VISIBILITY PROBLEMS RELATED TO SKIP LISTS

KAMILLA OLIVER AND HELMUT PRODINGER

ABSTRACT. For sequences (words) of geometric random variables, visibility problems related to a sun in north-west are considered. This leads to a skew version of such words. Various parameters are analyzed, like left-to-right maxima, descents and inversions.

1. INTRODUCTION

Assume that X is a geometrically distributed random variable, $\Phi\{X = k\} = pq^{k-1}$, with $p + q = 1$, and a word $x = a_1 a_2 \dots a_n$ of n independent outcomes of such a variable is given. It is typically displayed as in Figure 1, with $n = 14$, and the word is 31552252341111. Assume that there is a sun standing straight in north-west. Then certain nodes are lit, and others are not, whence the two types of nodes in Figure 3.

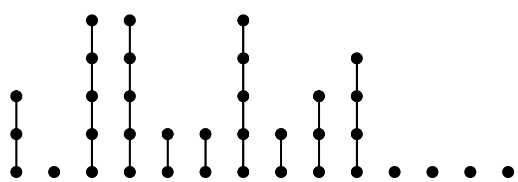

FIGURE 1. A word of length 13

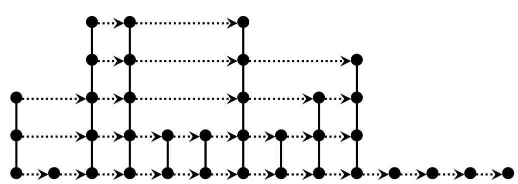

FIGURE 2. The same word, now with horizontal pointers akin to the skip-list structure

Date: September 23, 2017.
1991 *Mathematics Subject Classification.* 11B39.
Key words and phrases. Skip lists, geometric random variable, visibility, q-series.

Such a scenario was recently studied in [2] in the context of bargraphs. Further papers of Mansour and some of his team members about visibility questions are [4]; compare also [3].

A graphical depiction of geometrically distributed words (a combinatorial class that was extensively studied in the past) as in Figure 1 stems from a data structure called skiplist. It has also horizontal pointers, and they are related to a visibility problem, since the pointers are interrupted as indicated in Figure 2.

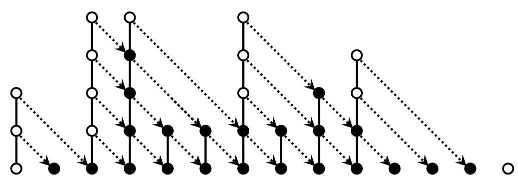

FIGURE 3. The sun stands in north-west, and some nodes are lit, others aren't.

Motivated by the recent paper [2], we study a sun standing in north-west and the number of lit nodes. The example in Figure 3 makes this very clear, 12 nodes are lit.

A moments reflexion tells us that the *skew* word $x^* = b_1 b_2 \dots b_n$ with $b_i = a_i + i - 1$ is relevant here. In our running example this is $x^* = 3\,2\,7\,8\,6\,7\,11\,9\,11\,13\,11\,12\,13\,14$.

Let us indicate the left-to-right maxima in this skew word:
$x^* = \mathbf{3}\,2\,\mathbf{7}\,\mathbf{8}\,6\,7\,\mathbf{11}\,9\,11\,\mathbf{13}\,11\,12\,13\,\mathbf{14}$. The differences of two consecutive such records are $3, 4, 1, 3, 2, 1$ (for this, we patched the word with a leftmost 0). The sum of these numbers is $14 = 3+4+1+3+2+1$, which, by telescoping, is also the largest value (the maximum) that occurs in the the skew word.

We are thus led to study in this paper the *maximum* of a random skew word of lenght n and the number of left-to-right maxima (which is 6 is the running example). The modification of the words that we call *skew* in this paper unfortunately does not allow us the elegant method of generating functions as in [5, 6].

We will take the opportunity and treat a few other combinatorial questions related to skew words as well, such as descents and inversions. Again, these questions are somewhat harder to deal with related to the non-skew (classical) versions.

We need some basic notation from q-calculus [1]: $(x)_n := (1-x)(1-xq)\dots(1-xq^{n-1})$ for $n \ge 0$ or $n = \infty$, as well as Cauchy's identity

(q-binomial theorem)

$$\sum_{n\ge 0}\frac{(a)_n}{(q)_n}t^n=\frac{(at)_\infty}{(t)_\infty}.$$

2. THE MAXIMUM OF RANDOM SKEW GEOMETRICALLY DISTRIBUTED WORDS

Let $\mathscr{M}_n$ be the maximum of a skew word. Its expectation can be computed as follows:

$$\mathbb{P}(\mathscr{M}_n\le k)=(1-q^k)(1-q^k)\dots(1-q^{k-n+1})=\frac{(q)_k}{(q)_{k-n}};$$

for $k\ge n$, otherwise it is zero.

