Ensembles inévitables, mots de Lyndon et graphes de De Bruijn

(avec Guoniu Han et Eduardo Moreno)

Ensemble inévitable: tout mot assez long a un facteur dedans.

$s_n(k)$ = (resp. $q_n(k)$) le nombre de classes de conjugués de mots (resp. de mots primitifs) de longueur n sur un alphabet à k lettres.

$$s_n(k) = \sum_{d|n} q_n(k).$$

$$q_n(k) = \frac{1}{n} \sum_{d|n} \mu(n/d) k^d$$

où μ désigne la fonction de Möbius, et aussi

$$s_n(k) = \frac{1}{n} \sum_{d|n} \varphi(n/d) k^d$$

où φ désigne la fonction d'Euler. Ceci permet un calcul direct des s_n.

Pour $k = 2$,

n	1	2	3	4	5	6	7	8	9	10	11	12	13
$q_n(2)$	2	1	2	3	6	9	18	30	56	99	186	335	630
$s_n(2)$	2	3	4	6	8	14	20	36	60	108	188	352	632

Pour $k = 3$, on obtient

n	1	2	3	4	5	6	7	8	9	10	11	12
$q_n(3)$	3	3	8	18	48	116	312	810	2184	5880	16104	44220
$s_n(3)$	3	6	11	24	51	130	315	834	2195	5934	16107	44368

Un ensemble inévitable de mots de longueur n sur un alphabet à k lettres rencontre chaque classe de conjugués de mots de longueur n. Il a donc au moins $s_n(k)$ éléments.

Fait: Pour $n \leq 9$ il existe un ensemble inévitable de mots binaires de longueur n avec s_n éléments. Ceci plaide en faveur de la conjecture suivante.
Conjecture Pour tout entier $n \geq 1$, il existe un ensemble inévitable de $s_n(k)$ mots de longueur n sur k lettres.

Nous présentons ci-dessous les différents cas pour $n \leq 7$.

Les cas $n \leq 4$

Pour $n = 1$, il y a un seul ensemble inévitable qui est l'alphabet lui-même. Pour $n = 2$, sur $A = \{a_0, a_1, \ldots, a_{k-1}\}$, on a un ensemble inévitable de $s_2(k) = k(k+1)/2$ éléments

$$X = \{a_i a_j \mid i \leq j\}.$$

Pour $n = 3, 4$, on a $s_3(k) = (k^3 + 2k)/3$ et $s_4(k) = (k^4 + k^2 + 2k)/4$.

Sur deux lettres, on a les ensembles suivants. Chaque mot est donné en regard de la puissance de mot de Lyndon correspondante.

	Lyndon	Inévitable
0	aaa	aaa
1	aab	aba
2	abb	abb
3	bbb	bbb

	Lyndon	Inévitable
0	$aaaa$	$aaaa$
1	$aaab$	$aaab$
2	$aabb$	$baab$
3	$abab$	$baba$
4	$abbb$	$bbab$
5	$bbbb$	$bbbb$

Le nombre d'ensembles inévitables est donné par le tableau ci-dessous.

n	1	2	3	4
	1	2	4	30

Pour $n = 3$ et $k = 3, 4$, on a les ensembles

inévitables suivants.

	Lyndon	Inévitable
0	aaa	aaa
1	aab	aab
2	aac	caa
3	abb	bab
4	abc	cab
5	acb	acb
6	acc	cac
7	bbb	bbb
8	bbc	bcb
9	bcc	ccb
10	ccc	ccc

	Lyndon	Inévitable
0	aaa	aaa
1	aab	aba
2	aac	caa
3	aad	daa
4	abb	abb
5	abc	cab
6	abd	dab
7	acb	cba
8	acc	cac
9	acd	cda
10	adb	dba
11	adc	cad
12	add	dad
13	bbb	bbb
14	bbc	cbb
15	bbd	dbb
16	bcc	cbc
17	bcd	dbc
18	bdc	cbd
19	bdc	dbd
20	ccc	ccc
21	ccd	cdc
22	cdd	cdd
23	ddd	ddd

Pour $n = 4$, on trouve

	Lyndon	Inévitable
0	$aaaa$	$aaaa$
1	$aaab$	$aaab$
2	$aaac$	$acaa$
3	$aabb$	$baab$
4	$aabc$	$caab$
5	$aacb$	$acba$
6	$aacc$	$aacc$
7	$abab$	$baba$
8	$abac$	$acab$
9	$abbb$	$babb$
10	$abbc$	$bcab$
11	$abcb$	$bcba$
12	$abcc$	$ccab$
13	$acac$	$caca$
14	$acbb$	$cbba$
15	$acbc$	$acbc$
16	$accb$	$ccba$
17	$accc$	$cacc$
18	$bbbb$	$bbbb$
19	$bbbc$	$cbbb$
20	$bbcc$	$cbbc$
21	$bcbc$	$cbcb$
22	$bccc$	$cbcc$
23	$cccc$	$cccc$

Le cas $n = 5$

Pour $n = 5$, on a l'ensemble suivant, dû à Christopher Saker (communication personnelle, juin 2000).

