MATHÉMATIQUES ET SCIENCES HUMAINES

P. BOVET

Génération automatique de mots circulaires et équilibres

Mathématiques et sciences humaines, tome 49 (1975), p. 29-41

<http://www.numdam.org/item?id=MSH_1975__49__29_0>

Math. Sci. hum. (*13*^{e} année, n°*49*, *1975*, p.*29-41*)

GENERATION AUTOMATIQUE DE MOTS CIRCULAIRES ET EQUILIBRES

par

P. BOVET *

INTRODUCTION

Depuis plusieurs années (cf. Barbut 1966 a, b, c et Durup 1967) nous utilisons des mots circulaires et équilibrés, pour déterminer l'enchaînement des essais de situations de psychologie expérimentale. De telles structures permettent en effet de joindre heureusement des propriétés d'équilibre et d'aléatorisation, propriétés qui correspondent respectivement aux exigences de contrebalancement et d'imprévisibilité imposées dans l'expérimentation en psychologie.

Si la définition et le dénombrement de ces structures circulaires constituent des problèmes aujourd'hui résolus, leur génération systématique par contre, n'est pas l'objet à notre connaissance, d'un algorithme simple. Le point est d'importance car seule une méthode rigoureuse nous permet de sauvegarder le caractère aléatoire des suites d'éléments obtenus.

C'est là précisément le but du programme présenté ici : obtenir des suites aléatoires issues de mots circulaires et équilibrés de dimensions quelconques, qui soient directement utilisables pour l'expérimentation.

I. DEFINITIONS ET PROPRIETES

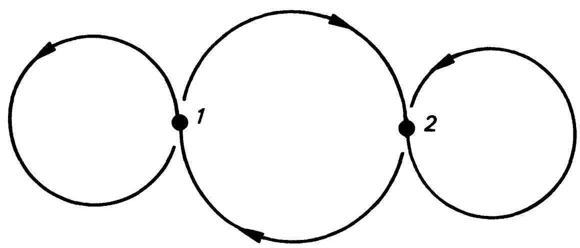
1.1. <u>Soit un ensemble S de cardinal n dont nous identifierons les éléments par les entiers</u>

$$1, 2, \ldots, i, \ldots n.$$

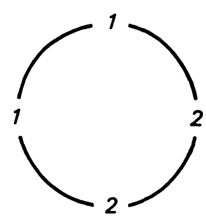
* Laboratoire de Psychologie Expérimentale, Département de Psychologie, Associé au C. N. R. S., Université de Provence, 13621 Aix-en-Provence

Nous appelons mot circulaire et équilibré * de type (n, r), un circuit de longueur n^r sur l'ensemble S tel que chacun des n^r r-uplets d'éléments de S apparaisse dans ce circuit une fois et une seule.

Voici par exemple un mot circulaire et équilibré de type (2, 2) :



Nous pourrions écrire ce mot circulairement et selon le sens trigonométrique ainsi :



Mais nous utilisons de préférence l'écriture en ligne consistant à répéter en dernière position et entre parenthèses le premier élément écrit.

Soit, pour notre exemple :

1 1 2 2 (1) = 1 2 2 1 (1) = etc.

Illustrons par un autre exemple la propriété caractéristique des mots circulaires : le circuit

2 1 1 1 2 1 2 2 (2)

est un mot circulaire de type (2, 3). Construit sur un ensemble à 2 éléments, il a pour longueur $n^r = 2^3 = 8$. Et on vérifie que tous les triplets {(1 1 1), (1 1 2), (1 2 1), (1 2 2), (2 1 1), (2 1 2), (2 2 1), (2 2 2)} apparaissent bien une fois et une seule chacun dans ce circuit.

---

* ou simplement : mot circulaire, dans le présent texte.

De la propriété caractéristique des mots circulaires et équilibrés (apparition unique de tous les n^r r-uplets d'éléments de S) découle immédiatement le corollaire suivant valable pour chacune des valeurs entières k strictement inférieures à r.

1.2. <u>Dans un mot circulaire de type (n, r) chacun des n^k k-uplets d'éléments de S apparaît n^{r-k} fois (pour k = 1, 2,..., r-1)</u>.

On notera qu'ainsi, en particulier, chacun des n éléments de S apparaît n^{r-1} fois. Et chacun des n^2 couples d'éléments de S apparaît n^{r-2} fois, etc.

