant universel* de $Hom(m_1, \dots, m_k)$ est obtenu en posant, dans (S_k) , $t_{ij} = 1$ si i et j sont inférieurs ou égaux à h , et $t_{ij} = 0$ sinon. C'est donc

$$\frac{1}{2}(m_1 \square 1, \dots, m_h \square 1; m_{h+1}, \dots, m_k)$$

Montrons que $Hom(m_1, \dots, m_k)$ et $Hom(m_1 - 1, \dots, m_k - 1 \square m_{k+1}, \dots, m_k)$ sont identiques.

Soit, en effet $\underline{t} = (t_1, \dots, t_k \square \dots, t_{ij}, \dots)$ une configuration satisfaisant à (S_k) .

Pour $1 \square i \square h$, on a

$$m_i \square 2t_i = \square_{j \neq i} t_{ij} \geq 1$$

Donc il existe au moins un indice $j_0 (1 \square j_0 \square h)$ tel que

$$t_{ij_0} \geq 1$$

Supposons la numérotation faite de telle sorte que, pour $i = 1$, on ait $j_0 = 2$, et numérotions 3 le plus petit des indices restant $\square$ pour $i = 3$, on désignera par 4 le j_0 correspondant $\square$ etc...

Compte tenu de cette permutation des h premiers indices, on a

$$t_{12} \geq 1, t_{34} \geq 1, \dots, t_{h \square 1, h} \geq 1; \quad t_{ij} \geq 0 \quad \text{si} \quad i > h \quad \text{et} \quad j > h$$

Posons $t'_{12} = t_{12} \square 1 \geq 0, \dots, t'_{h \square 1, h} = t_{h \square 1, h} \square 1 \geq 0$

Et $t'_{ij} = t_{ij}$ pour tous les autres couples (i, j) . Alors la configuration

$$\underline{t}' = (t_1, \dots, t_k \square \dots, t'_{ij}, \dots)$$

satisfait au système linéaire $(S'_k) \square$

$$(S'_k) \begin{cases} m_i \square 1 = 2t_i + \square_{j \neq i} t'_{ij} & \text{pour } 1 \square i \square h \\ m_i = 2t_i + \square_{j \neq i} t'_{ij} & \text{pour } h+1 \square i \square k \end{cases}$$

Donc la même suite $\underline{t} = (t_1, \dots, t_k)$ d'homogénéité, élément de $Hom(m_1, \dots, m_k)$ est aussi élément de $Hom(m_1 - 1, \dots, m_h - 1 \square m_{h+1}, \dots, m_k)$.

Soit inversement $\square$

$$\mathbf{u}' = (u_1, \dots, u_k \square \dots, u'_{ij}, \dots)$$

Une configuration satisfaisant au système $(S'_k) \square$

$$(S'_k) \begin{cases} m_i \square 1 = 2u_i + \square_{j \neq i} u'_{ij} & \text{pour } 1 \leq i \leq h \\ m_i = 2u_i + \square_{j \neq i} u'_{ij} & \text{pour } h+1 \leq i \leq k \end{cases}$$

avec $u'_{ij} \geq 0, \square (i, j).$

Alors, en posant $u_{12} = u'_{12} + 1, u_{34} = u'_{34} + 1, \dots, u_{h \square 1, h} = u'_{h \square 1, h} + 1$ et $u_{ij} = u'_{ij}$ pour tous les autres couples (i, j) , la configuration $\square \mathbf{u} = (u_1, \dots, u_k \square \dots, u_{ij}, \dots)$ satisfait à $(S_k) \square$

$$(S_k) \square i, m_i = 2u_i + \square_{j \neq i} u_{ij}$$

et la même suite $\underline{u} = (u_1, \dots, u_2)$ de $Hom(m_1 - 1, \dots, m_h - 1 \square m_{h+1}, \dots, m_k)$ est élément de $Hom(m_1, m_2, \dots, m_k)$.

Conclusion. Pour démontrer que Hom est un treillis, on peut supposer que tous les m_i sont pairs, ce qu'on fait dans la suite.

6.2. NOTATIONS ET REMARQUE TRIVIALE

On va démontrer par récurrence sur le nombre k de catégories que $Hom(m_1, \dots, m_k)$ est un treillis, dans lequel l'infimum et le supremum de deux suites $\underline{t}$ et $\underline{u}$ d'homogénéité seront notés respectivement $\underline{t} \square \underline{u}$ et $\underline{t} \sqcup \underline{u}$.