Consequently

$$\begin{aligned}
\mathbb{E}(\mathscr{M}_n)&=n+\sum_{k\ge 0}\left[1-\frac{(q)_{n+k}}{(q)_k}\right]=n+\lim_{t\to 1}\sum_{k\ge 0}\left[t^k-\frac{(q)_n(q^{n+1})_k}{(q)_k}t^k\right]\\
&=n+\lim_{t\to 1}\left[\frac{1}{1-t}-(q)_n\frac{(q^{n+1}t)_\infty}{(t)_\infty}\right]\\
&=n+\lim_{t\to 1}\left[\frac{1}{1-t}-\frac{(q)_n(q^{n+1}t)_\infty}{(1-t)(qt)_\infty}\right]\\
&=n+(q)_n\frac{d}{dt}\frac{(q^{n+1}t)_\infty}{(qt)_\infty}\bigg|_{t=1}=n+(q)_n\frac{d}{dt}\prod_{k\ge 1}\frac{1-q^{n+k}t}{1-q^kt}\bigg|_{t=1}\\
&=n+(q)_n\frac{d}{dt}\prod_{k=1}^{n}\frac{1}{1-q^kt}\bigg|_{t=1}=n-\frac{1}{(q)_n}\frac{d}{dt}\prod_{k=1}^{n}(1-q^kt)\Big|_{t=1}\\
&=n+\sum_{k=1}^{n}\frac{q^k}{1-q^k}.
\end{aligned}$$

The sum is a q-analogue of a harmonic number, and it is customary to denote the limit by

$$\alpha_q:=\sum_{k\ge 1}\frac{q^k}{1-q^k}.$$

Of course,

$$H_n(q)=\sum_{k=1}^{n}\frac{q^k}{1-q^k}=\alpha_q+O(q^n).$$

Further,

$$\mathbb{E}(\mathscr{M}_n^2)=\sum_{k=0}^{n-1}(2k+1)+\sum_{k\ge n}\left[1-\frac{(q)_k}{(q)_{k-n}}\right](2k+1)$$

$$
\begin{aligned}
&= n^2 + \sum_{k\ge 0}\Big[1-\frac{(q)_{n+k}}{(q)_k}\Big](2k+2n+1)\\
&= n^2 + (2n+1)H_n(q) + 2\sum_{k\ge 0}\Big[1-\frac{(q)_{n+k}}{(q)_k}\Big]k\\
&= n^2 + (2n+1)H_n(q) + 2\lim_{t\to 1}\Big[\frac{t}{(1-t)^2}-(q)_n\sum_{k\ge 0}\frac{(q^{n+1})_k}{(q)_k}kt^k\Big]\\
&= n^2 + (2n+1)H_n(q) + 2\lim_{t\to 1}\Big[\frac{t}{(1-t)^2}-(q)_n t\frac{d}{dt}\frac{(q^{n+1}t)_\infty}{(t)_\infty}\Big]\\
&= n^2 + (2n+1)H_n(q) + 2\lim_{t\to 1} t\frac{d}{dt}\Big[\frac{1}{1-t}-(q)_n\frac{(q^{n+1}t)_\infty}{(1-t)(qt)_\infty}\Big]\\
&= n^2 + (2n+1)H_n(q) + (q)_n\frac{d^2}{dt^2}\frac{(q^{n+1}t)_\infty}{(qt)_\infty}\Big|_{t=1}.
\end{aligned}
$$

We compute the second derivate alone:

$$
\begin{aligned}
(q)_n\frac{d^2}{dt^2}\frac{(q^{n+1}t)_\infty}{(qt)_\infty}\Big|_{t=1} &= (q)_n\frac{d^2}{dt^2}\prod_{k\ge 1}\frac{1-q^{n+k}t}{1-q^kt}\Big|_{t=1}\\
&= (q)_n\frac{d^2}{dt^2}\prod_{k=1}^{n}\frac{1}{1-q^kt}\Big|_{t=1}\\
&= \frac{2}{(q)_n^2}\Big(\frac{d}{dt}\prod_{k=1}^{n}(1-q^kt)\Big|_{t=1}\Big)^2-\frac{1}{(q)_n}\frac{d^2}{dt^2}\prod_{k=1}^{n}(1-q^kt)\Big|_{t=1}\\
&= 2H_n^2(q)-2\sum_{1\le i<j\le n}\frac{q^i}{1-q^i}\frac{q^j}{1-q^j}\\
&= 2H_n^2(q)-H_n^2(q)+H_n^{(2)}(q)=H_n^2(q)+H_n^{(2)}(q),
\end{aligned}
$$

with a q-analogue of a harmonic number of second order

$$H_n^{(2)}(q)=\sum_{k=1}^{n}\Big(\frac{q^k}{1-q^k}\Big)^2.$$

Summarizing,

$$\mathbb{E}(\mathscr{M}_n^2)=n^2+(2n+1)H_n(q)+H_n^2(q)+H_n^{(2)}(q).$$

Therefore we have the variance:

$$
\begin{aligned}
\mathbb{V}(\mathscr{M}_n) &= \mathbb{E}(\mathscr{M}_n^2)-\mathbb{E}^2(\mathscr{M}_n)\\
&= n^2+(2n+1)H_n(q)+H_n^2(q)+H_n^{(2)}(q)-(n+H_n(q))^2\\
&= H_n(q)+H_n^{(2)}(q).
\end{aligned}
$$

