THE VISIBILITY PARAMETER FOR WORDS AND PERMUTATIONS

LIGIA L. CRISTEA* AND HELMUT PRODINGER**

Dedicated to Philippe Flajolet (1948–2011)

Abstract. In this paper we investigate the *visibility parameter*, i.e., the number of *visible pairs*, first for words over a finite alphabet, then for permutations of the finite set $\{1, 2, \ldots, n\}$, and finally for words over an infinite alphabet whose letters occur with geometric probabilities. The results obtained for permutations correct the formula for the expectation obtained in a recent paper by Gutin et al. [3], and for words over a finite alphabet the formula obtained in the present paper for the expectation is more precise than that obtained in the cited paper. More importantly, we also compute the variance for each case.

1. Introduction

Consider a word $w_1 \ldots w_n$ of length n where the letters are positive integers. The visibility parameter (=the number of visible pairs) is defined as

$$\mathsf{Vis}(w_1 \ldots w_n) = \#\{(j,k) \mid 1 \le j < k \le n,\ w_l > \max\{w_j, w_k\} \text{ for all } j < l < k\}.$$

Using indicator variables $\chi_{j,k}$, defined by

$$\chi_{j,k}(w_1 \ldots w_n) = \begin{cases} 1 & \text{if } w_l > \max\{w_j, w_k\} \text{ for all } j < l < k, \\ 0 & \text{otherwise,} \end{cases}$$

we can write

$$\mathsf{Vis} = \sum_{1 \le j < k \le n} \chi_{j,k}.$$

Gutin et al. [3] have investigated this parameter motivated by horizontal visibility graphs (HVG), which provide a method for studying time series by investigating graphs associated to them. In the mentioned paper, the authors give a necessary and sufficient condition for a graph to be a HVG and characterise subfamilies of HVGs by approaching ordered sets as words, thus combinatorics on words becomes a useful tool. The visibility parameter is introduced in this context and is thus natural as a combinatorial parameter.

Date: April 2, 2012.

1991 *Mathematics Subject Classification.* 05A05, 05A30.

Key words and phrases. Words, permutations, q-enumeration.

*This author is supported by the Austrian Research Foundation (FWF), Project P20412-N18.

*Parts of this work were done during a research stay of this author at the Mathematics Department of the Stellenbosch University.

**This author is supported by an incentive grant of the South African Science Foundation NRF.

We have changed the '$<$' in the definition of the visibility parameter given by Gutin et al. [3] to '$>$' since it is more convenient: For words over the alphabet $\{1, 2, \ldots, M\}$ where each word of length n is equally likely and for permutations (written as words $\pi_1 \ldots \pi_n$, with $\pi_i \in \{1, \ldots, n\}$), where each permutation appears with probability $1/n!$, it makes (statistically) no difference, but we also investigate the model of words with letters in $\{1, 2, 3, \ldots\}$, where the letter k appears (independently) with probability pq^{k-1}, and $p + q = 1$. This model is quite important in Computer Science. To justify this claim, we mention two areas: the *skip list* [7] and *probabilistic counting* [1].

Parameters similar to the visibility parameter *have* already appeared in the literature:

The first one is "Knuth's parameter a" (which might also be called left-sided path-length) [5]: it is defined as

$$\mathsf{a}(w_1 \ldots w_n) = \#\{(j, k) \mid 1 \le j < k \le n,\ w_j < w_l \text{ for all } j < l \le k\}.$$

The other one [6] is a q-analogue of the path-length in binary search trees:

$$\rho(w_1 \ldots w_n) = \#\{(j, k) \mid 1 \le j < k \le n,\\ w_j = \min\{w_j, \ldots, w_k\} \text{ or } w_k = \min\{w_j, \ldots, w_k\}\}.$$

For the case of a finite alphabet with M letters, Gutin et al. have obtained the average as $\mathbb{E}(\mathsf{Vis}) \sim 2n - \frac{H_M}{M}n$, for $n \to \infty$ and fixed M, with harmonic numbers

$$H_M = \sum_{1 \le k \le M} \frac{1}{k}.$$

We will obtain a more precise formula that also includes a constant and an exponentially small error term. More importantly, we also compute the variance; a precise statement follows later in the paper.

For permutations, Gutin et al. give the average as $\mathbb{E}(\mathsf{Vis}) \sim 2n - H_n$. Here, we correct the formula to $\mathbb{E}(\mathsf{Vis}) \sim 2n - 2H_n$ and also compute the variance.

For the words where the letters are equipped with geometric probabilities, we also compute expectation and variance. As it was explained in many papers, the limit $q \to 1$ reproduces the quantities for the instance of permutations (equal letters become impossible in the limit, and each relative ordering of the letters is equally likely in the limit). There are too many papers to be cited, but this is the first one in the series: [4].

We would like to emphasise that the computations for the expectations are quite simple in all instances, but that the computation of the variance is an arduous task that requires skills and patience.

2. Finite alphabet

We consider the model of an alphabet $\{1, \ldots, M\}$ where each letter occurs with probability $\frac{1}{M}$, and different letters are independent from each other.

We note that $\mathbb{E}(\chi_{j,k})$ is just the probability that (j, k) is a visible pair. This probability is not hard to compute: Let b be the larger of w_j and w_k. If the other one is strictly smaller, we get a factor 2 by symmetry; the other instance is that they are both equal to b. The letters in between must all be $> b$; hence we find

$$\mathbb{E}(\mathsf{Vis}) = \sum_{1\le j<k\le n} \frac{1}{M^{k+1-j}} \sum_{1\le b\le M} (M-b)^{k-1-j}\Big(2(b-1)+1\Big).$$

This instance is simple enough to compute it *exactly*:

$$\mathbb{E}(\mathsf{Vis}) = \sum_{1\le b\le M} \Big(-\frac{2}{b}+\frac{1}{b^2}+2\frac{n}{M}-\frac{n}{Mb}+2\frac{(M-b)^n}{bM^n}-\frac{(M-b)^n}{b^2M^n}\Big).$$