	Lyndon	Inévitable
0	$aaaaa$	$aaaaa$
1	$aaaab$	$aaaab$
2	$aaabb$	$baaab$
3	$aabab$	$abaab$
4	$aabbb$	$bbaab$
5	$ababb$	$babab$
6	$abbbb$	$bbabb$
7	$bbbbb$	$bbbbb$

Il contredit l'affirmation donnée dans Lothaire , Exercise 6.2 page 99) suivant laquelle 9 mots sont nécéssaires. Il n'est jamais trop tard pour s'amender. Il y a en fait 28 ensembles inévitables pour $n = 5$.

Le cas $n = 6$

Pour $n = 6$, Christopher Saker (communication personnelle, septembre 2000,) a trouvé l'ensemble figurant dans la seconde colonne ci-dessous.

Une énumération systématique de tous les choix possibles de $6^9 \times 3^2 \times 2 \approx 10^8$ pour conjuguer les puissances de mots de Lyndon permet d'obtenir la troisième colonne après $\approx 3 \times 10^6$ itérations (quelques minutes).

Un tirage au hasard des 14 entiers donne la troisième colonne en moins de 10^3 essais (instantané).

	Lyndon	Saker	Enum	Random
0	*aaaaaa*	*aaaaaa*	*aaaaaa*	*aaaaaa*
1	*aaaaab*	*aaaaba*	*aaaaab*	*abaaaa*
2	*aaaabb*	*aaaabb*	*baaaab*	*abbaaa*
3	*aaabab*	*baaaba*	*abaaab*	*abaaab*
4	*aaabbb*	*baaabb*	*bbaaab*	*aabbba*
5	*aabaab*	*abaaba*	*abaaba*	*abaaba*
6	*aababb*	*bbaaba*	*aababb*	*abbaab*
7	*aabbab*	*abaabb*	*abaabb*	*bbabaa*
8	*aabbbb*	*bbaabb*	*bbaabb*	*abbbba*
9	*ababab*	*bababa*	*ababab*	*bababa*
10	*ababbb*	*bababb*	*abbbab*	*babbba*
11	*abbabb*	*bbabba*	*babbab*	*babbab*
12	*abbbbb*	*bbabbb*	*babbbb*	*abbbbb*
13	*bbbbbb*	*bbbbbb*	*bbbbbb*	*bbbbbb*

Le nombre total de systèmes de représentants des classes circulaires est $6^9 \times 3^2 \times 2 \approx 10^8$. Le nombre d'ensembles inévitables minimaux est 68288. La densité des ensembles inévitables par rapport au nombre total de systèmes de représentants des classes circulaires pour $n \leq 6$ est donnée ci-dessous.

n	1	2	3	4	5	6
	1	1	$0,4$	$0,2$	10^{-2}	6×10^{-3}

Le cas $n = 7$

Le nombre total de systèmes de représentants des classes circulaires est $7^{18} \approx 10^{15}$. Nous avons obtenu l'ensemble ci-dessous par une méthode mixte énumération/tirage au hasard. Le nombre de tirages porte sur $\approx 10^6$ ensembles (exécution en moins de une heure). L'énumération consiste à choisir les mots de petite période. Par exemple, le mot de période 2 doit être un conjugué du mot de Lyndon d'indice 10 ou 15 (avec une seule possibilité pour chacun). L'énumération pure peut tourner pendant plusieurs heures sans résultat.

	Lyndon	Unavoidable
0	$aaaaaaa$	$aaaaaaa$
1	$aaaaaab$	$aaabaaa$
2	$aaaaabb$	$aaabbaa$
3	$aaaabab$	$aababaa$
4	$aaaabbb$	$abbbaaa$
5	$aaabaab$	$aabaaba$
6	$aaababb$	$ababbaa$
7	$aaabbab$	$aabbaba$
8	$aaabbbb$	$bbbbaaa$
9	$aabaabb$	$baabbaa$
10	$aababab$	$abababa$
11	$aababbb$	$abbbaab$
12	$aabbabb$	$abbabba$
13	$aabbbab$	$abbbaba$
14	$aabbbbb$	$bbbbaab$
15	$ababbb$	$babbaba$
16	$abababb$	$bbbbaba$
17	$abbabbb$	$bbbabba$
18	$abbbbbb$	$bbbabbb$
19	$bbbbbbb$	$bbbbbbb$

Les cas $n = 8, 9$

Pour $n = 8, 9$ sur deux lettres, des ensembles inévitables minimaux (respectivement 36 et 60 mots) ont été trouvés par Guoniu Han.

