Les mots de Kolakoski

La suite de Kolakoski contient-elle, en moyenne, autant de 1 que de 2 ?

Un **mot** est une succession, finie ou non, de symboles (généralement appelé lettres) pris dans un certain alphabet :

a a b a b b b c d d a a a b . . .

En repérant les répétitions, on peut découper un mot en **blocs** :

aa b a bbb c dd aaa b …

En écrivant bout à bout les longueurs de ces blocs, on obtient un nouveau mot formé de nombres :

2 1 1 3 1 2 3 ...

Ce deuxième mot peut être appelé la **lecture** du premier.

Si les symboles de l'alphabet représentent des nombres, il se peut que **certains mots coïncident avec leur lecture**. Kolakoski a donné en 1966 le premier exemple de ces "self reading words", avec l'alphabet {**1, 2**} :

Déterminons un tel mot en supposant qu'il commence par le chiffre **2**. Ce "**2**" en lecture signifie que le mot doit commencer par un bloc de 2 symboles identiques ; le début du mot est donc, nécessairement :

2 2 1

le second chiffre **2** indique la présence d'un bloc de 2 symboles qui succède au premier bloc ; après le bloc **22** le mot se poursuit donc par **112** Nous voici donc avec le début :

2 2 1 1 2

Le troisième symbole de cette suite indique que le troisième bloc du mot est de longueur 1, ce qui permet de trouver le symbole suivant. Le procédé de construction se poursuit ainsi, indéfiniment :

2 2 1 1 2 1 2 2 1 2 2 1 1 ...

Cette suite, illimitée, est bien formée de **2** chiffres 2, **2** chiffres 1, **1** chiffre 2, **1** chiffre 1, **2** chiffres 2, **1** chiffre 1, **2** chiffre 2, **2** chiffres 1, etc.

Sa lecture est donc bien

2 2 1 1 2 1 2 2 ...

c'est à dire ... elle-même ! Le mot infini obtenu est connu sous le nom de **suite de Kolakoski**. Plus généralement, en l'honneur de l'inventeur, nous appelons **mot de Kolakoski** tout mot qui est sa propre lecture.

Le sujet

Malgré la simplicité de sa fabrication (que nous vous invitons à justifier plus rigoureusement par vous même), la suite de Kolakoski reste bien mystérieuse.

On se demande en particulier si les chiffres 1 et 2 apparaissent aussi souvent l'un que l'autre dans cette suite.

La suite de Kolakoski contient-elle, en moyenne, autant de 1 que de 2 ?

La question se généralise aux mots de Kolakoski :

Quels sont ces étranges mots qui se lisent eux-mêmes ?

Dans quelles proportions les symboles y figurent-ils ?

Quelques entrées possibles

On peut construire un autre mot de Kolakoski sur le même alphabet {**1,2**} , et, plus généralement, des mots de Kolakoski sur d'autres alphabets. Il serait intéressant de pouvoir comparer ces mots et les classer.

On peut chercher et étudier différentes façons de fabriquer la suite de Kolakoski ou d'autres mots d'autres mots de Kolakoski. Peut-on prévoir le n-ième symbole ?

On peut recenser diverses propriétés de la suite de Kolakovski. Quels sont les mots **finis** qui apparaissent comme morceaux de cette suite ? Quels sont ceux qui apparaissent une fois ? Deux fois ? ... Un nombre de fois déterminé, k ? ... Une infinité de fois ? **Exemple** : le mot **2 1 2** , de longueur 3, apparait deux fois, consécutivement, dans les premiers termes. Y apparaît-il une infinité de fois (non nécessairement consécutives) ? Y apparaît-il 3 fois consécutivement ? ...