We will drop exponentially small terms, since we will also do this for the variance; otherwise, the complexity of the formulæ becomes unbearable:

$$\sum_{1\le b\le M} \Big(-\frac{2}{b}+\frac{1}{b^2}+\frac{2n}{M}-\frac{n}{Mb}\Big) = 2n - n\frac{H_M}{M} - 2H_M + H_M^{(2)}.$$

This is the promised formula that also includes a constant term and an exponentially small error term. For all our future computations, we mean ρ^n, where we can choose $1-\frac{1}{M}<\rho<1$. Here, we see harmonic numbers of the second order, but we introduce, more generally,

$$H_M^{(d)} = \sum_{1\le k\le M} \frac{1}{k^d}.$$

Now for the computation of the variance, we compute the second (factorial) moment: we have to compute

$$\mathbb{E}(\chi_{j,k}\cdot\chi_{l,m}),$$

where $1\le j<k\le n$, $1\le l<m\le n$, and $(j,k)\ne(l,m)$. Unfortunately, there are many cases to be distinguished, according to the pairs of indices. We distinguish 6 cases, and 6 other ones, which are symmetric, so that the cumulated results of the 6 cases (listed below) must be multiplied by 2. This comment applies as well for the other models studied later in this paper.

Here are the 6 ranges of summation:

(1) $\{1\le j<k<l<m\le n\}$,
(2) $\{1\le j<l<m<k\le n\}$,
(3) $\{1\le j<l<k<m\le n\}$,
(4) $\{1\le j<k=l<m\le n\}$,
(5) $\{1\le j<l<k=m\le n\}$,
(6) $\{1\le j=l<k<m\le n\}$;

the 6 other ones are obtained by the replacements $j\leftrightarrow l$ and $k\leftrightarrow m$.

In the following computations, we cannot give too many intermediate steps, otherwise the length of this paper would not be acceptable. The contribution of the first range is

$$\sum_{1\le j<k<l<m\le n} \frac{1}{M^{k+1-j+m+1-l}} \sum_{1\le b\le M} (2b-1)(M-b)^{k-1-j} \sum_{1\le d\le M} (2d-1)(M-d)^{m-1-l}.$$

With Maple, we can perform the sums on j,k,l,m and discard exponentially small terms, which leads to

$$\sum_{1\le b\le M}\sum_{1\le d\le M}\Big(-\frac{2}{d^3}-\frac{2}{b^3}+\frac{1}{d^2b^2}+\frac{4}{d^2}+\frac{4}{b^2}+\frac{2n^2}{M^2}-\frac{4n}{Md}-\frac{4n}{Mb}+\frac{2n}{Mb^2}$$

$$
\begin{aligned}
&+\frac{n}{M^2d}-\frac{n}{Md^2b}+\frac{n^2}{2M^2db}-\frac{n^2}{M^2d}-\frac{n^2}{M^2b}+\frac{n}{M^2b}-\frac{n}{2M^2db}-\frac{2n}{M^2}\\
&-\frac{n}{Mdb^2}+\frac{2n}{Md^2}+\frac{4n}{Mdb}+\frac{4}{db}+\frac{1}{d^3b}+\frac{1}{db^3}-\frac{4}{d^2b}-\frac{4}{db^2}\Big).
\end{aligned}
$$

The commands to achieve that are not difficult, one would for instance type in

```
sum(sum(sum(sum...,m=l+1..n),l=k+1..n-1),k=j+1..n-2),j=1..n-3);,
```

where the dots stand for the term in the sum. Then one gets many terms and asks to replace each occurrence of $(M-b)^n/M^n$ and $(M-d)^n/M^n$ by zero. What remains are the terms just mentioned. The steps as indicated will be performed for each of the following instances.

The further simplification of the sums must be done by hand, with the result

$$
\begin{aligned}
&2n^2+8H_M^{(2)}M-4MH_M^{(3)}+\big(H_M^{(2)}\big)^2-\frac{H_M^2n}{2M^2}+4H_M^2+\frac{H_M^2n^2}{2M^2}-\frac{2H_M^{(2)}H_Mn}{M}\\
&-2n-\frac{2H_Mn^2}{M}+4H_M^{(2)}n+2H_M^{(3)}H_M-8H_M^{(2)}H_M+\frac{4H_M^2n}{M}-8H_Mn+\frac{2H_Mn}{M};
\end{aligned}
$$

there is nothing difficult about it, it is a straight-forward term-by-term translation.

Now we move to the second range and perform similar operations:

$$
\sum_{1\le j<l<m<k\le n}\frac{1}{M^{k+1-j}}\times\sum_{1\le b<d\le M}(2b-1)(M-b)^{k-1-m+l-1-j}(2d-2b-1)(M-d)^{m-1-l}.
$$

Summed, and without exponentially small terms, the last sum becomes:

$$
\sum_{1\le b<d\le M}\Big(-\frac{4n}{Md}+\frac{4}{d^2}+\frac{4n}{bM}+\frac{4}{bd}-\frac{1}{b^2d^2}-\frac{2n}{b^2M}+\frac{n}{Mb^2d}-\frac{8}{b^2}+\frac{2}{b^2d}-\frac{2}{b^3d}+\frac{4}{b^3}\Big),
$$

or

$$
\begin{aligned}
&-8n+6\frac{nH_M}{M}+12H_M+4nH_M+2H_M^2-10H_M^{(2)}-\frac12\big(H_M^{(2)}\big)^2+\frac12H_M^{(4)}\\
&-2nH_M^{(2)}+\frac{n}{M}H_M^{(2,1)}-8MH_M^{(2)}+2H_M^{(2,1)}-2H_M^{(3,1)}+4MH_M^{(3)};
\end{aligned}
$$

the new notation refers to

$$
H_M^{(a,b)}=\sum_{1\le j<k\le M}\frac{1}{j^ak^b}.
$$

The contribution of the third range is zero for combinatorial reasons, and the fourth one leads to

$$
\begin{aligned}
\sum_{1\le j<k<m\le n}\frac{1}{M^{m+1-j}}\Big[&\sum_{1\le b<c\le M}(2b-1)(M-b)^{k-1-j}(M-c)^{m-1-k}\\
&+\sum_{1\le c<b\le M}(2c-1)(M-b)^{k-1-j}(M-c)^{m-1-k}
\end{aligned}
$$