Systèmes de type fini

Soit S un système de type fini pour la définition de ce terme). On note $s_n = s_n(S)$ le nombre d'orbites de points de période n dans S (aussi appelées *colliers* de longueur n). On note $p_n = p_n(S)$ le nombre de points de période n et $q_n = q_n(S)$ le nombre d'orbites de période exatement n. On pose

$$\zeta(S) = \exp \sum_{n\geq 1} \frac{p_n}{n} z^n = \frac{1}{1 - \Sigma_{n\geq 1}\, u_n z^n}.$$

On a alors pour tout $n \geq 1$

$$p_n = nu_n + \sum_{1\leq i\leq n-1} p_i u_{n-i}$$

et

$$p_n = \sum_{d|n} d q_d.$$

Comme $s_n = \Sigma_{d|n}\, q_d$ ces formules permettent de calculer s_n en fonction des p_n et donc des u_n. De plus, on a

$$s_n = \frac{1}{n} \sum_{d|n} \varphi(n/d) p_d$$

où φ désigne la fonction d'Euler. Ceci permet un calcul direct des s_n.

Si S est une partie de A^Z et X une partie de A^*, on dit que X est inévitable dans S si tout élément de S a un facteur dans X.

Proposition Soit S un système de type fini sur l'alphabet A et $n \geq 0$ un entier. Si X est un ensemble inévitable de mots de longueur n, il contient au moins un représentant de chaque collier de période d pour $d|n$. On a donc $Card(X) \geq s_n(S)$.

Démonstration. Soit $s = w^\zeta$ un mot de période n de S. Comme X est inévitable, il existe un facteur de s qui est dans X. Ainsi, X contient un conjugué de w.

Probleme Soit S un systeme de type fini. Pour quels entiers n existe-t-il un ensemble inévitable de $s_n(S)$ mots de longueur n?

La réponse n'est pas toujours positive. En effet, si $S = (ab)^\zeta$, un tel ensemble n'existe que pour n pair (on a $s_n = 0$ pour n impair).

Pas de bb

Si S est le système S_{bb} des mots qui évitent bb sur deux lettres, on a

$$\zeta(S) = \frac{1}{1 - z - z^2}.$$

On en déduit les valeurs suivantes.

n	1	2	3	4	5	6	7	8	9	10	11	12	13
p_n	1	3	4	7	11	18	29	47	76	123	199	322	521
q_n	1	1	1	1	2	2	4	5	8	11	18	25	40
s_n	1	2	2	3	3	5	5	8	10	15	19	31	41

Pour S_{bb}, on a les ensembles inévitables suivants pour $n \leq 9$.

n	1	2	3	4	5	6	7	8	9
	a	aa	aaa	$aaaa$	$aaaaa$	$aaaaaa$	$aaaaaaa$	$aaaaaaaa$	$aaaaaa$
		ab	aba	$aaba$	$aabaa$	$aaaaba$	$aaabaaa$	$aaaabaaa$	$aaaaba$
				$baba$	$ababa$	$abaaab$	$aababaa$	$ababaaaa$	$aababa$
						$abaaba$	$aabaaba$	$aabaabaa$	$aaaaba$
						$bababa$	$abababa$	$abaaabaa$	$aabaaa$
								$ababaaab$	$aababa$
								$ababaaba$	$aababa$
								$abababab$	$aababa$
									$baabaa$
									$ababa$

Il est curieux que pour le système S_{aa} sur trois lettres, on n'ait pas d'ensemble inévitable de 8 mots.

0.1 Pas de *bbb*

Pour S_{bbb}, on a

$$\zeta(z) = \frac{1}{1 - z - z^2 - z^3}.$$

On en déduit les valeurs suivantes.

n	1	2	3	4	5	6	7	8	9	10	11	12	13
p_n	1	3	7	11	21	39	71	131	241	443	2757		
q_n	1	1	2	2	4	5	10	15	26	42	74	121	212
s_n	1	2	3	4	5	9	11	19	29	48	75	132	213

On a pour $n \leq 7$ les ensembles inévitables minimaux suivants.

n	1	2	3	4	5	6	7	8	9
	a	aa	aaa	$aaaa$	$aaaaa$	$aaaaaa$	$aaaaaaa$		
		ab	aab	$aaba$	$aabaa$	$aaaaba$	$aaabaaa$		
			bab	$abba$	$aabba$	$abaaab$	$aababaa$		
				$abab$	$ababa$	$abaaba$	$aabaaba$		
					$babba$	$bababa$	$abababa$		
						$bbabba$	$abbabba$		
						$ababba$	$aababba$		
						$baabba$	$aabbaba$		
						$aaabba$	$aabbaaa$		
							$bababba$		
							$aabbaab$		