La démonstration reposera sur cette *remarque triviale* que si $x_1, x_2, \dots, x_k$ sont des réels positifs non tous nuls, on a pour au moins $(k - 1)$ d'entre eux l'inégalité stricte $\square$

$$(6.1) \quad x_i < \square_{j \neq i} x_j$$

On ne peut avoir l'inégalité contraire, pour un indice $i_0 \square x_{i_0} \geq \square_{j \neq i_0} x_j$ que si x_{i_0} est le maximum des x_i , et si $\square$

$$i \neq i_0 \square x_i < x_{i_0}$$

Dans la suite, on notera $t_i \square u_i$ et $t_i \sqcup u_i$ le plus petit et le plus grand respectivement de deux nombres t_i et u_i , et par $\underline{t} \square \underline{u}$ et $\underline{t} \sqcup \underline{u}$ les deux suites $\square$

$$\underline{t} \square \underline{u} = (\dots, t_i \square u_i, \dots) \quad \underline{t} \sqcup \underline{u} = (\dots, t_i \sqcup u_i, \dots)$$

6.3. INFIMUM DE DEUX SUITES D'HOMOGENÉITÉ

Soient $\underline{t} = (t_1, \dots, t_k)$ et $\underline{u} = (u_1, \dots, u_k)$ les deux suites d'homogénéité de configurations $\underline{t}$ et $\underline{u}$ satisfaisant à (S_k)

$$\forall i, \quad m_i = 2t_i + \sum_{j \neq i} t_{ij} = 2u_i + \sum_{j \neq i} u_{ij}$$

On a alors, dans $Hom(m_1, \dots, m_k)$, où tous les m_i sont pairs

$$\underline{t} \sqcup \underline{u} = \underline{t} \sqcup \underline{u}$$

Supposons en effet que parmi les nombres positifs ou nuls $(m_i - 2(t_i - u_i))$ celui d'indice 1 satisfait à l'inégalité stricte (6.1)

$$(6.1)' \quad m_1 - 2(t_1 - u_1) < \sum_{i \neq j} (m_i - 2(t_i - u_i))$$

Hypothèse de récurrence. Il existe des $v_{ij} \geq 0$ tels que soit vérifié le système (S_{k-1})

$$(S_{k-1}) \quad m_i = 2(u_i - t_i) + \sum_{j \neq i} v_{ij} \quad (2 \leq i \leq k)$$

(S_{k-1}) et $(6.1)'$ impliquent

$$(6.2) \quad \sum_{i=2}^k (m_i - 2(u_i - t_i)) = 2 \sum_{i=2}^k \sum_{j>i} v_{ij} > m_1 - 2(u_1 - t_1) \geq 0$$

On peut donc choisir des entiers $\square_{ij}$ tels que $v_{ij} \geq \square_{ij} \geq 0$ et

$$(6.3) \quad 2 \sum_{i=2}^k \sum_{j>i} \square_{ij} = m_1 - 2(t_1 - u_1)$$

Posons $v'_{ij} = v_{ij} - \square_{ij}$ pour $\begin{pmatrix} 2 \leq i \leq k \\ 2 \leq j \leq k \end{pmatrix}$ et $i \neq j$

et $v'_{1i} = \sum_{j \neq i} \square_{ij}$

On a alors $\sum_{i=2}^k v'_{1i} = 2 \sum_{i=2}^k \sum_{j>i} \square_{ij}$

On vérifie alors immédiatement que la configuration

$$(t_1 - u_1, t_2 - u_2, \dots, t_k - u_k; \dots, v'_{ij}, \dots)$$

satisfait au système (S_k)

$$(S_k) \quad \begin{cases} m_1 = 2(t_1 - u_1) + \sum_{i=2}^k v'_{1i} \\ m_i = 2(t_i - u_i) + \sum_{j \neq i} v'_{ij} \quad (2 \leq i \leq k) \end{cases}$$

$\underline{t} \sqcup \underline{u}$ est donc un minorant commun de $\underline{t}$ et $\underline{u}$ dans $Hom(m_1, \dots, m_k)$ mais comme pour tout minorant commun $\square$ dans Hom on a nécessairement

$$\bigwedge_i \bigwedge u_i \bigwedge t_i, \quad \bigwedge i$$

on a bien $\bigwedge \underline{t} \bigwedge \underline{u} = \underline{t} \bigwedge \underline{u}$.