Le dénombrement des mots circulaires peut être étudié sur la base du dénombrement de chemins dans des graphes bien définis (cf. Durup, op. cit.). Dans cette optique les mots circulaires (ou <u>circuits</u>) sont considérés comme des <u>cycles</u> munis d'une orientation ; les cycles pouvant eux-mêmes être regroupés en classes d'équivalence nommées <u>structures cycliques</u> par relation entre les permutations des éléments de S.

Nous nous contenterons ici de donner la formule définitive du dénombrement des mots circulaires et équilibrés, rappelée par Barbut (op. cit.).

1.3. <u>Il existe $n^{-r}(n!)^{n^{r-1}}$ mots circulaires de type (n, r)</u>.

II. MOTS CIRCULAIRES ET SUITES EQUILIBREES

On peut imaginer des situations expérimentales où les mots circulaires sont utilisés tels quels pour l'enchaînement de la suite des situations-stimulus. Dans de telles situations (conditionnement animal par exemple) le caractère circulaire de nos structures pourrait même bel et bien se trouver matérialisé au plan du dispositif d'expérience (ruban perforé rebouclé par exemple).

Mais en général, ce ne sont pas précisément les mots circulaires qui sont utilisés dans les expériences, mais des suites extraites de tels mots par simple coupure et déroulement. Ainsi la suite

1 1 2 1 2 2 2 1

est extraite du mot circulaire

2 1 1 1 2 1 2 2 (2)

de type (2, 3).

2.1. Nous appelons suite équilibrée de type (n, r) toute suite ainsi obtenue à partir d'un mot circulaire de type (n, r).

Un mot circulaire étant composé de n^r lettres, il peut être coupé en n^r points ; et on peut donc obtenir à partir de lui, n^r suites équilibrées.

2.2. Le nombre total de suites équilibrées de type (n, r) n'est d'ailleurs autre que le nombre de mots circulaires de type (n, r) multiplié par n^r.
Soit :

$$n^{-r}\ (n!)^{n^{r-1}}\ \times n^r = (n!)^{n^{r-1}}$$

On notera que pour des valeurs croissantes de n et de r ce nombre atteint très vite des grandeurs astronomiques (cf. Tableau 1).

TABLEAU 1

Nombre de suites équilibrées de type (n, r)
(l'écriture $>10^m$ entraîne ici $<10^{m+1}$)

	r = 2	r = 3	r = 4
n = 2	4	16	256
n = 3	216	10.077.696	$>10^{21}$
n = 4	331.776	$>10^{22}$	$>10^{78}$

Soit (S_i) $i = 1, 2 \ldots n^r$ une suite équilibrée de type (n, r).

Par définition du mot circulaire dont elle est extraite on a :

$$(S_j, S_{j+1}, \ldots, S_{j+r-1}) \neq (S_k, S_{k+1}, \ldots, S_{k+r-1})$$

$$\forall\ j,\ \forall\ k \neq j \quad \text{avec } j,\ k \in \{1, 2, \ldots n^r-(r-1)\}$$

C'est-à-dire que toute sous-suite de (S_i) de longueur r est unique.

Cependant, à la différence de ce que l'on trouve dans un mot circulaire, l'ensemble des sous-suites de longueur r extractibles dans une suite équilibrée de type (n, r) n'épuise pas l'ensemble des n^r r-uplets d'éléments de S.

Il manque r-1 r-uplets qui ne sont autres que les sous-suites chevauchantes :

$$(S_{n^r-(r-2)}, S_{n^r-(r-3)}, \ldots, S_{n^r}, S_1)$$

$$(S_{n^r-(r-3)}, \ldots, S_{n^r}, S_1, S_2)$$

$$\vdots$$

$$(S_{n^r}, S_1, \ldots, S_{r-2}, S_{r-1})$$

C'est pourquoi, pratiquement, nous utilisons toujours pour l'expérimentation des suites équilibrées de type (n, r) précédées de r-1 éléments supplémentaires appelés communément stimulus "bidons". Ces r-1 stimulus ne sont autres que les r-1 derniers éléments de la suite équilibrée considérée. Ayant uniquement pour fonction d'équilibrer parfaitement le passé expérimental des sujets, ils donnent lieu à des réponses qui ne sont généralement pas analysées.

2.3. <u>Nous appelons suite équilibrée complétée de type (n, r) une suite équilibrée de type (n, r) précédée de ces r-1 éléments supplémentaires</u>.

