RECORDS IN GEOMETRICALLY DISTRIBUTED WORDS: SUM OF POSITIONS

HELMUT PRODINGER

Abstract. Kortchemski introduced a new parameter for random permutations: the sum of the positions of the records. We investigate this parameter in the context of random words, where the letters are obtained by geometric probabilities. We find a relation for a bivariate generating function, from which we can obtain the expectation, exactly and asymptotically. In principle, one could get all moments from it, but the computations would be huge.

1. Introduction

The recent paper [5] deals with records (left-to-right maxima) in (random) permutations. An element x_i in $x_1 \dots x_n$ is a record if it is larger than $x_1, \dots, x_{i-1}$. Furthermore, the position of this record is i. The new statistic defined in [5] is the sum over the positions of all records.

Example. 47516823 has the records $4, 7, 8$, in positions $1, 2, 6$, whence the new parameter is $1 + 2 + 6 = 9$.

The probability generating function was worked out, and is

$$\prod_{k=1}^{n} \frac{z^k + k - 1}{k},$$

from which the expectation can be computed as

$$\sum_{k=1}^{n} \frac{d}{dz} \frac{z^k + k - 1}{k}\Big|_{z=1} = \sum_{k=1}^{n} \frac{k}{k} = n.$$

In this paper, we want to study this parameter for words over the alphabet $\{1, 2, 3, \dots\}$, equipped with geometric probabilities $p, pq, pq^2, \dots$, with $p + q = 1$.

There are two versions: Strong records (must be strictly larger than elements to the left) and weak records (must be only larger or equal to elements to the left). Otherwise the definition of position and sum of positions is the same. This model can be considered to be a q-version; for $q \to 1$ the model of permutations is approached. This holds true, since the positions and their sums are independent of the value.

Date: April 10, 2008.

Key words and phrases. Words, geometric probabilities, records, Rice's method.

We are able to find a functional relation for the double generating function $F(z,u)$, where the coefficient of $z^n u^k$ is the probability that a random word of length n has new parameter k. (In [5], the new parameter is denoted srec.)

From it, we are able to deduce the expected value, both, exactly and asymptotically. Higher moments could in principle be *pumped out*, but only with a considerable effort. In [3] and [2], such computations has been performed; this will be an indicator of the work involved. We cannot say anything about limiting distributions, but this is also still open in the instance of skip lists just cited, and the $\mathsf{path\ length}$ there has a similar flavour to the present srec.

The methodology is described in detail in [2] and [4], and the interested reader can consult these papers first, to be prepared for the computations to come.

We will find the abbreviation

$$[\![k]\!] := 1 - z(1-q^k)$$

quite useful; it was also used in our earlier papers dealing with words equipped with geometric probabilities.

2. Strong records

Let $f_k(z,u)$ be the following generating function: The coefficient of $z^n u^l$ is the probability that a random word of length n has a record k in its last letter, and the srec parameter is $= l$.

The recursion

$$f_k(z,u) = \sum_{1\le j<k} f_j(zu,u)\frac{zupq^{k-1}}{1-(1-q^j)zu} + zupq^{k-1}$$

is straightforward: The previous record was j, and we do an arbitrary number of letters $\le j$, followed by the last letter k. Additionally, all z's are replaced by zu, to keep track of the sum of the positions. The last term stands for the instance that there was no previous record.

The promised generating function is now

$$F(z,u) = \sum_{k\ge1} f_k(z,u)\frac{1}{[\![k]\!]},$$

since any k can be the last record, and after it, only letters $\le k$ can occur.

Now note that

$$f_k(z,1) = \frac{zpq^{k-1}}{[\![k-1]\!]},$$

which follows directly from the combinatorial description: The last letter is k, the letters before are $< k$. We will need the functions

$$a_k(z) = \frac{\partial}{\partial u} f_k(z,u)\Big|_{u=1}.$$

We rewrite the recursion as

$$\frac{f_k(z,u)}{zupq^{k-1}} = \sum_{1\le j<k} \frac{f_j(zu,u)}{1-(1-q^j)zu} + 1$$

and take differences:

$$\frac{f_k(z,u)}{zpq^{k-1}} - \frac{f_{k-1}(z,u)}{zpq^{k-2}} = \frac{f_{k-1}(zu,u)u}{1-(1-q^{k-1})zu}.$$