Cette égalité est vraie pour $k = 3$ (§ 5.3.1, *supra*), elle l'est donc pour tout $k \geq 4$.

6.4. SUPREMUM DE DEUX SUITES D'HOMOGENÉITÉ

Soient $\underline{t} = (t_1, t_2, \dots, t_k)$ et $\underline{u} = (u_1, u_2, \dots, u_k)$ deux suites de $Hom(m_1, \dots, m_k)$.

On a vu que, pour $k = 3$, leur supremum $\bar{t} \bigwedge \bar{u}$ dans Hom peut être strictement supérieur à $\underline{t} \bigwedge \underline{u} = (\dots, t_i \bigwedge u_i, \dots)$.

Ce supremum existe quel que soit le nombre $k \geq 4$ de catégories.

En effet, $\underline{t}$ et $\underline{u}$ ont dans $Hom(m_1, \dots, m_k)$ des majorants communs, ne serait-ce que le majorant universel $\frac{1}{2}(m_1, m_2, \dots, m_k)$.

Si l'ensemble $S(\underline{t}, \underline{u})$ des majorants communs de $\underline{t}$ et $\underline{u}$ dans Hom avait plusieurs éléments minimaux distincts, soient $\underline{v}$ et $\underline{w}$ deux d'entre eux.

D'après ce qu'on vient de démontrer § 6.3, ils ont dans Hom un infimum $\underline{v} \bigwedge \underline{w} = \underline{v} \bigwedge \underline{w}$.

Comme $\bigwedge v_i \geq t_i \bigwedge u_i$ et $\bigwedge w_i \geq t_i \bigwedge u_i$

implique $\bigwedge v_i \bigwedge w_i \geq t_i \bigwedge u_i$, i.e. $\underline{v} \bigwedge \underline{w} \geq \underline{t} \bigwedge \underline{u}$, $\underline{v} \bigwedge \underline{w}$ est un majorant commun de $\underline{t}$ et $\underline{u}$ et par suite $\underline{v}$ et $\underline{w}$ ne sont pas minimaux, ou sont confondus.

Cette démonstration élémentaire prouve seulement l'existence du supremum

$$\underline{t} \bigwedge \underline{u} = \min S(\underline{t}, \underline{u})$$

Elle ne donne pas de procédé de construction de $\underline{t} \bigwedge \underline{u}$ connaissant $\underline{t}$ et $\underline{u}$.

Voici encore un problème pour les chercheurs.

7. RANGER LES CONFIGURATIONS SUR UNE ÉCHELLE

7.1. QUELQUES INDICATEURS BASÉS SUR LE PRÉORDRE PAR «SCORES»

Rappelons que pour toute configuration $\underline{t}$, le «score en homogénéité» est le nombre s de ses paires de type homogène (cf. § 5.3)

$$s = \sum_{i=1}^k t_i$$

On a de même un «score en mixité» s'

$$s' = \sum_{i=1}^k \sum_{j>i} t_{ij}$$

Les valeurs moyennes (ou «théoriques»), $\bar{s}$ et $\bar{s}'$ respectivement, de s et s' sont données par les formules (3.8) du § 3.5.

Pour $m = 2n = m_1 + m_2$, on a

$$s + s' = \bar{s} + \bar{s}' = n$$

À partir de s et s' , on peut définir divers indices d'homogénéité croissante.

Remarques préalables. Aux fins de comparaison de situations d'effectif total m différentes, on ne considèrera que des indices homogènes (de dimension 0, au sens des physiciens).

- si deux indices sont distincts, c'est-à-dire ne sont pas l'un fonction monotone croissante de l'autre, il y aura nécessairement des contradictions dans les comparaisons effectuées au moyen de chacun d'entre eux

- mais comme tous les nombres en jeu sont des entiers, les indices varient par sauts discrets il pourra donc se faire qu'en pratique ces contradictions n'apparaissent pas, ou n'apparaissent que peu.