Autres systèmes

Le problème peut se présenter differemment pour des systèmes de type finis isomorphes. Ainsi, pour $S = (a + bc)^\varsigma$ qui est isomorphe a $(a + ba)^\varsigma$, on a tout d'abord pas d'ensemble inévitable de s_n mots de longueur n pour n impair. En effet, le mot $(bc)^\varsigma$ n'a pas de facteur répétable de longueur impaire. Les ensembles suivants sont inévitables pour $n = 2, 4, 6$.

n	2	4	6
	aa	$aaaa$	$aaaaaa$
	bc	$abca$	$bcbcbc$
		$bcbc$	$aabcaa$
			$abcbca$
			$bcabca$

Pour $n = 8$, il n'existe pas d'ensemble inévitable de 8 mots de longueur 8. En effet, le seul facteur répétable de longueur 8 de $(abc)^\varsigma$ est $w = bcabcabc$. Mais les seuls facteurs répétables de longueur 8 du mot $(abcbc)^\varsigma$ sont $abcbcabc$ et $bcabcbca$ qui sont conjugués de w.

Emboitements

On peut imaginer construire des ensembles inévitables pour le système plein en complétant progressivement des ensembles inévitables pour des systèmes de type fini plus petits.

Plus précisément, on peut construire un ensemble inévitable minimal $X = \{aaaa, baba, aaba, abba, bbba, bbbb\}$ pour $n = 4$ de la façon suivante. On considère la suite

$$\emptyset \subset S_b \subset S_{bb} \subset S_{bbb} \subset .$$

et aussi

$$X_3 = \{bbbb, bbba\} \subset X_2 = \{bbbb, bbba, abba\} \subset X_1 = \{bbbb, bbba, ab$$

On a

$$\begin{aligned} S_{bbb} &= S_{X_3} \\ S_{bb} &= S_{X_2} \\ S_b &= S_{X_1} \\ \emptyset &= S_X \end{aligned}$$

On peut compléter un ensemble inévitable du système S_{bb} en un ensemble inévitable du système S_{bbb} en construisant un ensemble inévitable pour les mots qui ont le facteur bb mais pas le facteur bbb.

Les résultats pour $n \leq 9$ sont représentés ci-dessous.

n	1	2	3	4	5	6	7	8
		bb	abb	$abba$	$abbaa$	$abbaaa$	$aaabbaa$	$abbaaaaa$
					$abbab$	$abbaab$	$aababba$	$abbaaaab$
						$abbaba$	$baabbaa$	$abbaaaba$
						$abbabb$	$abbabba$	$abbaaabb$
							$aabbaba$	$abbaabaa$
							$bababba$	$abbaabab$
								$abbaabba$
								$aaabbaba$
								$ababbaba$
								$abaabbab$
								$abbabbab$

Pour le système plein sur deux lettres $S = \{a, b\}^Z$, on parvient à construire pour chaque $n \leq 6$ un ensemble inévitable par emboitement des systèmes

S_{b^i} pour $i \leq n$.

n	1	2	3	4	5	6	7	8
	a	aa	aaa	$aaaa$	$aaaaa$	$aaaaaa$	$aaaaaaa$	
	b	ab	aba	$aaba$	$aabaa$	$aaaaba$	$aaabaaa$	
		bb	abb	$baba$	$ababa$	$abaaab$	$aababaa$	
			bbb	$abba$	$abbaa$	$abaaba$	$aabaaba$	
				$bbba$	$abbab$	$bababa$	$abababa$	
				$bbbb$	$abbba$	$aaaabb$	$aaabbaa$	
					$bbbba$	$bbaaab$	$aababba$	
					$bbbbb$	$babbaa$	$baabbaa$	
						$abaabb$	$abbabba$	
						$bbaabb$	$aabbaba$	
						$ababbb$	$bababba$	
						$babbab$	$abbbaaa$	
						$babbab$	$abbbaab$	
						$bbbabb$	$abbbaba$	
						$bbbbbb$	$abbbabb$	
							$abbbbaa$	
							$abbbbab$	
							$abbbbba$	
							$bbbbbba$	
							$bbbbbbb$	

Le théorème de Fredericksen et Maiorana

Théorème Si on concatène dans l'ordre alphabétique les mots de Lyndon de longueur divisant n, on obtient un cycle de De Bruijn.

0.01.1

0.001.011.1

0.0001.0011.01.0111.1

0.00001.00011.00101.00111.01111.1