Now we differentiate this relation w.r.t. u, plug in $u=1$ and get

$$\frac{a_k(z)}{zpq^{k-1}} - \frac{a_{k-1}(z)}{zpq^{k-2}} = \frac{f_{k-1}(z,1)}{[\![k-1]\!]} + \frac{a_{k-1}(z)}{[\![k-1]\!]} + \frac{z}{[\![k-1]\!]}\frac{d}{dz}f_{k-1}(z,1) + \frac{f_{k-1}(z,1)(1-q^{k-1})z}{[\![k-1]\!]^2}.$$

Simplifying, we get

$$\frac{a_k(z)[\![k-1]\!]}{zpq^{k-1}} - \frac{a_{k-1}(z)[\![k-2]\!]}{zpq^{k-2}} = \frac{zpq^{k-2}}{[\![k-1]\!][\![k-2]\!]} + \frac{zpq^{k-2}}{[\![k-2]\!]^2}.$$

These relations can be summed:

$$\frac{a_k(z)[\![k-1]\!]}{zpq^{k-1}} - 1 = \sum_{j=2}^{k} \frac{zpq^{j-2}}{[\![j-1]\!][\![j-2]\!]} + \sum_{j=2}^{k} \frac{zpq^{j-2}}{[\![j-2]\!]^2}.$$

Further,

$$\frac{a_k(z)}{[\![k]\!]} = \frac{zpq^{k-1}}{[\![k-1]\!][\![k]\!]} + \frac{zpq^{k-1}}{[\![k-1]\!][\![k]\!]}\sum_{j=1}^{k-1} \frac{zpq^{j-1}}{[\![j]\!][\![j-1]\!]} + \frac{zpq^{k-1}}{[\![k-1]\!][\![k]\!]}\sum_{j=2}^{k} \frac{zpq^{j-2}}{[\![j-2]\!]^2}.$$

Noticing that

$$\frac{zpq^{k-1}}{[\![k-1]\!][\![k]\!]} = \frac{1}{[\![k]\!]} - \frac{1}{[\![k-1]\!]},$$

we get

$$\frac{a_k(z)}{[\![k]\!]} = \frac{zpq^{k-1}}{[\![k-1]\!][\![k]\!]} + \frac{zpq^{k-1}}{[\![k-1]\!][\![k]\!]}\left[\frac{1}{[\![k-1]\!]} - 1\right] + \frac{zpq^{k-1}}{[\![k-1]\!][\![k]\!]}\sum_{j=2}^{k} \frac{zpq^{j-2}}{[\![j-2]\!]^2}.$$

Now we can compute $G(z)$, the generating functions of the expectations:

$$\begin{aligned}
G(z) &= \sum_{k\ge1} \frac{a_k(z)}{[\![k]\!]}\\
&= \sum_{k\ge1} \frac{zpq^{k-1}}{[\![k-1]\!]^2[\![k]\!]} + \sum_{1\le j<k}\left[\frac{1}{[\![k]\!]} - \frac{1}{[\![k-1]\!]}\right]\frac{zpq^{j-1}}{[\![j-1]\!]^2}\\
&= \sum_{k\ge1} \frac{zpq^{k-1}}{[\![k-1]\!]^2[\![k]\!]} + \sum_{1\le j}\left[\frac{1}{1-z} - \frac{1}{[\![j]\!]}\right]\frac{zpq^{j-1}}{[\![j-1]\!]^2}\\
&= \sum_{k\ge1} \frac{zpq^{k-1}}{[\![k-1]\!]^2[\![k]\!]} + \frac{1}{1-z}\sum_{1\le j} \frac{zpq^{j-1}}{[\![j-1]\!]^2} - \sum_{1\le j} \frac{zpq^{j-1}}{[\![j-1]\!]^2[\![j]\!]}
\end{aligned}$$

$$= \frac{1}{1-z}\sum_{k\ge1}\frac{zpq^{k-1}}{[\![k-1]\!]^2}.$$

Now we use the substitution $z = w/(w-1)$:

$$G(z) = -p(1-w)^2\sum_{k\ge0}\frac{wq^k}{(1-wq^k)^2}.$$

The formula for coefficients is

$$[z^n]G(z) = (-1)^n[w^n](1-w)^{n-1}G(z(w)),$$

which is easy to derive via Cauchy's integral formula, see the earlier papers.