Voici quelques indices qui ont été effectivement utilisés

Indice	Définition	Signification	Domaine de variation	Remarques
i_1	s/n	Proportion de paires homogènes	0 à 1	
i'_1	s'/n	Idem, pour les mixtes	0 à 1	$i'_1 = 1 - i_1$
i_2	$\frac{s - \bar{s}}{n} = \frac{\bar{s}' - s'}{n}$	Ecart à la moyenne	-1 à 1	$i_2 = i_1 - \bar{i}_1$
i_3	$\frac{s - s_0}{s_1 - s_0}$	i_1 rapporté au domaine effectif de variation de s	0 à 1 bornes comprises	On peut avoir (§ 3.5) $0 < s_0 \leq s \leq s_1 \leq n$
i_4	s/s'	Rapport du nombre des homogènes à celui des mixtes	0 à l'infini	
i_5	$s/\bar{s}$	Homologue de i_2 en échelle multiplicative	0 à l'infini	
i_6	$(s/\bar{s}) : (s'/\bar{s}')$	Bi-rapport. Le plus synthétique de tous	0 à l'infini	$i_6 = i_4 \frac{\bar{s}'}{\bar{s}} = \frac{i_1(1 - \bar{i}_1)}{\bar{i}_1(1 - i_1)}$

Tableau 7.1

Un dernier indicateur de degré d'homogénéité, d'une nature différente des précédents, est la proportion $\text{Pr}(s/n)$ des configurations de score inférieur ou égal à s . Cette proportion peut se calculer au moyen des formules du § 3.4 on cumule les proportions de configurations de scores allant de 0 à s .

Dans le cadre d'un *modèle probabiliste* où les configurations seraient considérées comme « tirées au hasard », ces proportions peuvent s'interpréter comme des probabilités.

À titre d'exemple, le Tableau 7.2 donne la valeur de ces indices pour les 13 configurations du Tableau 5.1, et de la Figure 5.1.

Configurations	s	Pr %	Pr Cum.	i_1	i_2	i_3	i_4	i_5	i_6
000	0	3,75	3,75	0	- 0,33	0	0	0	0
100	1	22,4	26,15	0,143	- 0,187	0,167	0,167	0,433	0,339
200, 110, 101	2	33,6	59,75	0,286	- 0,044	0,333	0,4	0,867	0,813
300, 201, 210	3	24,15	83,9	0,427	0,099	0,5	0,75	1,3	1,525
310, 211, 220	4	12,6	96,5	0,571	0,242	0,667	1,333	1,733	2,711
311	5	2,8	99,3	0,714	0,385	0,833	2,5	2,167	5,083
321	6	0,7	100	0,857	0,527	1	6	2,6	12,2
néant	7	0							

Tableau 7.2

Il est conseillé au lecteur de faire le même travail pour les 27 configurations du Tableau 5.2 et de la Figure 5.3□ il pourra constater quelques contradictions entre indices.

Par exemple, appelons A_s la classe de score s dans le premier cas, et B_s dans le second. On a□

- $A_1 < B_1$ pour i_4 et i_6 □ $B_1 < A_1$ pour les autres
- $A_4 < B_5$ pour i_5 et Pr Cum.□ $B_5 < A_4$ pour les autres

7.2. UN ORDRE TOTAL□L'ÉCHELLE LEXICO-HIÉRARCHIQUE

L'ordre lexicographique dans lequel on écrit les configurations pour les énumérer toutes «sans omission ni répétition□ est compatible avec l'ordre partiel du treillis Hom (§ 6) des k -uplets $(t_1, t_2, \dots, t_k)$ de nombres de paires homogènes lorsque E est divisé en k catégories $E_1, E_2, \dots, E_k$ en ce sens que si $(t_1, t_2, \dots, t_k) \geq (t'_1, t'_2, \dots, t'_k)$ dans Hom , alors $(t_1, t_2, \dots, t_k)$ est supérieur ou égal à $(t'_1, t'_2, \dots, t'_k)$ dans l'ordre lexicographique.

Par contre, il n'est pas compatible en général, avec le préordre total par scores du paragraphe 7.1 précédent. Par exemple, pour les configurations du Tableau 7.1 on a, dans l'ordre lexicographique, $300 > 220$, alors que le premier est de score 3 inférieur au score 4 du second.

Pour qu'il y ait compatibilité, on peut décider d'ordonner lexicographiquement toutes les configurations d'une même classe du préordre par «scores□. *Encore faut-il que cela ait un sens en termes d'appariements* dans E □c'est le cas si les catégories $E_1, E_2, \dots, E_k$ entre lesquelles E est divisé, sont *hiérarchisées*, comme le sont par exemple, pour l'aïkido, la division par grades, ou celle par classes d'âge.

C'est pourquoi nous appelons l'ordre total ainsi obtenu *ordre lexico-hiérarchique*, de façon à bien souligner cette condition *nécessaire* (la hiérarchie des classes) pour ranger les configurations sur cette échelle.