$$+\sum_{1\le b\le M}(b^2+b-1)(M-b)^{k-1-j}(M-b)^{m-1-k}\Bigg].$$

Summed, and without exponentially small terms, this sum becomes:

$$\sum_{1\le b<c\le M}\Big(-\frac{4}{c^2}+\frac{4n}{Mc}+\frac{2}{bc^2}-\frac{2n}{bMc}-\frac{4}{cb}+\frac{2}{b^2c}\Big)$$
$$+\sum_{1\le b\le M}\Big(-\frac{2}{b}-\frac{2}{b^2}+\frac{2}{b^3}+\frac{n}{M}+\frac{n}{bM}-\frac{n}{b^2M}\Big),$$

or

$$-6H_M+4H_M^{(2)}+5n-3\frac{nH_M}{M}-\frac{n}{M}H_M^2-2H_M^2+2H_MH_M^{(2)}.$$

The fifth range leads to

$$\sum_{1\le j<l<k\le n}\frac{1}{M^{k+1-j}}\sum_{1\le c<b\le M}(2c-1)(M-c)^{l-1-j}(M-b)^{k-1-l}.$$

The sixth range leads to

$$\sum_{1\le j<k<m\le n}\frac{1}{M^{m+1-j}}\sum_{1\le c<b\le M}(2c-1)(M-b)^{k-1-j}(M-c)^{m-1-k},$$

which is the same as for the fifth range, so that we combine them.

Summed, and without exponentially small terms, this sum becomes:

$$\sum_{1\le c<b\le M}\Big(\frac{2}{bc^2}+\frac{4n}{bM}-\frac{2n}{bMc}-\frac{4}{cb}+\frac{2}{cb^2}-\frac{4}{b^2}\Big),$$

or

$$2H_MH_M^{(2)}-2H_M^{(3)}+4n-\frac{4nH_M}{M}-\frac{n}{M}H_M^2+\frac{n}{M}H_M^{(2)}-2H_M^2+2H_M^{(2)}-4H_M+4H_M^{(2)}. \quad (1)$$

Now we add the contributions from the 6 ranges as just computed, multiply by 2 (as explained), add the expectation, and subtract the square of expectation, which gives us the variance. There are many cancellations (as perhaps expected), and the result is

$$n\Big(\frac{2H_M^{(2)}}{M}-\frac{H_M^2}{M^2}-\frac{2H_MH_M^{(2)}}{M}+\frac{2H_M^{(2,1)}}{M}+\frac{H_M}{M}\Big)$$
$$+\Big(-4H_M^{(3,1)}-4H_M^{(3)}+H_M^{(4)}+H_M^{(2)}+2H_M+4H_MH_M^{(3)}-4H_MH_M^{(2)}+4H_M^{(2,1)}\Big).$$

For the reader's convenience, we summarise the findings of this section.

Theorem 1. *The visibility parameter (=number of visible pairs), in words of length n over an alphabet with M letters, has expectation and variance as follows:*

$$\mathbb{E}(\mathsf{Vis})=2n-n\frac{H_M}{M}-2H_M+H_M^{(2)}+O(\rho^n),$$

$$
\begin{aligned}
\mathbb{V}(\mathit{Vis}) &= n\Big(\frac{2H_M^{(2)}}{M} - \frac{H_M^2}{M^2} - \frac{2H_M H_M^{(2)}}{M} + \frac{2H_M^{(2,1)}}{M} + \frac{H_M}{M}\Big) \\
&\quad + \Big(-4H_M^{(3,1)} - 4H_M^{(3)} + H_M^{(4)} + H_M^{(2)} + 2H_M + 4H_M H_M^{(3)} - 4H_M H_M^{(2)} + 4H_M^{(2,1)}\Big) \\
&\quad + O(\rho^n).
\end{aligned}
$$

3. Permutations

Now we consider random permutations: The words of length n use letters $1, \dots, n$, each occurs exactly once, and all such words are equally likely.

We start with the expected value of the number of visible pairs. It is

$$2\sum_{1\le j<k\le n} \frac{1}{n^{\underline{k+1-j}}} \sum_{1\le a<b\le n} (n-b)^{\underline{k-1-j}}.$$

Here, we use the notation of falling factorials [2]:

$$x^{\underline{n}} = x(x-1)\dots(x-n+1).$$

The explanation is simple: 2 comes from symmetry, and the probability that the pair (j,k) is visible is computed as the number of favourable cases divided by number of all cases, as in elementary probability. Maple can compute the inner sums (with the $\mathsf{SumTools}$ package):

$$2\sum_{1\le j<k\le n} \frac{1}{(k-j)(k+1-j)} = 2\sum_{1\le h\le n} \frac{n-h}{h(h+1)} = 2n - 2H_n.$$

Again, we briefly mention what one must do. After loading the $\mathsf{SumTools}$ package, one types in

Summation(Summation(. . . ,b=a+1..n),a=1..n-1);,

where the dots stand for the term in the sum, and asks for simplification. The reduction from a sum over j and k to a sum just over h is not too hard in this instance; in this section we work out a full example which is much more complicated and occurs in the computation of the variance.

And now, for the variance, we consider again the 6 ranges of indices.