Consequently,

$$\begin{aligned}
[z^n]G(z) &= p(-1)^{n+1}[w^n](1-w)^{n+1}\sum_{k\ge0}\frac{wq^k}{(1-wq^k)^2}\\
&= p(-1)^{n+1}[w^n]\sum_{l=0}^{n+1}\binom{n+1}{l}(-w)^{n+1-l}\sum_{k\ge0}\frac{wq^k}{(1-wq^k)^2}\\
&= p\sum_{l=1}^{n+1}\binom{n+1}{l}(-1)^l[w^{l-1}]\sum_{k\ge0}\frac{wq^k}{(1-wq^k)^2}\\
&= p\sum_{l=2}^{n+1}\binom{n+1}{l}(-1)^l\sum_{k\ge0}(q^k)^{l-1}[w^{l-1}]\frac{w}{(1-w)^2}\\
&= p\sum_{l=2}^{n+1}\binom{n+1}{l}(-1)^l\frac{l-1}{1-q^{l-1}}.
\end{aligned}$$

Theorem 1. *The expected value of the sum of the positions of records, in random words of length n, is given by*

$$E_n = p\sum_{k=2}^{n+1}\binom{n+1}{k}(-1)^k\frac{k-1}{1-q^{k-1}}.$$

This expression possesses a representation as a contour integral (Rice's method [1]):

$$E_n = \frac{p}{2\pi i}\oint_{\mathcal{C}}\frac{(-1)^{n+1}(n+1)!}{z(z-1)\dots(z-n-1)}\frac{z-1}{1-q^{z-1}};$$

the curve $\mathcal{C}$ encircles the poles $2,3,\dots,n+1$ and no others.

An asymptotic expansion can be obtained by changing the contour, and taking the extra residues into account (with a negative sign). There are poles at $z = 1+\frac{2\pi ik}{\log Q} =: 1+\chi_k$, for $k\in\mathbb{Z}$, and $Q := \frac{1}{q}$, and also at $z=0$ (this one we ignore). Collecting the contributions, we find

$$\frac{pn}{\log Q} + \frac{pn}{\log Q}\sum_{k\ne0}\chi_k\Gamma(-1-\chi_k)e^{2\pi ik\cdot\log_Q n} + O(1).$$

Theorem 2. *The expected value of the sum of the positions of records, in random words of length n, has the asymptotic expansion*

$$E_n = \frac{pn}{\log Q}\big(1+\delta(\log_Q n)\big) + O(1),$$

where the periodic function (of small amplitude) is given by

$$\delta(x) = \sum_{k\neq 0} \chi_k \Gamma(-1-\chi_k) e^{2\pi i k x}.$$

Finally, let us perform the limit $q \to 1$ in the explicit formula for E_n:

$$\sum_{k=2}^{n+1} \binom{n+1}{k} (-1)^k = n,$$

as predicted.

3. Weak records

The computations are somewhat similar, so that we can be a bit shorter. We will use the same notation, but with the new meaning.

The recursion is

$$f_k(z,u) = \sum_{j\le k} f_j(zu,u) \frac{zupq^{k-1}}{1-zu(1-q^{j-1})} + zupq^{k-1},$$

leading us to the generating function of interest

$$F(z,u) = \sum_{k\ge 1} f_k(z,u) \frac{1}{[\![k-1]\!]}.$$

The recursion can be written as

$$\frac{f_k(z,u)}{zpq^{k-1}} - \frac{f_{k-1}(z,u)}{zpq^{k-2}} = \frac{f_k(zu,u)u}{1-zu(1-q^{k-1})},$$

whence

$$\frac{a_k(z)}{zpq^{k-1}} - \frac{a_{k-1}(z)}{zpq^{k-2}} = \frac{f_k(z,1)}{[\![k-1]\!]} + \frac{a_k(z)}{[\![k-1]\!]} + \frac{z\frac{d}{dz}f_k(z,1)}{[\![k-1]\!]} + \frac{f_k(z,1)z(1-q^{k-1})}{[\![k-1]\!]^2}$$

or

$$\frac{a_k(z)[\![k]\!]}{zpq^{k-1}} - \frac{a_{k-1}(z)[\![k-1]\!]}{zpq^{k-2}} = \frac{zpq^{k-1}}{[\![k]\!][\![k-1]\!]} + \frac{zpq^{k-1}}{[\![k]\!]^2}.$$