Le Tableau 7.3 fournit l'échelle lexico-hiérarchique pour les 13 configurations du Tableau 7.2, dans deux axes, parmi les 6 possibles, de hiérarchie des classes.

Les échelles lexico-hiérarchiques ne sont qu'ordinales, comme l'ordre lexicographique lui-même.

Le rang r d'une configuration repère sa position sur l'échelle, et rien de plus. Il serait absurde d'en déduire des indices numériques d'homogénéité croissante, comme on l'a fait au paragraphe précédent pour les scores, qui, eux, sont des nombres.

Le seul indicateur cohérent avec le caractère purement ordinal de l'ordre lexicohiérarchique, et qui permette la comparaison des «degrés d'homogénéité» pour des situations variées (i.e., à effectifs totaux ou à effectifs de classes variables, pour des divisions différentes en classes, pour divers ordres hiérarchiques des classes, etc...) c'est la *proportion* (ou la *probabilité*) cumulée des configurations inférieures ou égales à chaque configuration.

Par exemple, selon le Tableau 7.3, la configuration 101 est beaucoup moins homogène (32 % environ) pour la hiérarchie $E_1 > E_2 > E_3$ que pour la hiérarchie $E_2 > E_3 > E_1$ (48,5 %).

s'	t_{12}	t_{13}	t_{23}	t_1	t_2	t_3	s	r	Pr Cum.	t_1	t_2	t_3	Pr Cum.
7	4	3	0	0	0	0	0	1	3,75	0	0	0	3,75
6	3	2	1	1	0	0	1	2	26,15	1	0	0	26,15
5	4	1	0	1	0	1	2	3	31,75	2	0	0	43
-	2	3	0	1	1	0	-	4	42,95	1	0	1	48,5
-	2	1	2	2	0	0	-	5	59,75	1	1	0	59,75
4	3	0	1	2	0	1	3	6	65,35	3	0	0	61,5
-	1	2	1	2	1	0	-	7	82,15	2	0	1	67,9
-	1	0	3	3	0	0	-	8	83,9	2	1	0	83,9
3	2	1	0	2	1	1	4	9	92,3	3	1	0	86,8
-	0	3	0	2	2	0	-	10	93,7	2	1	1	95,2
-	0	1	2	3	1	0	-	11	96,5	2	2	0	96,5
2	1	0	1	3	1	1	5	12	99,3	3	1	1	99,3
1	0	1	0	3	2	1	6	13	100	3	2	1	100

Tableau 7.3

Remarques. 1) L'ordre $E_1 > E_2 > \dots > E_k$ entre classes implique pour les paires de type homogène T_i , l'ordre $T_1 > T_2 > \dots > T_k$. C'est d'ailleurs sur cette considération qu'est fondée la construction de l'ordre lexico-hiérarchique des configurations.

Par contre pour les paires de type mixte T_{ij} , si l'on a bien, pour 3 classes

$$T_{12} > T_{13} > T_{23}$$

l'ordre est partiel dès que le nombre de classes dépasse 3 il n'y a aucune raison a priori de ranger une paire de type T_{14} avant, ou après, une paire de type T_{23} . On aura ainsi

Chaque échelon d'une échelle lexico-hiérarchique, ordonnée sur les k -uplets $(t_1, t_2, \dots, t_k)$ de nombres de paires de type homogène, est donc, pour $k \geq 4$, non une seule configuration, mais la classe de toutes les configurations ayant le même k -uplet d'homogénéité (cf. § 3.3, remarques).

2) Puisque l'échelle est seulement ordinale, cela n'aurait aucun sens de pondérer les paires homogènes avec des poids p_i tels que

$$p_1 > p_2 > \dots > p_k > 0$$

et de donner à la configuration $(t_1, t_2, \dots, t_k) = \underline{t}$ le poids (i.e., la «note» pédagogique qui lisez ceci, prenez en bonne... note)

$$p(\underline{t}) = \prod_{i=1}^k p_i t_i$$

On peut d'ailleurs démontrer, grâce aux formules de dénombrement du § 5.3 (supra), que si l'on fait cela pour $k = 3$, l'ordre des «notes» $p(\underline{t})$ n'est compatible avec l'ordre lexico-hiérarchique que si le triplet des p_i tend, lorsque les effectifs de classe augmentent indéfiniment, vers l'équipondération ($p_1 = p_2 = p_3$).