Theorem 1. *The expected value and the variance of the parameter $\mathscr{M}_n$ of a random skew geometrically distributed word of length n, are given by*

$$\mathbb{E}(\mathscr{M}_n) = n + H_n(q),$$
$$\mathbb{V}(\mathscr{M}_n) = H_n(q) + H_n^{(2)}(q).$$

3. Left-to-right maxima

Now we want to study the number of (strict) left-to-right maxima of the skew word x^*. As a preparation, let $\mathscr{Y}_m$ be the indicator variable of the event "$a_m + m - 1$ is a left-to-right maximum in the skew word x^*."

For the standard case, such computations appear in [5, 6]. However, as explained in the Introduction, this is more challenging here, and we managed only to get the expected value.

The expected value is computed as follows:

$$\begin{aligned}\mathbb{E}(\mathscr{Y}_m) &= \sum_{j\ge1} pq^{j-1}(1-q^{j+m-2})\dots(1-q^j)\\ &= p\sum_{j\ge0} q^j \frac{(q)_{m-1+j}}{(q)_j} = p(q)_{m-1}\sum_{j\ge0} q^j \frac{(q^m)_j}{(q)_j}\\ &= p(q)_{m-1}\frac{(q^{m+1})_\infty}{(q)_\infty} = \frac{p}{1-q^m}.\end{aligned}$$

Consequently, the expected value of the number of left-to-right maxima is

$$\mathbb{E}(\mathscr{Y}_1 + \cdots + \mathscr{Y}_n) = p\sum_{j=1}^{n} \frac{1}{1-q^j} = pn + pH_n(q) = pn + p\alpha + O(q^n).$$

4. Descents and inversions

First, we want to count the number of pairs, such that $a_i + i - 1 > a_{i+1} + i$, which means $a_i > a_{i+1} + 1$. Let $\mathscr{D}_i$ be the corresponding indicator variable.

$$\mathbb{E}(\mathscr{D}_i) = \sum_{k\ge1} pq^{k-1} \sum_{j>k+1} pq^{j-1} = \frac{q^2}{1+q}.$$

Thus the expected value of the total number of descents is $(n-1)\frac{q^2}{q+1}$.

In a similar style, assume that $1 \le i < j \le n$ and let $\mathscr{D}_i$ be the corresponding indicator variable "$a_i + i - 1 > a_j + j - 1$." Then

$$\mathbb{E}(\mathscr{D}_{i,j}) = \sum_{k\ge1} pq^{k-1} \sum_{1\le h<\max\{1,k+i-j\}} pq^{h-1} = \frac{q^{1+j-i}}{1+q}.$$

The expected number of inversions is then

$$\mathbb{E}(\mathsf{inversions}) = \sum_{1\le i<j\le n} \mathbb{E}(\mathscr{D}_{i,j}) = \sum_{1\le i<j\le n} \frac{q^{1+j-i}}{1+q} = \frac{1}{1+q}\sum_{1\le i,h<n} q^{1+h}$$
$$= \frac{n-1}{1+q}\sum_{1\le h<n} q^{1+h} = \frac{(n-1)q^2(1-q^{n-1})}{1-q^2}.$$

For $q \to 1$, this expression tends to $\frac{(n-1)^2}{2}$.

References

[1] George E. Andrews. *The theory of partitions*. Addison-Wesley Publishing Co., Reading, Mass.-London-Amsterdam, 1976. Encyclopedia of Mathematics and its Applications, Vol. 2.

[2] M. Archibald, A. Blecher, C. Brennan, A. Knopfmacher, and T. Mansour. Shedding light on words. *Applicable Analysis and Discrete Math*, 11:216–231, 2017.

[3] Ligia L. Cristea and Helmut Prodinger. The visibility parameter for words and permutations. *Cent. Eur. J. Math.*, 11(2):283–295, 2013.

[4] Gregory Gutin, Toufik Mansour, and Simone Severini. A characterization of horizontal visibility graphs and combinatorics on words. *Phys. A*, 390(12):2421–2428, 2011.

[5] Helmut Prodinger. Combinatorics of geometrically distributed random variables: left-to-right maxima. In *Proceedings of the 5th Conference on Formal Power Series and Algebraic Combinatorics (Florence, 1993)*, volume 153, pages 253–270, 1996.

[6] Helmut Prodinger. On the m-enumeration of maximum and left-to-right maxima in geometrically distributed words. *Pure Math. Appl. (PU.M.A.)*, 23(2):143–154, 2012.

(Kamilla Oliver) 9 Stellenberg Road, 7130 Somerset West, South Africa
E-mail address: olikamilla@gmail.com

(Helmut Prodinger) Department of Mathematical Sciences, Stellenbosch University, 7602 Stellenbosch, South Africa
E-mail address: hproding@sun.ac.za