The first one leads to

$$
\begin{aligned}
4\sum_{1\le j<k<l<m\le n} &\frac{1}{n^{\underline{(k+1-j)+(m+1-l)}}} \times \\
&\Big[\sum_{1\le a<b<c<d} (n-d)^{\underline{m-1-l}}(n-b-m-1+l)^{\underline{k-1-j}} \\
&\quad + \sum_{1\le a<c<b<d} (n-d)^{\underline{m-1-l}}(n-b-m+l)^{\underline{k-1-j}} \\
&\quad + \sum_{1\le a<c<d<b} (n-b)^{\underline{k-1-j}}(n-d-k+j)^{\underline{m-1-l}} \\
&\quad + \sum_{1\le c<a<b<d} (n-d)^{\underline{m-1-l}}(n-b-m+l)^{\underline{k-1-j}}
\end{aligned}
$$

$$+\sum_{1\le c<a<d<b}(n-b)^{\underline{k-1-j}}(n-d-k+j)^{\underline{m-1-l}}$$

$$+\sum_{1\le c<d<a<b}(n-b)^{\underline{k-1-j}}(n-d-k-1+j)^{\underline{m-1-1}}\Bigg],$$

which Maple (using the SumTools package, as just described) can transform into

$$\sum_{1\le j<k<l<m\le n}\frac{4}{(m-l)(m+1-l)(k-j)(k+1-j)}=\sum_{1\le k<l\le n}\frac{4(k-1)(n-l)}{k(n+1-l)}.$$

The second range leads to

$$4\sum_{1\le j<l<m<k\le n}\frac{1}{n^{\underline{k+1-j}}}\sum_{1\le a<b<c<d\le n}(n-d)^{\underline{m-1-l}}(n-b-m-1+l)^{\underline{(l-1-j)+(k-1-m)}},$$

which Maple evaluates as

$$\sum_{1\le j<l<m<k\le n}\frac{4}{(m-l)(m+1-l)(k-j)(k+1-j)}.$$

Since the third range does not contribute, we move to the fourth:

$$\sum_{1\le j<k<m\le n}\frac{1}{n^{\underline{m+1-j}}}\times$$

$$\Bigg[\sum_{1\le a<b<c\le n}(n-b-m+k)^{\underline{k-1-j}}(n-c)^{\underline{m-1-k}}+\sum_{1\le a<c<b\le n}(n-b)^{\underline{m-2-j}}$$

$$+\sum_{1\le b<a<c\le n}(n-a-m+k)^{\underline{k-1-j}}(n-c)^{\underline{m-1-k}}$$

$$+\sum_{1\le b<c<a\le n}(n-a)^{\underline{k-1-j}}(n-c-k+j)^{\underline{m-1-k}}$$

$$+\sum_{1\le c<a<b\le n}(n-b)^{\underline{m-2-j}}+\sum_{1\le c<b<a\le n}(n-a)^{\underline{k-1-j}}(n-b-k+j)^{\underline{m-1-k}}\Bigg],$$

which Maple brings into this form:

$$\sum_{1\le j<k<m\le n}\frac{2}{(m+1-j)(m-j)}\left[\frac{1}{m-k}+\frac{1}{m-1-j}+\frac{1}{k-j}\right].$$

The fifth range leads to

$$\sum_{1\le j<l<k\le n}\frac{1}{n^{\underline{k+1-j}}}\Bigg[\sum_{1\le a<c<b\le n}(n-b)^{\underline{k-1-l}}(n-c-k+l)^{\underline{l-1-j}}$$

$$+\sum_{1\le c<a<b\le n}(n-b)^{\underline{k-1-l}}(n-a-k+l)^{\underline{l-1-j}}\Bigg],$$

which is

$$\sum_{1\le j<l<k\le n}\frac{2}{(k-l)(k-j)(k+1-j)};$$

the sixth range produces the same result.

The collection of the contributions of the 6 ranges is

$$\begin{aligned}&\sum_{1\le k<l\le n}\frac{4(k-1)(n-l)}{k(n+1-l)}+\sum_{1\le j<k<l\le n}\frac{4}{(k-j)(l-j)(l+1-j)}\\&+\sum_{1\le j<l<m<k\le n}\frac{4}{(m-l)(m+1-l)(k-j)(k+1-j)}\\&+\sum_{1\le j<k<l\le n}\frac{2}{(l+1-j)(l-j)}\Big[\frac{1}{l-k}+\frac{1}{l-1-j}+\frac{1}{k-j}\Big].\end{aligned}$$

The simplification of this is a long and tedious computation that we cannot produce in full here. It is done by hand; computers are only used to test that no errors occurred during the individual steps. The simplification is based on the following intermediate results:

$$\sum_{1\le k<l\le n}\frac{(k-1)(n-l)}{k(n+1-l)}=\frac{n(n-1)}{2}-2nH_n+2n+H_n^2-H_n^{(2)},$$

$$\sum_{1\le j<k<l\le n}\frac{1}{(l-j)(l+1-j)(k-j)}=n-H_n-\frac{1}{2}H_n^2+\frac{1}{2}H_n^{(2)},$$

$$\sum_{1\le j<k<l\le n}\frac{1}{l-j}\frac{1}{l+1-j}\frac{1}{l-1-j}=\frac{n}{2}-H_n+\frac{1}{2},$$

$$\sum_{1\le j<k<l\le n}\frac{1}{(l-j)(l+1-j)(l-k)}=n-H_n-\frac{1}{2}H_n^2+\frac{1}{2}H_n^{(2)},$$

$$\sum_{1\le j<k<l<m\le n}\frac{1}{(l-k)(l+1-k)(m-j)(m+1-j)}=nH_n-4n+3H_n+\frac{1}{2}H_n^2-\frac{1}{2}H_n^{(2)}.$$

To give the reader an idea how such formulæ can be obtained, we show how to compute one of the ingredients in full detail:

$$\begin{aligned}X&=\sum_{1\le j<k<l<m\le n}\Big[\frac{1}{l-k}-\frac{1}{l+1-k}\Big]\Big[\frac{1}{m-j}-\frac{1}{m+1-j}\Big]\\&=\sum_{1\le j<k<l\le n-1}\Big[\frac{1}{l-k}-\frac{1}{l+1-k}\Big]\Big[\frac{1}{l+1-j}-\frac{1}{n+1-j}\Big]\\&=\sum_{1\le j<l\le n-1}\Big[1-\frac{1}{l-j}\Big]\Big[\frac{1}{l+1-j}-\frac{1}{n+1-j}\Big]\\&=\sum_{1\le j\le n-2}\sum_{1\le l\le n-1-j}\Big[1-\frac{1}{l}\Big]\Big[\frac{1}{l+1}-\frac{1}{n+1-j}\Big]\end{aligned}$$