Mais si tous les p_i sont égaux entre eux, $p(\underline{t})$ est proportionnel au score $s = \sum t_i$ on retombe donc dans le préordre par scores.

7.3. CONCLUSIONS

S'il n'y a pas de hiérarchie des classes, le choix est dans l'alternative suivante

- Ou bien préordonner les configurations au moyen des scores. Mais cela se paye d'une considérable «perte d'information», chaque classe du préordre contenant plusieurs configurations, et éventuellement beaucoup
- Ou bien ordonner partiellement les configurations selon le treillis *Hom* (il y a une bien moindre «perte d'information»), ou par le demi-treillis *Mix* (on conserve toute l'information) des § 5 et 6. Mais l'interprétation devient beaucoup plus complexe.

S'il y a hiérarchie des classes, utiliser l'échelle lexico-hiérarchique graduée selon les probabilités (ou les proportions, selon les questions posées) cumulées.

8. QUESTIONS AUX MATHÉMATICIENS

8.1. UNE GÉNÉRALISATION NATURELLE

Le problème des appariements est un cas particulier de la classe suivante de problèmes : un ensemble E étant partitionné en k classes $E_1, E_2, \dots, E_k$ données (partition Π), on regroupe ses éléments non par paires, mais de façon qu'ils constituent une seconde partition Π' soumise à certaines conditions. Par exemple

- le nombre des classes et les effectifs de celles-ci dans Π' sont donnés : c'est l'un des problèmes de «tableaux dont les marges sont données», comme Maurice Fréchet les appelait jadis
- le nombre de classes est fixé, mais non leurs effectifs
- toutes les classes, sauf une éventuellement, ont même nombre d'éléments : on forme des trios, ou des quatuors, etc...

- y compris aucune condition sur Π' .

L'outil de base pour étudier de tels problèmes est l'infimum $\inf \Pi'$ de Π et de Π' dans le treillis des partitions de E . Chaque *configuration* sera définie par le «Tableau croisé» des effectifs des classes de $\Pi \Pi'$.

Si les classes données $E_1, E_2, \dots, E_k$ de E correspondent aux colonnes du tableau, les conditions imposées à Π' porteront sur les lignes dans le cas des appariements, tous les totaux de ligne valent 2.

Quant au système (S_k) d'équations linéaires en nombres entiers positifs ou nuls (§ 3.3), il s'obtiendra en écrivant que chaque total de colonne i est égal à $|E_i|$, effectif de la classe E_i .

La construction, pour la statistique inférentielle, de «Plans en blocs», doit d'ailleurs pouvoir être incluse dans la même classe de problèmes.

Question aux lecteurs qui auront eu le courage de nous lire jusqu'ici que savent-ils de tels problèmes, et de leurs solutions? Un florilège serait fort utile.

8.2. RETOUR AUX APPARIEMENTS QUESTIONS EN SUSPENS

Nous les récapitulons ici, paragraphe par paragraphe

§ 2 • trouver pour les rapports r_n un encadrement qui converge rapidement vers leur limite r (§ 2.3)

- étudier la vitesse de convergence de la chaîne de Markov (§ 2.5) vers la situation stationnaire.

§ 4 • étudier la convergence éventuelle vers la loi de Laplace-Gauss de la distribution de la variable t (nombre des paires mixtes) centrée et réduite (§ 4.3).

§ 5 • trouver une méthode ou un algorithme d'énumération et de dénombrement des «Motifs de parité» (§ 5.5.1 et 5.2.2, remarques).

Leur nombre est $2^{\frac{(k-1)(k-2)}{2}}$ si $\min_i(m_i) \geq k-1$. Mais il peut être très inférieur.

Dans le cas extrême où tous les m_i valent 1, ce nombre n'est plus que a_k du § 2.1.

- étudier plus avant le demi-treillis *Mix* (§ 5.2.2)

- trouver pour un nombre $k \geq 4$ de classes un algorithme de dénombrement des configurations (§ 5.3 pour $k = 3$).

La méthode suivie au § 5.3 est spécifique au cas de 3 classes, et ne semble pas généralisable. Pour le seul dénombrement des suites $t = (t_1, t_2, \dots, t_k)$ d'homogénéité, on peut conjecturer que le rôle joué par les nombres triangulaires pour $k = 3$ le sera par les nombres binomiaux de forme

$$\frac{(h+1)(h+2)\dots(h+k-1)}{(k-1)!}$$

§ 6 • trouver un algorithme de construction du supremum de deux configurations dans le treillis *Hom* (§ 6.4)

- plus généralement, y a-t-il un algorithme, et un programme informatique associé, qui permettrait de «Sortir» le treillis Hom liste des sommets, et, pour chacun, liste de ses éléments couvrants et de ses éléments couverts, chaînes, etc...»

