Applicable Analysis and Discrete Mathematics
available online at http://pefmath.etf.rs

Appl. Anal. Discrete Math. **5** (2011), 55–59. doi:10.2298/AADM110205005O

WORDS CODING SET PARTITIONS

Kamilla Oliver, Helmut Prodinger

The words in the title are characterized by the fact that a smaller number must (first) appear earlier than a larger number, and that all numbers $1, \dots, k$ are present (for some k). Under the assumption that the letters are drawn from a geometric distribution, the probability that a word of length n enjoys these properties is determined, both exactly and asymptotically.

1. INTRODUCTION

For a set partition of $\{1, 2, \dots, n\}$ into k blocks, a natural coding is as follows: Element 1 is in block 1, and the smallest number not in block 1 is in block 2, and the smallest number not in blocks 1 or 2, is in block 3, etc. In this way, to every element i a number a_i is attached, namely the block in which it lies. Writing these numbers as a word $a_1 \dots a_n$, the set partition is coded in a natural way. One particular reference for this is [**3**].

Forgetting now about set partitions, we are talking about words where the letters are the positive integers, and, assuming that k is the largest letter that appears in the word, then the letters $1, \dots, k-1$ must also appear, and the word has exactly k (strict) left-to-right maxima, which is the same as saying that, if $i < j$, the first appearance of i is earlier than the first appearance of j. As one referee has kindly pointed out, such words are known as *restricted growth strings* in the literature [**6**].

Now we assign the (geometric) probability pq^{i-1} (where $p+q = 1$) to the letter i and consider P_n, the probability that a random word of length n has the *restricted growth* property. We are thus in the context of *combinatorics of geometrically distributed words*, a series of papers started with [**4**] and continued by the second writer as well as many others; a recent contribution is the paper [**5**].

2000 Mathematics Subject Classification. 00A05.
Keywords and Phrases. q-series, enumeration, asymptotics, periodic oscillation.

The present question is not only appealing from a combinatorial point of view (easy to formulate but not trivial to solve) but the approach used here (with the parameter q) leads to "richer" results, and often the instance $q = 1$ corresponds to the classical combinatorial instance, especially, when the parameter is of the *order statistics* type.

We will prove the following theorems.

Theorem 1. *The probability P_n that a random word of length n has the* restricted growth *property is* (*exactly*) *given by*

$$P_n = p \sum_{j=0}^{n-1} (-1)^j \binom{n-1}{j} q^j (p;q)_j.$$

Here we use the (standard) notation $(x;q)_m = (1-x)(1-xq)\dots(1-xq^{m-1})$. We will also need the limit of it as $m \to \infty$, denoted by $(x;q)_\infty$, as well as the Gaussian q-binomial coefficients

$$\begin{bmatrix} n \\ k \end{bmatrix}_q := \frac{(q;q)_n}{(q;q)_k (q;q)_{n-k}}.$$

We need the following standard formulae:

$$\sum_{k=0}^{N} \begin{bmatrix} N \\ k \end{bmatrix}_q (-1)^k q^{\binom{k}{2}} x^k = (x;q)_N, \qquad \frac{1}{(w;q)_\infty} = \sum_{n \ge 0} \frac{w^n}{(q;q)_n}.$$

All this can be found in [**1**].

The asymptotic evaluation leads to our second theorem.

Theorem 2. *The probability that a random word of length n has the* restricted growth *property is asymptotically given by*

$$P_n \sim \frac{(p;q)_\infty}{L(q;q)_\infty} \Gamma\Big(\frac{\log p}{\log q}\Big) n^{-\frac{\log p}{\log q}} + n^{-\frac{\log p}{\log q}} \Phi(\log_Q n),$$

where $\Phi(x)$ is a 1-periodic function with mean zero. The abbreviations $Q = 1/q$ and $L = \log Q$ are used. The function is given by its Fourier series

$$\Phi(x) = \frac{(p;q)_\infty}{L(q;q)_\infty} \sum_{k \ne 0} \Gamma\Big(\frac{\log p}{\log q} + \frac{2\pi i k}{L}\Big) e^{-2\pi i k x}.$$

In the symmetric case $p = q$, this looks better:

$$\frac{1}{L} n^{-1} + n^{-1} \Phi(\log_2 n).$$

2. ANALYSIS

We use the natural decomposition

$$1\{\le 1\}^*2\{\le 2\}^*3\{\le 3\}^*\dots k\{\le k\}^*,$$

which translates into

$$\frac{zp}{1-(1-q)z}\,\frac{zpq}{1-(1-q^2)z}\cdots\frac{zpq^{k-1}}{1-(1-q^k)z}=z^kp^kq^{\binom{k}{2}}\prod_{j=1}^{k}\frac{1}{1-(1-q^j)z}.$$