$$= \sum_{1\le j\le n-2}\sum_{1\le l\le n-1-j}\frac{1}{l+1} - \sum_{1\le j\le n-2}\sum_{1\le l\le n-1-j}\frac{1}{n+1-j}$$
$$- \sum_{1\le j\le n-2}\sum_{1\le l\le n-1-j}\frac{1}{l}\frac{1}{l+1} + \sum_{1\le j\le n-2}\sum_{1\le l\le n-1-j}\frac{1}{l}\frac{1}{n+1-j}$$
$$= \sum_{1\le j\le n-2}[H_{n-j}-1] - \sum_{1\le j\le n-2}\frac{n-1-j}{n+1-j} - \sum_{1\le j\le n-2}[1-\frac{1}{n-j}] + \sum_{1\le j\le n-2}\frac{H_{n-1-j}}{n+1-j}$$
$$= \sum_{1\le j\le n-2}H_{n-j} - \sum_{1\le j\le n-2}\frac{n-1-j}{n+1-j} + \sum_{1\le j\le n-2}\frac{1}{n-j} + \sum_{3\le j\le n}\frac{H_{j-2}}{j} - 2(n-2)$$
$$= \sum_{2\le j\le n-1}H_j + 2\sum_{1\le j\le n-2}\frac{1}{n+1-j} + \sum_{2\le j\le n-1}\frac{1}{j} + \sum_{3\le j\le n}\frac{H_{j-1}}{j} - \sum_{3\le j\le n}\frac{1}{j(j-1)} - 3(n-2)$$
$$= \sum_{1\le j\le n-1}H_j + 2\sum_{3\le j\le n}\frac{1}{j} + \sum_{1\le j\le n-1}\frac{1}{j} + \sum_{3\le j\le n}\frac{H_{j-1}}{j} - \sum_{3\le j\le n}\frac{1}{j(j-1)} - 3n+4$$
$$= \sum_{1\le j\le n-1}H_j + 2\sum_{3\le j\le n}\frac{1}{j} + H_n - \frac{1}{n} + \sum_{1\le j\le n}\frac{H_{j-1}}{j} - \frac{1}{2} - \frac{1}{2} + \frac{1}{n} - 3n+4$$
$$= nH_n - 4n + 3H_n + \frac{1}{2}H_n^2 - \frac{1}{2}H_n^{(2)}.$$

The reader is advised to consult the book by Graham, Knuth and Patashnik [2] for properties of harmonic numbers.

The result of the collection is

$$2n^2 - n + 1 - 4nH_n + 2H_n + 2H_n^2 - 2H_n^{(2)}.$$

Taking this times 2 (because of symmetry, as discussed before), adding the expectation, and subtracting the square of the expectation leads after simplification to the variance $2H_n + 2 - 4H_n^{(2)}$.

We summarise the results of this section:

Theorem 2. *The visibility parameter (=number of visible pairs), in permutations of n elements has expectation and variance as follows:*

$$\mathbb{E}(\mathit{Vis}) = 2n - 2H_n,$$
$$\mathbb{V}(\mathit{Vis}) = 2H_n + 2 - 4H_n^{(2)}.$$

Notice that the variance is very small, and thus the distribution highly concentrated.

4. Geometrically distributed words

Our model, to repeat it, is that each letter k appears with probability pq^{k-1}, independently from each other. For the expected value, we do the very same approach as before:

$$\mathbb{E}(\mathsf{Vis}) = \frac{p^2}{q^2}\sum_{1\le j<k\le n}\left[2\sum_{1\le a<b}q^{a+b}q^{b(k-1-j)} + \sum_{a\ge 1}q^{2a}q^{a(k-1-j)}\right]$$

$$= \frac{p^2}{q^2} \sum_{1\le j<k\le n} \left[2\frac{q^{k-j}}{1-q^{k-j}} \frac{q^{k+1-j}}{1-q^{k+1-j}} + \frac{q^{k+1-j}}{1-q^{k+1-j}} \right].$$

Even at that stage it becomes clear that we need to introduce some notation:

$$a_k := \frac{q^k}{1-q^k}$$

and

$$\sigma_d := \sum_{1\le j\le n} a_j\, j^d, \qquad \tau_d := \sum_{1\le j<k\le n} a_j\, a_k\, k^d,$$
$$\upsilon_d := \sum_{1\le j<k\le n} a_j\, a_k\, j^d, \qquad \mu := \sum_{1\le j<k\le n} a_j\, a_k\, jk.$$

Note that

$$a_k\, a_{k+1} = \frac{q}{p} a_k - \frac{1}{q} a_{k+1}, \tag{2}$$

and

$$a_k\, a_{k+2} = \frac{q^2}{p(1+q)} a_k - \frac{1}{p(1+q)} a_{k+2}, \tag{3}$$

as is easy to check. Consequently,

$$\begin{aligned}
\mathbb{E}(\mathsf{Vis}) &= \frac{p^2}{q^2} \sum_{1\le j<k\le n} \Big[2a_{k-j}\, a_{k+1-j} + a_{k+1-j} \Big] \\
&= \frac{p^2}{q^2} \sum_{1\le j\le n} (n-j) \Big[2a_j\, a_{j+1} + a_{j+1} \Big] \\
&= \frac{p^2}{q^2} \sum_{1\le j\le n} (n-j) \left[\frac{2q}{p} a_j - \frac{2}{p} a_{j+1} + a_{j+1} \right] \\
&= \frac{1+q}{q} n - \frac{p^2}{q^2} n\, \sigma_0 - \frac{p(1+q)}{q^2} \sigma_0 + \frac{p^2}{q^2} \sigma_1.
\end{aligned}$$

In the limit for $q \to 1$, this turns into

$$-2 \sum_{1\le j\le n} \frac{1}{j} + 2n = 2n - 2H_n,$$

as predicted from the model of permutations.

Now we turn to the computation of the second factorial moment.