Etant donné, par exemple, la suite équilibrée de type (3, 3) suivante :

1 3 1 1 2 3 2 1 2 2 2 3 3 1 3 2 2 1 1 1 3 3 3 2 3 1 2

Nous noterons la suite équilibrée complétée qui lui correspond ainsi :

{1 2} 1 3 1 1 2 3 2 1 2 2 2 3 3 1 3 2 2 1 1 1 3 3 3 2 3 1 2

III. PRINCIPE DE GENERATION ALEATOIRE DE SUITES EQUILIBREES

Une première méthode envisageable pour construire une suite équilibrée de type (n, r) consiste en un tirage de n^r éléments de S avec la contrainte suivante applicable à partir du $(r+1)^{\text{ème}}$ élément tiré :

3.1. <u>Tout élément tiré S_i (pour $i = r+1, r+2, \ldots n^r$) doit constituer avec ses r-1 antécédents un r-uplet</u>

$$\underline{(S_{i-(r-1)}, S_{i-(r-2)}, \ldots, S_{i-1}, S_i)}$$

<u>qui ne soit encore jamais apparu</u>.

3.2. <u>Si on arrive à tirer ainsi n^r éléments successifs, il reste, pour établir que la suite des éléments tirés constitue bien une suite équilibrée, à vérifier que les r-1 sous-suites chevauchantes décrites ci-dessus §II sont précisément les r-uplets qui manquent pour épuiser l'ensemble des n^r r-uplets d'éléments de S</u>.

On notera que cette méthode peut être définie comme un tirage aléatoire exhaustif des n^r r-uplets d'éléments de S, qui respecte la contrainte suivante :

Soit $(S_i, S_{i+1}, \ldots, S_{i+(r-2)}, S_{i+(r-1)})$ le précédent r-uplet tiré (avec $i = 1, 2 \ldots n^r-1$) ; seuls seront alors acceptables les r-uplets de la forme : $(S_{i+1}, S_{i+2}, \ldots, S_{i+(r-1)}, x)$ ou x figure un élément quelconque de S.

Présentée sous cette forme, la méthode conduit directement à l'obtention d'une suite équilibrée complétée.

Il nous faut maintenant remarquer, que notre première méthode de génération aléatoire de suites équilibrées conduit à un grand nombre d'impasses et d'échecs, c'est-à-dire respectivement à des suites bloquées par le jeu de la contrainte (3.1.) ou à des suites incorrectes au regard du critère (3.2.).

Ainsi par exemple au niveau de suites équilibrées de type (2, 3) si les tirages suivants ont pu être poursuivis jusqu'au bout et constituent bien des suites équilibrées (comme le lecteur peut s'en assurer facilement),

<u>Exemple n°1</u> 1 1 1 2 1 2 2 2
<u>Exemple n°1 bis</u> 1 2 1 1 1 2 2 2

un tirage débutant ainsi :

<u>Exemple n°2</u> 1 1 1 2 1 1 . .

doit être interrompu car il n'y a plus de triplets disponibles commençant par 1 1.

Quant aux tirages ci-dessous :

Exemple n°3	1 2 1 1 2 2 1 2
Exemple n°4	1 2 1 1 1 2 2 1

ils constituent des suites qui bien que de longueur 8 ne sont pas équilibrées car les triplets chevauchants sont des triplets déjà utilisés.

De quelque type qu'ils soient, les échecs et les impasses sont très coûteux du point de vue de la production de suites équilibrées aléatoires. Entraînant l'abandon de la suite en cours de construction, ils augmentent considérablement le temps de traitement nécessaire à l'obtention de suites satisfaisantes.

Un grand nombre d'impasses et d'échecs peuvent être évités avec l'adoption d'un algorithme plus rigoureux que celui de notre première méthode. Le principe de l'amélioration est le suivant :

3.3. Contrôler en cours de génération d'une suite équilibrée de type (n, r) non plus seulement l'exhaustivité des r-uplets mais encore les fréquences respectives de tous les k-uplets, pour k = 1, 2,..., r-1.

Nous avons vu (supra 1.2.) que dans un mot circulaire tout k-uplet apparaît n^{r-k} fois. Il en est de même dans une suite équilibrée complétée. Si donc un k-uplet (k = 1, 2,..., r-1) figure déjà n^{r-k} fois dans une suite en cours de construction, on s'interdira dès lors tout tirage entraînant une nouvelle apparition de ce k-uplet.

Nous illustrerons l'efficacité de ces contrôles supplémentaires en reprenant nos exemples ci-dessus d'impasses et d'échecs dans des suites de type (2, 3).

Ainsi dans l'exemple n°2, l'impasse

1 1 1 2 1 1

apparue à la suite du tirage du dernier élément correspondant au triplet admissible 2 1 1, est évitée

- soit par contrôle du nombre d'éléments isolés (ou 1-uplets) apparus,
- soit par contrôle du nombre de couples (ou 2-uplets) apparus.

