Annals of Discrete Mathematics 6 (1980) 251–255

A NEW STATISTICS ON WORDS

A. LASCOUX and M.P. SCHÜTZENBERGER

Départment de Mathématiques, Paris VII, Tour 55, 2 Place Jussieu, 75005 Paris (France)

The study of various statistics on words is a chapter of combinatorics and it has been recently surveyed by Foata [1]. One ot these statistics is the so-called "Major Index" of a word. It plays a role in the theory of representations of the symmetric group. We present here a generalization, the "load", which we used to prove Foulkes' conjecture, i.e. to supply a q-analogue of Kostka numbers [2].

It turns out that the load can be defined in terms of a minimization problem. We develop this approach here without entering into the algebraïc background or applications. A special feature of the problem is that a solution is given by the most simple minded algorithm, a not too frequent occurrence. Therefore, we suspect that it is a very degenerate special case of a more general question already encountered in operations research. Since we have not been able to trace it in the literature, we apologize to those authors, whose articles may have been unwillingly overlooked.

1. The minimization problem

Throughout this paper consider two finite totally ordered alphabets,

$$A=\{a<b<c<\cdots\} \quad \text{and} \quad \Gamma=\{\alpha<\beta<\cdots\}.$$

As usual the number of occurrences of a letter x in a word w is denoted by $|w|_x$, and one says that a word is *standard* iff no letter occurs more than once in it.

We also consider a fixed word v in the free monoid A^*, and we make the basic assumption that its multidegree is non-increasing, i.e. that $|v|_a \geq |v|_b \geq \cdots$.

A *lifting* of v is a standard word w in the free monoid $(A\times\Gamma)^*$ such that, on the one hand, v is its projection on A^* and, on the other hand, its minimal alphabet, B, is an initial interval of $A\times\Gamma$, i.e. $c\gamma\in B$ only if $c'\gamma'\in B$ for any $c'\leq c$ and $\gamma'\leq\gamma$. Our basic assumption on the multidegree of v insures that v has at least one lifting (and, in fact it has only one iff v is standard). Note that B is the union over all greek letters γ of sets $A_\gamma\times\{\gamma\}$ where A_γ is an initial intervalle of A.

Let w be a lifting of v. For each $c\gamma\in B$ we say that the letter $c\gamma$ is *marked* iff it is on the *right* of the letter $c_-\gamma$ where c_- denotes the (immediate) predecessor of c

in A. Thus, a being always the first letter of A, we have that no letter $a\gamma$ is marked. We deduce by two summations the *delay function* $\mu = \mu_w$ and the *integrated delay function* $\bar{\mu}$ of the lifting w. The first one has domain B and $(c\gamma)\mu$ is defined as the number of marked $c'\gamma$ with $c' \leqslant c$. The second one has domain $A \times \Gamma$ and $(c\gamma)\bar{\mu}$ is the sum of $(c'\gamma')\mu$ over all $c' \leqslant c$ and $\gamma' \leqslant \gamma$.

The problem is to choose a lifting so as to minimise the maximum of the integrated delay function, i.e. the sum of the values of the delay function itself. This minimax quantity is the *load* of the word v.

Example. Let $v = a\,b\,d\,a\,b\,c\,b\,a\,c\,d$, of length $|v| = 10$. The partition of 10 expressing the multidegree is $K = (2\,2\,3\,3)$. One of the $2!\,2!\,3!\,3!$ possible liftings is:

$$a\gamma,\ b\beta,\ d\alpha,\ a\alpha,\ \underline{b}\gamma,\ c\alpha,\ a\beta,\ b\alpha,\ \underline{c}\beta,\ \underline{d}\beta.$$

Marked letters are underlined. The delay and integrated delay functions are:

$$\begin{array}{c|ccc}
d & 1 & 2 & \\
c & 1 & 1 & \\
b & 1 & 0 & 1 \\
a & 0 & 0 & 0 \\
\hline
 & \alpha & \beta & \gamma
\end{array}
\qquad
\begin{array}{c|ccc}
d & 3 & 6 & 7 \\
c & 2 & 3 & 4 \\
b & 1 & 1 & 2 \\
a & 0 & 0 & 0 \\
\hline
 & \alpha & \beta & \gamma
\end{array}$$

The natural lifting (discussed below) and corresponding functions are:

$$a\gamma,\ b\beta,\ d\alpha,\ a\beta,\ b\alpha,\ \underline{c\beta},\ a\alpha,\ \underline{b\gamma},\ \underline{c\alpha},\ \underline{d\beta}.$$

$$\begin{array}{c|ccc}
d & 1 & 2 & \\
c & 1 & 1 & \\
b & 0 & 0 & 1 \\
a & 0 & 0 & 0 \\
\hline
 & \alpha & \beta & \gamma
\end{array}
\qquad
\begin{array}{c|ccc}
d & 2 & 5 & 6 \\
c & 1 & 2 & 3 \\
b & 0 & 0 & 1 \\
a & 0 & 0 & 0 \\
\hline
 & \alpha & \beta & \gamma
\end{array}$$

We shall prove that 6 is indeed the load of v.