The first range of summation contributes:

$$\begin{aligned}
\frac{p^4}{q^4} \sum_{1\le j<k<l<m\le n} &\left(2 \sum_{1\le a<b} q^a q^b q^{b(k-1-j)} + \sum_{1\le b} q^{2b} q^{b(k-1-j)} \right) \\
&\times \left(2 \sum_{1\le c<d} q^c q^d q^{d(k-1-j)} + \sum_{1\le d} q^{2d} q^{d(m-1-l)} \right)
\end{aligned}$$

$$= 4\frac{p^4}{q^4} \sum_{1\le j<k<l<m\le n} a_{k-j}\, a_{k-j+1}\, a_{m-l}\, a_{m-l+1}$$
$$+ 2\frac{p^4}{q^4} \sum_{1\le j<k<l<m\le n} a_{k-j}\, a_{k-j+1}\, a_{m-l+1}$$
$$+ 2\frac{p^4}{q^4} \sum_{1\le j<k<l<m\le n} a_{m-l}\, a_{m-l+1}\, a_{k-j+1}$$
$$+ \frac{p^4}{q^4} \sum_{1\le j<k<l<m\le n} a_{k-j+1}\, a_{m-l+1}.$$

The second range of summation contributes:

$$\begin{aligned}
\frac{p^4}{q^4} \sum_{1\le j<l<m<k\le n} \Bigg[& 4 \sum_{1\le a<d<b<c} q^a q^b q^c q^d q^{c(m-l-1)} q^{d(l-j-1+k-m-1)} \\
& + 2 \sum_{1\le a<c<b} q^a q^{2b} q^c q^{b(m-l-1)} q^{c(l-j-1+k-m-1)} \\
& + 2 \sum_{1\le a<b<c} q^{2a} q^b q^c q^{c(m-l-1)} q^{a(l-j-1+k-m-1)} \\
& + \sum_{1\le a<b} q^{2a} q^{2b} q^{b(m-l-1)} q^{a(l-j-1+k-m-1)} \Bigg] \\
& = 4\frac{p^4}{q^4} \sum_{1\le j<l<m<k\le n} a_{m-l}\, a_{m-l+1}\, a_{k-j}\, a_{k-j+1} \\
& + 2\frac{p^4}{q^4} \sum_{1\le j<l<m<k\le n} a_{m-l+1}\, a_{k-j}\, a_{k-j+1} \\
& + 2\frac{p^4}{q^4} \sum_{1\le j<l<m<k\le n} a_{m-l}\, a_{m-l+1}\, a_{k-j+1} \\
& + \frac{p^4}{q^4} \sum_{1\le j<l<m<k\le n} a_{m-l+1}\, a_{k-j+1}.
\end{aligned}$$

The fourth range is a bit long and originally consisted of 13 sums. Thus, we only present the simplified form:

$$\begin{aligned}
& 2\frac{p^3}{q^3} \sum_{1\le j<k<m\le n} a_{m-k}\, a_{m-j}\, a_{m-j+1} + 3\frac{p^3}{q^3} \sum_{1\le j<k<m\le n} a_{k-j+m-k}\, a_{k-j+m-k+1} \\
& + \frac{p^3}{q^3} \sum_{1\le j<k<m\le n} a_{m-k}\, a_{m-j+1} + 2\frac{p^3}{q^3} \sum_{1\le j<k<m\le n} a_{m-j-1}\, a_{m-j}\, a_{m-j+1} \\
& + \frac{p^3}{q^3} \sum_{1\le j<k<m\le n} a_{k-j+m-k-1}\, a_{k-j+m-k+1} + 2\frac{p^3}{q^3} \sum_{1\le j<k<m\le n} a_{k-j}\, a_{m-j}\, a_{m-j+1}
\end{aligned}$$

$$+\frac{p^3}{q^3}\sum_{1\le j<k<m\le n} a_{k-j}\,a_{m-j+1}+\frac{p^3}{q^3}\sum_{1\le j<k<m\le n} a_{m-j+1}.$$

The sixth range contributes

$$\frac{p^3}{q^3}\sum_{1\le j<k<m\le n}\Big[2\sum_{1\le a<c<b} q^a q^b q^c q^{b(k-1-j)} q^{c(m-1-k)}+\sum_{1\le c<b} q^b q^{2c} q^{b(k-1-j)} q^{c(m-1-k)}\Big]$$

$$=2\frac{p^3}{q^3}\sum_{1\le j<k<m\le n} a_{k-j}\,a_{m-j}\,a_{m+1-j}+\frac{p^3}{q^3}\sum_{1\le j<k<m\le n} a_{k-j}\,a_{m+1-j},$$

and this is also the contribution of the fifth range.

In the limit for $q\to 1$, the total contribution leads to the expression (1), which serves as a check.

The next task is to combine these 6 contributions (the third one is zero, as always), and to simplify. This computation is extremely long, and we cannot show all the steps. The formulæ (2) and (3) will be used to bring the second factorial moment (= twice the collected 6 contributions) into this form:

$$\begin{aligned}
&8\frac{p^2}{q^2}\sum_{1\le j<k<l<m\le n} a_{k-j}\ a_{m-l}-8\frac{p^2}{q^3}\sum_{1\le j<k<l<m\le n} a_{k-j}\ a_{m-l+1}\\
&+2\frac{p^2(1+q)^2}{q^4}\sum_{1\le j<k<l<m\le n} a_{k-j+1}\ a_{m-l+1}+8\frac{p^2}{q^2}\sum_{1\le j<k<l<m\le n} a_{l-k}\ a_{m-j}\\
&-4\frac{p^2(1+q)}{q^3}\sum_{1\le j<k<l<m\le n} a_{l-k}\ a_{m-j+1}+2\frac{p^2(1+q)^2}{q^4}\sum_{1\le j<k<l<m\le n} a_{l-k+1}\ a_{m-j+1}\\
&-4\frac{p^2(1+q)}{q^3}\sum_{1\le j<k<l<m\le n} a_{l-k+1}\ a_{m-j}+16\frac{p^2}{q^2}\sum_{1\le j<k<l\le n} a_{l-k}\ a_{l-j}\\
&-8\frac{p^2(1+q)}{q^3}\sum_{1\le j<k<l\le n} a_{l-k}\ a_{l-j+1}+2\frac{p(q^2+q-1)}{q^3}\sum_{1\le j<k\le n}(k-1-j)\ a_{k-j+1}\\
&+2\frac{p}{q}\sum_{1\le j<k\le n}(k-1-j)\ a_{k-j-1}+2\frac{p(1-3q)}{q^2}\sum_{1\le j<k\le n}(k-1-j)\ a_{k-j}.
\end{aligned}$$