9. COMMENTAIRES DU SOCIOLOGUE

Revenons sur le problème sociologique qui a motivé les développements mathématiques précédents. En considérant un critère particulier (genre, âge, grade..., le choix d'un partenaire en aikido se fait-il au hasard ou existe-t-il une tendance à former des paires homogènes ou au contraire hétérogènes – mixtes. Jusqu'à présent, les seuls moyens disponibles consistaient soit à calculer le rapport entre les paires homogènes et mixtes, soit à utiliser le test du khi-deux. Or, ni l'un ni l'autre n'étaient satisfaisants. Le premier parce qu'il ne peut tenir compte des contraintes de structure (importance des effectifs des catégories en présence) le second, parce qu'il ne peut aboutir à la construction d'une échelle d'homogénéité. Avec la possibilité d'établir l'ensemble des configurations possibles et calculer les probabilités cumulées, on dispose désormais des moyens de rendre les observations comparables et de créer une échelle permettant de les situer les unes par rapport aux autres. C'est ce que nous avons fait en étudiant les choix faits par une soixantaine d'aïkidokas à l'occasion d'un stage d'été. Au total, plus de 2000 paires ont été enregistrées à l'occasion de 91 séquences d'entraînement. Connaissant par ailleurs une série de propriétés susceptibles de faire sens dans le choix des partenaires (genre, âge, caractéristiques physiques, grade, ancienneté dans la pratique, antériorité dans le stage, club d'origine, rôle dans le stage, etc.), on a étudié systématiquement leur effet sur le niveau d'homogénéité atteint dans les différentes séquences.

On ne reviendra pas ici sur l'analyse sociologique de ces résultats présentée par ailleurs⁶ mais on voudrait souligner, pour le sociologue, l'intérêt et les limites que nous voyons à la démarche proposée.

Le premier intérêt porte sur la nature des variables prises en considération : la démarche par les configurations et les probabilités associées n'est pas limitée aux variables d'identité (avoir le même genre, le même âge, la même origine). Elle peut également s'appliquer à celles qui décrivent les relations entre individus (être connu par, être sous-traitant de, utiliser les services de). Être homogène ce n'est pas nécessairement être semblable et l'analyse ne se limite pas ici à tester la similitude. Concrètement, cela signifie que l'on peut tester non seulement si le choix du partenaire s'est fait sur la base de traits partagés mais aussi sur la base d'une relation préalable, plus généralement tester les effets de similitude comme tester ceux des liens sociaux. C'est ce qui a été fait dans le cas de l'aïkido.

Le second intérêt est de tenir compte des effets structurels. Même lorsqu'elles sont possibles, les mesures de distance entre les configurations par le nombre de paires homogènes ou mixtes qui les séparent ou par le rapport des paires homogènes/mixtes sont directement tributaires de paramètres structurels : effectif total du groupe et effectifs des différentes catégories. Ce qui rend les comparaisons difficiles dès que ces paramètres varient. À l'inverse, en situant une configuration observée parmi l'ensemble des configurations possibles et à partir des probabilités qui lui sont associées, on

⁶ «Apprécier l'homogénéité dans la formation de paires. Similitude et liens sociaux dans le choix du partenaire en aikido», *Revue française de Sociologie*, 48-2, 2007, p. 333-367.

neutralise en quelque sorte les effets structurels en ramenant la mesure à une échelle de variation contrôlée (de zéro à un) et à un critère commun (la probabilité cumulée).

Un troisième intérêt de la démarche proposée est de permettre les comparaisons à un double niveau□

- Entre différentes variables, ce qui permet de faire un premier tri parmi les variables jouant sur l'effet à expliquer. Concrètement, ceci nous a permis d'éliminer de l'explication des effets supposés jouer sur le choix du partenaire en aikido mais n'ayant de fait que peu d'importance□caractéristiques physiques (taille, poids), âge. À l'inverse, de mieux situer les effets des variables ayant le plus de poids (antériorité au stage, genre, origine géographique commune).
- Entre ce que nous avons appelé les séquences d'entraînement qui correspondent de fait à des observations dans le temps. La recherche d'une évolution d'un effet particulier devient alors possible et relativement aisée. Et c'est là un intérêt indéniable quand on connaît la difficulté d'étudier tout processus temporel surtout quand il s'agit de relations entre individus. Dans nos données, ceci nous a permis de montrer comment un effet majeur (la connaissance préalable des autres partenaires) s'atténuait dans le temps.

