MATHÉMATIQUES ET SCIENCES HUMAINES

W. FERNANDEZ DE LA VEGA
A. GUENOCHE

Construction de mots circulaires aléatoires uniformément distribués

Mathématiques et sciences humaines, tome 58 (1977), p. 25-29
<http://www.numdam.org/item?id=MSH_1977__58__25_0>

Math. Sci. hum. (15^{e}année, n°58, *1977*, p.25-29)

CONSTRUCTION DE MOTS CIRCULAIRES ALEATOIRES UNIFORMEMENT DISTRIBUES ⁎

W. FERNANDEZ DE LA VEGA ⁎⁎
A. GUENOCHE ⁎⁎

INTRODUCTION

Soient n et r deux entiers. On désigne par $[n]$ l'ensemble des n premiers entiers naturels:

$$[n] = \{1,2,\ldots n\}.$$

On appelle mot circulaire équilibré de type (n,r) une séquence circulaire orientée de longueur $L = n^r$ construite sur l'alphabet $[n]$ telle que chacun des n^r r-uplets de n y figure une fois (et donc une seule). Nous représenterons chaque séquence circulaire A comme un mot simple en le coupant en un point arbitraire:

$$A = a_1\ a_2 \ldots\ a_m,\ a_i \in [n],\ 1 \leqslant i \leqslant m = n^r,$$

bien que cette représentation ne soit pas univoque.

Par exemple le mot 11121222 est un mot circulaire équilibré de type (2,3).

Les mots circulaires équilibrés de type (n,r) ont été dénombrés par de Bruijn.

Il y en a $n^{-r}\ (n!)^{n^{r-1}}$. Ils ont de nombreuses applications, en psychologie expérimentale notamment : à l'ensemble $[n]$ correspond un ensemble de stimuli et les mots circulaires équilibrés coupés en un point arbitraire servent à définir des séquences de présentation des stimuli (cf.Barbut 1966 a,b,c et Durup 1967).

⁎ Les auteurs ont bénéficié pour cette recherche d'une aide du C.N.R.S., dans le cadre de l'A.T.P. "Informatique".

⁎⁎ Laboratoire d'Informatique pour les Sciences de l'Homme. C.N.R.S.

Pour éviter d'introduire des régularités systématiques dans la structure fine de ces séquences on a proposé d'utiliser des mots équilibrés aléatoires. Le problème de la construction de ces mots a été étudié par Bovet (1975) qui propose deux algorithmes. Pour chacun de ces algorithmes et pour n et r fixés les probabilités d'obtenir chacun des différents mots circulaires de type (n,r) ne sont pas égales (6). L'objet de cette note est de présenter un algorithme pour lequel ces probabilités sont égales autrement dit un algorithme de génération de mots circulaires aléatoires uniformément distribués.

Cet algorithme utilise la correspondance classique entre mots circulaires équilibrés et circuits eulériens.

MOTS CIRCULAIRES EQUILIBRES ET CIRCUITS EULERIENS

Soit S l'ensemble des (r-1) uplets de $[n]$:

$$S = \{(a_1\ a_2 \cdots\ a_{r-1}) / a_k \in [n], k = 1, \ldots\ r-1\}$$

et soit G le graphe dont S est l'ensemble des sommets et dans lequel deux sommets $s = (a_1\ a_2 \cdots\ a_{r-1})$ et $s' = (a'_1\ a'_2 \cdots\ a'_{r-1})$ sont reliés par un arc orienté de s vers s' si et seulement si on a $a_{i+1} = a'_i$ pour $1 \leqslant i \leqslant r-2$:

$$(s,s') \in G \iff a_2\ a_3 \cdots\ a_{r-1} = a'_1\ a'_2 \cdots\ a'_{r-2}.$$

Un circuit eulérien de G est un circuit qui parcourt une fois et une seule chaque arc de G. Ces circuits eulériens sont en correspondance biunivoque avec les mots circulaires équilibrés de type (n,r). En effet on peut associer à chaque circuit eulérien de G le mot circulaire formé par les dernières lettres de chacun des r-1 uplets, sommets du circuit, disposées dans l'ordre dans lequel ces sommets sont parcourus. Réciproquement un mot circulaire équilibré de type (n,r) définit le circuit eulérien de G dont les sommets sont les r-1 uplets qui en sont facteurs.