Note that we interpret $0\cdot a_0$ as 0 in the penultimate sum. The next step is the translation of these expressions in terms of the standard sums σ,τ,υ,μ as introduced before. For these we prepared a *catalogue* of translation formulæ. It is organized as a table (Table 1). Again, the computations leading to it are very long; we just show one computation that is more difficult than the others:

$$\sum_{1\le j<k<l<m\le n} a_{k-j}\,a_{m-l}=\sum_{1\le j<k<n}\frac{(n-k)(n-k-1)}{2}a_j\,a_{k-j}$$

$$=\sum_{1\le j<k\le n}\frac{(n-k)(n-k-1)}{2}a_k[1+a_j+a_{k-j}]$$

Table 1. The catalogue of formulæ

Sum	Formula
$\sum_{1\le j<k<l\le n} a_{k-j}\, a_{l-j}$	$n\tau_0 - \tau_1$
$\sum_{1\le j<k<l\le n} a_{k-j}\, a_{l+1-j}$	$(n+1)\tau_0 - \tau_1 + n\,\sigma_0 - n\frac{q}{p^2} - \sigma_1 + \frac{1}{p}\sigma_0$
$\sum_{1\le j<k<l<m\le n} a_{l-k}\, a_{m-j}$	$n\tau_1 - n\upsilon_1 - n\tau_0 - \tau_2 + \mu + \tau_1$
$\sum_{1\le j<k<l<m\le n} a_{l+1-k}\, a_{m-j}$	$\tau_1 - n\upsilon_1 - \tau_2 + \mu - \frac{q}{p}n\,\sigma_1 + \frac{q}{p}n\,\sigma_0 + \frac{q}{p}\sigma_2 - \frac{q}{p}\sigma_1$
$\sum_{1\le j<k<l<m\le n} a_{l-k}\, a_{m+1-j}$	$(n+3)\tau_1 - (n+1)\upsilon_1 - 2(n+1)\tau_0 - \tau_2 + \mu$ $- n\,\sigma_0 + \sigma_1 - \frac{1}{p}\sigma_0 + \frac{q}{p^2}n$
$\sum_{1\le j<k<l<m\le n} a_{l+1-k}\, a_{m+1-j}$	$(n+2)\tau_1 - (n+1)\upsilon_1 - (n+1)\tau_0 - \tau_2 + \mu$ $- \frac{q}{p}[(n+3)\sigma_1 - 2(n+1)\sigma_0 - \sigma_2] - \frac{q^2}{p^2}n$
$\sum_{1\le j<k<l<m\le n} a_{k-j}\, a_{m-l}$	$-\frac{n(n-1)}{2}\sigma_0 + \frac{n^2+n-1}{2}\sigma_1 - n\,\sigma_2 + \frac{1}{2}\sigma_3$ $+ n(n-1)\tau_0 - (2n-1)\tau_1 + \tau_2$
$\sum_{1\le j<k<l<m\le n} a_{k-j}\, a_{m+1-l}$	$-\frac{n(n+1)}{2}\sigma_0 + \frac{n^2+3n+1}{2}\sigma_1 - (n+1)\sigma_2 + \frac{1}{2}\sigma_3$ $+ n(n+1)\tau_0 - (2n+1)\tau_1 + \tau_2$ $- \frac{q}{2p}[n(n-1)\sigma_0 - (2n-1)\sigma_1 + \sigma_2]$
$\sum_{1\le j<k<l<m\le n} a_{k+1-j}\, a_{m-l}$	$-\frac{n(n+1)}{2}\sigma_0 + \frac{n^2+3n+1}{2}\sigma_1 - (n+1)\sigma_2 + \frac{1}{2}\sigma_3$ $+ n(n+1)\tau_0 - (2n+1)\tau_1 + \tau_2$ $- \frac{q}{2p}[n(n-1)\sigma_0 - (2n-1)\sigma_1 + \sigma_2]$
$\sum_{1\le j<k<l<m\le n} a_{k+1-j}\, a_{m+1-l}$	$-\frac{(n+2)(n+1)}{2}\sigma_0 + \frac{n^2+5n+5}{2}\sigma_1 - (n+2)\sigma_2 + \frac{1}{2}\sigma_3$ $+ (n+2)(n+1)\tau_0 - (2n+3)\tau_1 + \tau_2$ $- \frac{q}{p}n(n+1)\sigma_0 + \frac{q}{p}(2n+1)\sigma_1 - \frac{q}{p}\sigma_2 + \frac{q^2}{p^2}\frac{n(n-1)}{2}$
$\sum_{1\le j<k\le n}(k-1-j)a_{k-j}$	$(n+1)\sigma_1 - n\,\sigma_0 - \sigma_2$
$\sum_{1\le j<k\le n}(k-1-j)a_{k+1-j}$	$(n+3)\sigma_1 - 2(n+1)\sigma_0 - \sigma_2 + n\frac{q}{p}$
$\sum_{1\le j<k\le n}(k-1-j)a_{k-1-j}$	$(n-1)\sigma_1 - \sigma_2 + na_n$

$$= \sum_{1\le k\le n} \frac{(n-k)(n-k-1)(k-1)}{2} a_k + \sum_{1\le j<k\le n} (n-k)(n-k-1)a_j\, a_k$$
$$= -\frac{n(n-1)}{2}\sigma_0 + \frac{n^2+n-1}{2}\sigma_1 - n\,\sigma_2 + \frac{1}{2}\sigma_3 + n(n-1)\tau_0 - (2n-1)\tau_1 + \tau_2.$$

Eventually we have computed expectation and variance.