A case of special interest is when v is a standard word. Then the lifting and subsequent marking and functions are unique. We consider v as a permutation on the set of its letters and look at the inverse permutation v^{-1} on the places. For instance if $v = d\,a\,\underline{e}\,c\,f\,\underline{b}$, one has $v^{-1} = 2\,\underline{6}\,4\,1\,\underline{3}\,\underline{5}$. Marked letters correspond to numbers greater than their left neighbour. The load of v is by definition $0+1+1+1+2+3=8$. Up to reversal, this is the major index $5+2+1$ of v^{-1}, i.e. the sum of the lengths of the right segments of v^{-1} determined by the underlined numbers. No such connection between the major index and the load exists for non-standard words.

2. The main result

For any lifting w and $\gamma \in \Gamma$, we denote by w^{γ} the subword of w whose alphabet is $A_{\gamma} \times \{\gamma\}$. A straightforward attempt at a solution of our minimax problem consists in constructing successively these subwords for $\gamma = \alpha, \beta, \ldots$ by the following algorithm in which w is understood to be read from right to left.

Suppose that the subwords w^β have already been determined for $\beta < \gamma$. Then $a\gamma$ is the first occurrence of a (which has not yet been used) and, supposing that $c_-\gamma$ has been chosen, one takes for $c\gamma$ the first (non-used) occurrence of a c, on the left of $c_-\gamma$, if there is one. If none exists $c\gamma$ is the first occurrence of a (non used) c in w. This second possibility is the one which induces a marking of $c\gamma$.

The resulting lifting, marking and functions will be called *natural*. Our main result is the

Property 1. *The number of marked letters is minimal for the natural lifting; further, if $\bar{\nu}$ is the natural integrated delay function, one has $\bar{\nu} \leq \bar{\mu}$ for the integrated delay function $\bar{\mu}$ relative to any other lifting of v.*

Thus the load of v is given by the above algorithm.

3. A conjugacy lemma

We suppose here that $v = cv'$, where c is a letter which is not the first one of A and one of its liftings $w = c\gamma \cdot w'$. Let $v_1 = v'c$. It is clear that $w_1 = w' \cdot c\gamma$ is a lifting of v_1. The number of marked letters and delay function of w (resp. w_1) will be r and ν (resp. r_1 and ν_1).

Lemma 1. *The lifting w_1 is natural iff w is so. Further: $r_1 = r + 1$ or $= r$, depending upon whether c is or is not the last letter of A_γ; the delay functions ν and ν_1 are equal except for $(c\gamma)\nu = (c\gamma)\nu_1 + 1$.*

Proof. The truth of the first assertion follows immediately from the definitions. To check the other ones, note that the markings of w and w_1 are the same except, possibly for $c\gamma$ and $c_+\gamma$ (c_+ = the successor of c). In fact, because of $c \neq a$, the letter $c\gamma$ is marked in w_1 but not in w and if $c_+ \in A_\gamma$, i.e. if c is not the last letter of A_γ, the contrary is true of $c_+\gamma$.

Corollary 2. *Assume $v = v'a^m$ where $m = |v|_a$ and that Property 1 holds for v'. Then it holds for any word $v_1 = u''a^m u'$, where $u'u'' = v'$.*

Proof. The definitions apply to the word v' since it has a non-increasing multidegree on the alphabet $A \backslash a$.

The natural lifting of v is deduced from that of v' by left multiplication by the word $\cdots a\beta \cdot a\alpha$. Since this does not change the marking, Property 1 also holds for v. Thus, by induction on the length of u', it suffices to check the result when u' is a single letter. But this follows from the lemma since $u' \neq a$ by hypothesis.

4. Proof of the main result

We recall that the *plactic monoid* is the quotient of the free monoid A^* by the least congruence $\equiv$ such that for any letters b, c, d, one has:

$$\text{if}\quad b<c<d,\quad \text{then}\quad cbd\equiv cdb \quad\text{and}\quad bdc\equiv dbc;$$
$$\text{if}\quad b<c,\quad \text{then}\quad cbb\equiv bcb \quad\text{and}\quad ccb\equiv cbc.$$

It has been studied by G. de B. Robinson, C. Schensted, D.E. Knuth and Curtis Greene (Cf. [3]). We shall only require the fact that all the words in a congruence class have the same multidegree and that one of them has the form $u''a^m u'$ with a not occurring in u'' or u'.