Il existe aussi une correspondance entre l'ensemble des circuits eulériens de G et l'ensemble des arborescences ascendantes (arborescences dont les arcs sont orientés vers la racine) de sommet fixé s_o qui recouvrent G. Définissons d'abord un codage des circuits eulériens de G. Choisissons un sommet s_1 tel que (s_o, s_1) soit un arc de G. A chaque circuit eulérien C de G faisons correspondre la suite de ses sommets dans l'ordre dans lequel ils apparaissent dans ce circuit et en choisissant de commencer par s_o puis s_1 ($(s_o, s_1) \in C$ puisque C est eulérien)

$$C = (s_o, s_1, s_2, \ldots)$$

Ayant choisi s_o et s_1 cette représentation est évidemment unique. Pour tout sommet r distinct de s_o appelons arc de sortie de C le dernier arc du circuit issu de r. On vérifie alors que l'ensemble des arcs de sortie forment une ar-

borescence ascendante de sommet s_o. Réciproquement à chaque arborescence ascendante de sommet s_o on peut associer tous les circuits eulériens dont l'ensemble des arcs de sortie coïncide avec l'ensemble des arcs de cette arborescence. Il y en a $(n-1)!^{n^{r-1}}$, puisque les arcs de sorties étant fixés il y a $(n-1)!$ choix possibles pour les ordres dans lequels apparaissent dans la représentation considérée les autres arcs issus de chaque sommet de G.

Cette correspondance entre mots circulaires équilibrés et circuits eulériens est décrite dans Berge (1958).

UNE PROCEDURE DE CONSTRUCTION DE MOTS CIRCULAIRES EQUILIBRES ALEATOIRES.

Cette correspondance suggère de procèder en deux étapes pour tirer au hasard un circuit eulérien de G (et le mot circulaire équilibré associé)

1 - tirer au hasard une arborescence T de racine s_o fixée

2 - tirer au hasard un circuit eulérien parmi ceux associés à l'arborescence T obtenue.

En effet cette procédure n'est pas biaisée puisque le nombre de circuits eulériens associés à chaque arborescence de racine s_o est constant.

L'étape 2 est triviale. Il est possible de réaliser 1 directement en tirant au hasard successivement les arcs de T mais cela nécessite d'introduire les probabilités conditionnelles dont l'expression est compliquée. La méthode indirecte suivante est plus performante.

1'- Tirer au hasard un graphe partiel de G de demi degré extérieur égal à 1 en $S-s_o$, à 0 en s_o. Effectuer ce tirage autant de fois qu'il est nécessaire pour que le graphe partiel obtenu soit une arborescence.

1' - est simple à exécuter puisqu'il suffit de tirer au hasard en chaque sommet distinct de s_o un arc parmi les arcs incidents à ce sommet et de tester la présence de cycles dans le graphe obtenu.

La probabilité P d'obtenir une arborescence donnée dans un tirage est égale au produit des probabilités d'obtenir en chaque sommet distinct de s_o l'arc incident à ce sommet appartenant à cette arborescence. Ainsi

$$P = \left[\frac{1}{n}\right]^{n^{r-1}-1} = c^{te}$$

ce qui justifie la procédure proposée.

Notons que la procédure est également valable pour tirer au hasard une arborescence de racine s_o recouvrant les sommets d'un graphe quel-

conque H. On a alors

$$P = \prod \frac{1}{d_i}$$

où les d_i désignent les demi degrés extérieurs des sommets de H autres que s_o.

Calculs du nombre moyen N d'itérations nécessaires.

On obtient ce nombre en effectuant le quotient du nombre de graphes partiels de G de demi degré extérieur égal à 1 sur $S-s_o$, à 0 en s_o, soit $n^{n^{r-1}-1}$, au nombre d'arborescences de racine S_o recouvrant G, soit $n^{n^{r-1}-r}$ (De Bruijn 1946). Ainsi

$$N = \frac{n^{n^{r-1}-1}}{n^{n^{r-1}-r}} = n^{r-1} = \frac{L}{n}$$

où L désigne la longueur des mots circulaires cherchés. Ainsi le nombre moyen d'itérations est d'un ordre de grandeur tout-à-fait acceptable, et donne lieu à un algorithme de complexité polynominale en L.

Description du programme.

Un programme Fortran d'une centaine d'instructions a été écrit, qui réalise successivement les procédures 1' et 2. Pour tester si un graphe est une arborescence on efface tout arc dont le sommet initial n'est sommet terminal d'aucun autre arc. T est une arborescence si et seulement si l'application répétée de cette procédure permet de supprimer tous les arcs de T.

En ce qui concerne l'étape 2 on obtient aisément une permutation aléatoire de degré m en ordonnant m nombres tirés au hasard entre 0 et 1.

Le nombre moyen d'opérations exécutées par le programme pour construire un mot circulaire équilibré aléatoire de type (n,r) est de l'ordre de $n^3\ r^{2n-2}$.

BIBLIOGRAPHIE

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

2.BARBUT M., "Mots circulaires et équilibrés", Mat. et Sc. Hum., 16, (1966b).

3.BARBUT M., "Mots circulaires et équilibrés. Histoire du problème vu à travers

la bibliographie" Math. et Sc. Hum. 17, (1966c).

4.BERGE, La théorie des graphes et ses applications ,Dunod,1958.

5.BOVET, "Generation automatique de mots circulaires et équilibrés" Math. et Sc. Hum. 49, 29-42 (1975).

6.BOVET, Communication verbale.

7.DE BRUIJN, "A Combinational Problem", Indag. Math.8, 416-417 (1946).

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