Theorem 3. *The visibility parameter (=number of visible pairs), in words of length n over the positive integers, equipped with geometric probabilities, has expectation and* *variance as follows:*

$$\mathbb{E}(\mathsf{Vis}) = \frac{1+q}{q}n - \frac{p^2}{q^2}n\,\sigma_0 - \frac{p(1+q)}{q^2}\sigma_0 + \frac{p^2}{q^2}\sigma_1,$$

$$\begin{aligned}\mathbb{V}(\mathsf{Vis}) = n^2\Big(&-\frac{p^2(1+3q^2)}{q^4}\sigma_0 + \frac{p^4}{q^4}\sigma_1 + \frac{2p^4}{q^4}\tau_0 + \frac{(1+q)^2}{q^2}\Big)\\
&+ n\Big(-\frac{p(3-3q+3q^2+5q^3)}{q^4}\sigma_0 + \frac{p^2(5-2q+q^2)}{q^4}\sigma_1 - \frac{2p^4}{q^4}\sigma_2\\
&\qquad + \frac{4p^3(1+q)}{q^4}\tau_0 - \frac{2p^4}{q^4}\tau_1 - \frac{2p^4}{q^4}\upsilon_1 - \frac{1+q^2}{q^2} + \frac{2p}{q}a_n\Big)\\
&+ \frac{2p(-1+q+q^2+q^3)}{q^4}\sigma_0 - \frac{p^2(q^2-5)}{q^4}\sigma_1 - \frac{2p^3(2+q)}{q^4}\sigma_2\\
&+ \frac{p^4}{q^4}\sigma_3 + \frac{2p^2(1+q)^2}{q^4}\tau_0 - \frac{2p^3(1+q)}{q^4}\tau_1 - \frac{2p^3(1+q)}{q^4}\upsilon_1 + \frac{2p^4}{q^4}\mu\\
&+ \mathbb{E}(\mathsf{Vis}) - \big(\mathbb{E}(\mathsf{Vis})\big)^2.\end{aligned}$$

Remark 1. *From the previous expression one could get the impression that the variance is of order n^2. This is not so; the coefficient of n^2 is*

$$\frac{p^4}{q^4}\Big(2\tau_0 + \sigma_1 - \sigma_0 - \sigma_0^2\Big).$$

It is easy to see that

$$\sigma_0^2 - 2\tau_0 = \sum_{1\le k\le n} a_k^2,$$

and

$$\begin{aligned}\sigma_1 - \sigma_0 &= \sum_{1\le j\le n}(j-1)\frac{q^j}{1-q^j} = \sum_{j\ge1}(j-1)\frac{q^j}{1-q^j} + O(nq^n)\\
&= \sum_{j,k\ge1}(j-1)q^{jk} + O(nq^n) = \sum_{1\le k\le n}\sum_{j\ge1}(j-1)q^{jk} + O(nq^n)\\
&= \sum_{1\le k\le n}\frac{q^{2k}}{(1-q^k)^2} + O(nq^n) = \sum_{1\le k\le n} a_k^2 + O(nq^n).\end{aligned}$$

Therefore

$$\frac{p^4}{q^4}\Big(2\tau_0 + \sigma_1 - \sigma_0 - \sigma_0^2\Big) = O(n^3q^n),$$

and the variance is of order n, as expected.

Remark 2. *It can be noted that*

$$\tau_1 + \upsilon_1 = \sigma_0\,\sigma_1 - \sum_{1\le k\le n} k a_k^2 \qquad \text{and} \qquad \mu = \frac{1}{2}\Big(\sigma_1^2 - \sum_{1\le k\le n} k^2 a_k^2\Big).$$

Remark 3. *As is easy to check, the limit for $q\to1$ reproduces indeed the results from the section on permutations.*

5. Conclusion

We found the interaction between time series, graphs, combinatorics on words, computer algebra, and analysis of algorithms extremely fascinating. It is our hope that this paper will help to popularize the area.

For those who want to practise themselves how such computations can be done, here are some variations that one can consider:

Use the original definition with '$<$', as given in [3] in the instance of geometrically distributed words.

For words, work with $\geq$ instead of $>$ (resp. with $\leq$ instead of $<$).

Finally we mention (during the revision in April, 2012) that one of us is currently preparing a paper with a related parameter

$$\mathsf{Box} = \sum_{1\leq j<k\leq n} \chi_{j,k}$$

where

$$\chi_{j,k}(w_1 \dots w_n) = \begin{cases} 1 & \text{if } \min\{w_j, w_k\} < w_l < \max\{w_j, w_k\} \text{ for all } j < l < k, \\ 0 & \text{otherwise.} \end{cases}$$

Details will (first) appear on the web when the paper is finished.

References

[1] Ph. Flajolet and N. Martin. Probabilistic Counting Algorithms for Data Base Applications. *Journal of Computer and System Sciences*, 31(2):182–209, 1985.

[2] R. L. Graham, D. E. Knuth, and O. Patashnik. *Concrete Mathematics: A Foundation for Computer Science*. Addison-Wesley, 1992.

[3] G. Gutin, T. Mansour, and S. Severini. A characterization of horizontal visibility graphs and combinatorics on words. *Physica A: Statistical Mechanics and its Appications*, 390(12):2421–2428, 2011.

[4] H. Prodinger. Combinatorics of geometrically distributed random variables: Left-to-right maxima. *Discrete Mathematics*, 153:253–270, 1996.

[5] H. Prodinger. Combinatorics of geometrically distributed random variables: Inversions and a parameter of Knuth. *Annals of Combinatorics*, 5:241–250, 2001.

[6] H. Prodinger. A q-analogue of the path length of binary search trees. *Algorithmica*, 31:433–441, 2001.

[7] W. Pugh. Skip lists: a probabilistic alternative to balanced trees. *Communications of the ACM*, 33:668–676, 1990.

Ligia L. Cristea, Institut für Mathematik, Technische universität Graz, Steyrergasse 30, 8010-Graz, Austria.

E-mail address: `strublistea@gmail.com`

Helmut Prodinger, Mathematics Department, Stellenbosch University, 7602 Stellenbosch, South Africa.

E-mail address: `hproding@sun.ac.za`