En effet, avant le dernier 1 tiré, l'élément 1 figure déjà 2^{3-1} fois, et le couple 1 1 déjà 2^{3-2} fois.

On peut se poser la question du caractère nécessaire du contrôle des fréquences des k-uplets pour toutes les valeurs de k entre 1 et r-1. La considération de nos exemples n°3 et n°4 ci-dessus nous conduit à répondre affirmativement.

En effet, dans l'exemple n°3 le tirage du dernier élément peut être évité par le contrôle de la fréquence des couples et par lui seul (couple 1 2 apparu déjà 2 fois), tandis que dans l'exemple n°4 c'est le contrôle de la fréquence des éléments et non pas des couples qui permet d'éviter le dernier élément tiré (élément 1 apparu déjà 4 fois).

En définitive, la méthode que nous utilisons pour obtenir une suite équilibrée de type (n, r) est la suivante : la suite est constituée de proche en proche du rang 1 au rang n^r. Pour chaque rang p on tire au hasard un élément x de S (S = {1, 2,..., n}) et on cherche alors, successivement pour k = 1 à r *, si le k-uplet terminé par x est déjà appru n^{r-k} fois. Si c'est le cas, on tire alors un élément de S différent de x et on recommence le contrôle, toujours pour le rang p. Lorsqu'un élément de S satisfait au contrôle pour toutes les valeurs de k, il est retenu pour constituer l'élément de rang p de la suite équilibrée en cours de construction. Mais si tous les éléments de S ont été successivement testés sans succés (cas d'impasse), la suite en cours est purement et simplement abandonnée et un nouveau tirage est alors repris à partir du rang 1. Il en est de même si bien que parvenu jusqu'au rang n^r, le contrôle a posteriori que l'on effectue alors sur les r-uplets chevauchants révèle que l'ensemble des r-uplets n'est point épuisé (cas d'échec).

Il nous faut en effet souligner que notre méthode de génération de suites équilibrées avec contrôle de la fréquence de tous les k-uplets (et non des seuls r-uplets) n'évite complètement ni les impasses ni les échecs.

Ainsi, si les suites de nos exemples ci-dessus n°2 et n°4 sont transformées automatiquement avec notre méthode en les suites équilibrées de nos exemples n°1 et n°1 bis, il n'en est malheureusement pas de même pour la

---

* En fait, pour les différentes valeurs de k, le contrôle n'est effectué qu'à partir des rangs au moins égaux à k.

suite de l'exemple n°3 qui reste bloquée après le rang 7, ni de la suite ci-dessous

Exemple n°5 1 2 1 1 2 2 2 1

qui bien que parvenue sans encombres jusqu'à 8 éléments ne constitue pas une suite équilibrée.

IV. CARACTERISTIQUES DU PROGRAMME "SUITEQ"

Le programme génère des suites équilibrées aléatoires de type (n, r) pour des valeurs n et r quelconques.* Lors d'un même travail, on peut obtenir un nombre quelconque de suites équilibrées à condition qu'elles soient de même type.

Ces 3 nombres, n, r et le nombre (ns) de suites désirées sont au demeurant les seuls paramètres qu'un utilisateur a à fournir pour l'éxécution d'un travail.

Sans donner ici copie du programme lui-même, mentionnons-en les particularités techniques fondamentales.**

4.1. Respectant la méthode de génération aléatoire décrite ci-dessus, le programme SUITEQ réserve en mémoire centrale, pour un travail donné, des tableaux dont le nombre de dimensions dépend du paramètre r. Et ces tableaux sont ensuite utilisés dans un ensemble de boucles emboîtées, le nombre de ces boucles étant lui-même fonction de r.

En effet, pour une suite de type (n, r), le décompte en cours de génération de tous les k-uplets pour des valeurs de k allant de 1 à r nécessite un ensemble de r tableaux comportant respectivement de 1 à r dimensions (toutes les dimensions comportant n niveaux). Ces tableaux sont initialisés à l'aide du nombre définitif de k-uplets que l'on doit trouver, et c'est par décrémentations successives que le décompte des k-uplets déjà apparus est tenu à jour au cours de la construction. A la fin de celle-ci, le contrôle s'effectue au niveau du tableau des r-uplets par simple comparaison à zéro.

---

* Le programme SUITEQ ne s'accomoderait pas de valeurs de r supérieures à 31... Mais on est à ce niveau bien au delà de l'ordre de grandeur de suites équilibrées qui puissent être utilisables dans la pratique expérimentale.