Summing,

$$\frac{a_k(z)[\![k]\!]}{zpq^{k-1}} - \frac{a_1(z)[\![1]\!]}{zp} = \sum_{j=2}^{k} \frac{zpq^{j-1}}{[\![j]\!][\![j-1]\!]} + \sum_{j=2}^{k} \frac{zpq^{j-1}}{[\![j]\!]^2}.$$

Notice that

$$a_1(z) = \sum_{n\ge1} \binom{n+1}{2} (pz)^n = \frac{pz}{[\![1]\!]^3},$$

whence

$$\frac{a_k(z)}{[\![k-1]\!]} = \frac{zpq^{k-1}}{[\![k-1]\!][\![k]\!]^2} + \frac{(pz)^2 q^{k-1}}{(1-pz)^2[\![k]\!][\![k-1]\!]} + \frac{zpq^{k-1}}{[\![k-1]\!][\![k]\!]} \sum_{j=2}^{k} \frac{zpq^{j-1}}{[\![j]\!]^2}.$$

Summing,

$$\begin{aligned} G(z) &= \sum_{k\ge1} \frac{zpq^{k-1}}{[\![k-1]\!][\![k]\!]^2} + \frac{pz^2}{(1-pz)^2(1-z)} + \sum_{j\ge2} \frac{zpq^{j-1}}{[\![j]\!]^2} \sum_{k\ge j} \frac{zpq^{k-1}}{[\![k-1]\!][\![k]\!]} \\ &= \sum_{k\ge1} \frac{zpq^{k-1}}{[\![k-1]\!][\![k]\!]^2} + \frac{pz^2}{(1-pz)^2(1-z)} + \sum_{j\ge2} \frac{zpq^{j-1}}{[\![j]\!]^2} \left[\frac{1}{1-z} - \frac{1}{[\![j-1]\!]} \right] \\ &= \frac{zp}{(1-pz)^2} + \frac{pz^2}{(1-pz)^2(1-z)} + \frac{1}{1-z} \sum_{k\ge2} \frac{zpq^{k-1}}{[\![k]\!]^2} \\ &= \frac{pz}{(1-pz)^2(1-z)} + \frac{1}{1-z} \sum_{k\ge2} \frac{zpq^{k-1}}{[\![k]\!]^2}. \end{aligned}$$

This is very similar to the previous instance

$$G^{\text{strong}}(z) = \frac{1}{1-z} \sum_{k\ge0} \frac{zpq^k}{[\![k]\!]^2}.$$

And so

$$G(z) = \frac{1}{q} G^{\text{strong}}(z) - \frac{zp}{q(1-z)}.$$

This gives us the new expectations in terms of the old ones:

$$E_n = \frac{1}{q} E_n^{\text{strong}} - \frac{p}{q}.$$

For the asymptotic formula, we just divide the previous one by q:

$$E_n \sim \frac{pn}{q \log Q} \big(1 + \delta(\log_Q n)\big).$$

References

[1] P. Flajolet and R. Sedgewick. Mellin transforms and asymptotics: Finite differences and Rice's integrals. *Theoretical Computer Science*, 144:101–124, 1995.

[2] P. Kirschenhofer, C. Martínez, and H. Prodinger. Analysis of an optimized search algorithm for skip lists. *Theoretical Computer Science*, 144:199–220, 1995.

[3] P. Kirschenhofer and H. Prodinger. The path length of random skip lists. *Acta Informatica*, 31:775–792, 1994.

[4] A. Knopfmacher and H. Prodinger. Combinatorics of geometrically distributed random variables: Value and position of the rth left-to-right maximum. *Discrete Mathematics*, 226:255–267, 2001.

[5] I. Kortchemski. Asymptotic behavior of permutation records. arXiv:0804.044v1, pages: 12, 2008.

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

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