Lemma 3. *If v and v_1 are congruent there corresponds to any lifting of v a lifting of v_1 having the same marked letters.*

Proof. Consider first the case when $v=v'bdv''$ and $v_1=v'dbv''$ where $b<d\neq b_+$. It is clear that $w'\cdot b\gamma\cdot d\gamma'\cdot w''$ is a lifting of v iff $w'\cdot d\gamma'\cdot b\gamma\cdot w''$ is a lifting of v_1 and that they have the same marked letters. This case covers the first elementary congruence defining $\equiv$. Suppose now that $v=v'cbbv''$ and $v_1=v'=v'bcbv''$ where $c=b_+$. Let $w=w'\cdot c\gamma\cdot b\gamma'\cdot b\gamma''\cdot w''$ be a lifting of v, where γ, γ' and γ'' are any three greek letters. If $\gamma\neq\gamma'$ it is clear that $w'\cdot b\gamma'\cdot c\gamma\cdot b\gamma''\cdot w''$ is a lifting of v_1 having the same marked letters. If $\gamma=\gamma'$ we have $\gamma\neq\gamma''$ and we can begin by replacing w_1 by the lifting $w_1'=w'\cdot c\gamma\cdot b\gamma''\cdot b\gamma\cdot w''$ of v without changing the marking. A similar argument applies to the other cases needed to cover the second elementary defining congruences, and the result is proved.

To conclude the proof of Property 1 we proceed by induction on the number of letters of the alphabet, the initial case being trivial.

Take any word $v_1=u''a^m u'$ with $|u''u'|_a=0$ in the congruence class of v. By the induction hypothesis the property holds for the word $u'u''$ which is on the smaller alphabet $A\backslash a$. Thus, by Corollary 3, it holds for v_1, i.e. one has the inequality $\bar\nu_1\leq\bar\mu_1$ between any integrated delay function $\bar\mu_1$ and the natural one $\bar\nu_1$. Applying Lemma 3 gives the result, and a similar argument shows that the natural lifting has a minimal number of marked letters.

5. A further property of the natural delay function

We consider the natural lifting w of the word v, and two consecutive greek letters β and $\gamma=\beta_+$. We let M be the set of marked letters.

Lemma 4. *Assume that $d\beta\in M$ and $d\gamma\notin M$. There exists a letter $b<d$ such that $b\beta\notin M$ and $b\gamma\in M$ and that $c\beta, c\gamma\notin M$ for any intermediate c, i.e. $b<c<d$.*

Proof. Because of the algorithm defining the natural lifting we can assume for simplicity, that β is the first letter of the greek alphabet. The hypothesis $d\beta \in M$ implies that all the letters d are to the right of $d_-\beta$ and $d\beta$ is the first of them. Thus w has the subword $d_-\beta \cdot d\gamma \cdot d\beta$. Now if $d\gamma$ is not marked, $d_-\gamma$ must be to the right of $d\gamma$, and we have shown that w contains the subword $d_-\beta \cdot d_-\gamma$.

To conclude the proof we need only to verify, by induction on the number of intermediate letters, that if $c\beta \cdot c\gamma$ is a subword of w, then $c\beta$ is not marked and, either $c\gamma$ is marked or $c_-\beta \cdot c_-\gamma$ is again a subword of w. Indeed the hypothesis that $c\beta$ is to the left of $c\gamma$ prevents that it be marked, i.e. that $c_-\beta$ be to its left. Thus w contains the subword $c\beta \cdot c_-\beta \cdot c\gamma$. Suppose further that $c\gamma$ is not marked. Then $c_-\gamma$ must be on its right and the argument is concluded.

Corollary 5. *The natural delay function ν is non-decreasing in each of its two arguments.*

Proof. By definition, one has $(b\beta)\nu \leqslant (c\beta)\nu$ for $b < c$ and any β, and the inequality $(c\beta)\nu \leqslant (c\gamma)\nu$ follows from the above lemma by induction in the alphabet, since $(\mathrm{a}\beta)\nu = 0$ identically.

References

[1] D. Foata, Distributions eulériennes et mahonniennes, in: Proc. Advanced School on Higher Combinatorics. (M. Aigner Editor, Berlin, 1978).

[2] A. Lascoux and M.P. Schützenberger, Sur une conjecture de H.O. Foulkes, C.R. Acad. Sci. Paris 286 (1978) 323–324.

[3] G.P. Thomas, On Schensted's construction and the multiplication of Schur functions, Adv. in Math. 30 (1978) 8–32.