Les limites de la démarche résident d'abord dans les caractéristiques de la population étudiée. La situation que nous avons analysée en aikido relève manifestement de la micro-sociologie ($n = 60$) et nous sommes incapables de dire (au delà d'un vrai problème qui serait celui de l'observation concrète) comment la démarche pourrait s'appliquer à des échantillons importants. Plus fondamentalement, la démarche s'applique à une population «fermée□ ou à un réseau complet (classe, école, club, association, entreprise) et à des situations où le choix du partenaire reste libre et ouvert, autrement dit où des préférences peuvent s'exercer, ce qui n'est pas le cas de situations sur-déterminées par une condition préalable (un partenaire d'un genre différent dans le mariage, par exemple), ou celles où les individus sont en compétition (jeux antagonistes), ou appariés malgré eux (équipes de travail).

Une seconde série de limites tient dans le traitement des variables. Tout d'abord, dans notre application, nous n'avons pas dépassé la manipulation de plus de quatre catégories par variable et les difficultés pratiques à ce niveau sont déjà importantes. Théoriquement, rien ne s'oppose à traiter une variable ayant davantage de catégories. Mais il est difficile de dire jusqu'où il est possible d'aller concrètement. Ensuite, si la démarche permet de comparer différents effets, elle ne peut prendre en compte sans complications excessives les interactions entre variables (par ex. en aikido le lien entre grade et âge, ancienneté et grade). On est donc limité à un traitement variable par variable dont l'avantage principal est de nourrir des hypothèses et de venir à l'appui d'une argumentation. Enfin, l'approche reste fondamentalement «Par les catégories□ et non «Par les relations□. Ainsi, lorsqu'on considère la variable «Antériorité au stage□, il s'agit d'abord de l'appartenance à un groupe où l'interconnaissance générale est supposée et non de la mesure dans chaque paire d'une interconnaissance établie qui pourrait s'y trouver limitée. Ce qui veut dire que l'on continue à fonder l'analyse à partir des appartenances de groupes (ou de catégories) et non à partir des relations duelles. L'analyse proposée peut donc venir à l'appui d'une analyse de réseau mais elle ne peut pas s'y substituer.

Il reste à inviter les lecteurs sociologues à tester la démarche proposée dans l'étude des relations électives. Même si les contraintes que nous avons énoncées sont fortes (obligation d'être en paires appliquée à l'ensemble de la population concernée□ absence d'interdit dans la sélection d'un partenaire□ enjeu coopératif□ on choisit un

partenaire pour faire équipe, pas un adversaire), les situations correspondantes restent nombreuses. Énonçons-en quelques-unes. La plus prototypique est sans doute celle de la formation de binômes dans un cadre scolaire. Lorsqu'au cours d'une leçon le maître demande à ses élèves de se mettre par deux en vue de réaliser une tâche, on aboutit à la fin du processus à un ensemble de paires dont personne n'est exclu. Si on affaiblit l'exigence de temps⁷, la même situation d'appariement généralisé peut se retrouver dans des contextes variés. On en trouve des exemples dans des situations sportives ou ludiques où le groupe de base est la paire □ choisir un partenaire au tennis au sein de son club ou, d'une façon plus générale, jouer en double, faire équipe pour jouer aux cartes ou aux boules dans un club de loisirs... mais aussi dans la vie courante □ partager sa chambre ou choisir son voisin de siège dans un bus à l'occasion d'un voyage organisé, etc.

L'intérêt d'une investigation systématique de ces situations serait essentiel pour la sociologie. Il permettrait, à nos yeux, d'aboutir à préciser les effets majeurs pesant sur la formation des liens, et donc à mieux expliciter les logiques les plus à même de rendre compte des appariements dans la vie sociale.

---

⁷ Autrement dit, si la formation des paires ne se réalise pas de façon instantanée mais peut s'étaler dans le temps tout en se faisant au sein d'une population identifiée et délimitée (par exemple, dans un club de tennis, le choix de partenaires au cours d'une saison sportive).