** Le programme (275 lignes) ainsi que toutes indications complémentaires peuvent être obtenus en écrivant à l'Auteur.

Les ensembles d'instructions auxquels il faut faire appel sont donc très particuliers en ce sens que leur taille et leur structure dépend du paramètre r.

Voici 2 exemples d'ensemble d'instructions (écrites en PL1) s'appliquant respectivement aux réservations initiales et au contrôle final * :

```
           DCL U(N) ; U = N**(R-1) ;
           DCL UU(N,N) ; UU = N**(R-2) ;
           .
r lignes   .
           .
           DCL UU...U(N,N,...,N) ; UU...U = 1
               r fois  r fois      r fois

           DO M = 1 TO N ;
           DO MM = 1 TO N ;
           .
r lignes   .
           .
           DO MM...M = 1 TO N ;
              r fois
           IF UU...U(M,MM,...,MM...M) = 0
              r fois          r fois
           THEN GO TO ECHEC ;
           END ;
           END ;
           .
r lignes   .
           .
           END ;
```

* U, UU, UUU... sont les tableaux de contrôle des k-uplets, et N et R ne sont autres que les paramètres n et r.

ECHEC est une étiquette extérieure aux sactions de programme considérées ici.

Nous avons réussi à rendre automatique la génération de tels ensembles d'instructions grâce à l'utilisation du préprocesseur MACRO-PL1.

Par exemple, les 2 ensembles d'instructions MACRO ci-dessous ont précisément pour fonction la génération des 2 ensembles variables d'instructions présentés plus haut :

```
% DCL (I,J) FIXED ;
% DCL (A,B) CHAR ;
% B='' ;
% DO I=1 TO R ;
% A='(L' ;
% B='U'||B ;
% A=B||A ;
% DO J=1 TO I-1 :
% A=A||', L' ;
% END ;
% A=A||')' ;
DCL A ; B=L**(R-I) ;
% END ;
```

```
DO M=1 TO N ;
% A='M' ;
% B=B||'(M' ;
% DO I=2 TO R ;
% A=A||'M' ;
DO A=1 TO N ;
% B=B||', '||A ;
% END
% B=B||')' ;
IF B =0
THEN GO TO ECHEC ;
% DO I=1 TO R ;
END ;
% END ;
```

De la sorte, l'utilisateur a affaire à un programme unique quelque soit la valeur de r. La seule contrainte qu'il subit, après celle d'avoir à retenir l'option MACRO au niveau du système, est de fournir la valeur de r à l'aide d'une instruction MACRO à placer dans les premières cartes du programme.

Par exemple, pour obtenir 10 suites équilibrées de type (5, 3) il faudra écrire :

N = 5 ; % R = 3 ; NS = 10 ;

4.2. Indiquons ici très brièvement une autre particularité technique de notre programme : l'aléatorisation d'une suite équilibrée est obtenue par simulation d'un tirage <u>exhaustif</u> dans une urne contenant les éléments de S en leurs effectifs définitifs (utilisation de notre sous-programme TIRAGE). Jointe à l'élimination des éléments déjà tirés à un rang déterminé de la construction d'une suite, cette procédure entraîne un gain de temps considérable dans le tirage des éléments au cours de la construction d'une suite équilibrée.

Au plan du système, elle nécessite simplement l'accès à un sous-programme RANDOM du type classique, sous-programme que nous initialisons d'ailleurs à l'aide de la fonction TIME.

4.3. Terminons cet article par quelques illustrations (purement indicatives) des capacités du programme SUITEQ. Avec l'ordinateur IBM 370/168 du CIRCE, nous avons obtenu une suite équilibrée de type (4, 3), donc de longueur 64, en 4 secondes de calcul environ (compilation complète comprise). Dans les mêmes conditions une suite de type (5, 3) -longueur 125- a été générée en 5". Une suite de type (2, 9) -longueur 512- en 14" 1/2. Et une suite de type (4, 5) -longueur 1024- en 19".

BIBLIOGRAPHIE

BARBUT M., "Un exercice de combinatoire des mots circulaires et équilibrés", <u>Math. et Sc. Hum.</u>, 14, (1966a).

BARBUT M., "Mots circulaires et équilibrés", <u>Math. et Sc. Hum.</u>, 16, (1966b).

BARBUT M., "Mots circulaires et équilibrés. Histoire du problème vu à travers la bibliographie", Math. et Sc. Hum., 17, (1966c).

DURUP H., "Graphes et plans d'expériences temporels, mots circulaires et plans toriques", Math. et Sc. Hum., 18, (1967).