This has to be summed over all k, to get the generating function of the sought probabilities (P_n is the coefficient of z^n in this series):

$$\sum_{k\ge1}z^kp^kq^{\binom{k}{2}}\prod_{j=1}^{k}\frac{1}{1-(1-q^j)z}.$$

Substituting $z=w/(w-1)$, this becomes

$$\sum_{k\ge1}w^k(-1)^kp^kq^{\binom{k}{2}}\prod_{j=1}^{k}\frac{1}{1-wq^j}=\sum_{k\ge1}\frac{w^k(-1)^kp^kq^{\binom{k}{2}}}{(wq;q)_k}.$$

Reading off coefficients:

$$\begin{aligned}
P_n&=[z^n]\sum_{k\ge1}\frac{w^k(-1)^kp^kq^{\binom{k}{2}}}{(wq;q)_k}\\
&=\frac{1}{2\pi i}\oint\sum_{k\ge1}\frac{\mathrm{d}z}{z^{n+1}}\,\frac{w^k(-1)^kp^kq^{\binom{k}{2}}}{(wq;q)_k}\qquad\text{by Cauchy's integral formula}\\
&=\frac{1}{2\pi i}\oint\sum_{k\ge1}\frac{\mathrm{d}w(1-w)^{n-1}}{w^{n+1}}\,\frac{w^k(-1)^{n-k}p^kq^{\binom{k}{2}}}{(wq;q)_k}\\
&=\sum_{k=1}^{n}[w^{n-k}](1-w)^{n-1}\frac{(-1)^{n-k}p^kq^{\binom{k}{2}}}{(wq;q)_k}\\
&=\sum_{k=1}^{n}\sum_{j=0}^{n-1}\binom{n-1}{j}(-1)^j[w^{n-k-j}]\frac{(-1)^{n-k}p^kq^{\binom{k}{2}}}{(wq;q)_k}\\
&=\sum_{k=1}^{n}\sum_{j=0}^{n-1}\binom{n-1}{j}(-1)^{n-k-j}p^kq^{\binom{k}{2}}\begin{bmatrix}n-j-1\\k-1\end{bmatrix}_q q^{n-k-j}\qquad\text{the known expansion of the denominator}\\
&=p\sum_{j=0}^{n-1}\binom{n-1}{j}q^{n-j-1}(-1)^{n-j-1}\sum_{k=0}^{n-j-1}(-1)^kp^kq^{\binom{k}{2}}\begin{bmatrix}n-j-1\\k\end{bmatrix}_q
\end{aligned}$$

$$= p\sum_{j=0}^{n-1}\binom{n-1}{j}q^{n-j-1}(-1)^{n-j-1}(p;q)_{n-j-1} \qquad \text{the sum is known as Rothe's sum}$$

$$= p\sum_{j=0}^{n-1}\binom{n-1}{j}q^{j}(-1)^{j}(p;q)_{j}.$$

Is there a more direct way to prove this formula?

Here is an example for $n = 3$; the words enjoying the restricted growth property are $111, 112, 121, 122, 123$, and they appear with probabilities p^3, p^3q, p^3q, p^3q^2, p^3q^3. And

$$\begin{aligned} p^3 + p^3q + p^3q + p^3q^2 + p^3q^3 &= p\sum_{j=0}^{2}\binom{2}{j}q^{j}(-1)^{j}(p;q)_{j} \\ &= p\big(1 - 2q(1-p) + q^2(1-p)(1-pq)\big). \end{aligned}$$

For the asymptotic evaluation, we use the following integral representation as in [**2**]:

$$p\sum_{j=0}^{n-1}(-1)^{j}\binom{n-1}{j}q^{j}(p;q)_{j} = \frac{-p}{2\pi i}\int_{\mathcal{C}} q^{z}(p;q)_{z}\frac{\Gamma(n)\Gamma(-z)}{\Gamma(n-z)}\mathrm{d}z.$$

Here, $\mathcal{C}$ enclosed the poles $0, 1, \dots, n-1$ and no others, and the interpretation of $(p;q)_z$ is

$$(p;q)_z = \frac{(p;q)_\infty}{(pq^z;q)_\infty}.$$

For the readers' convenience we note that $n! = \Gamma(n+1)$, and thus

$$\frac{\Gamma(n)\Gamma(-z)}{\Gamma(n-z)} = \frac{\Gamma(n)}{(n-z-1)(n-z-2)\cdots(-z)} = \frac{(-1)^n(n-1)!}{z(z-1)\cdots(z+1-n)}.$$

Furthermore, the residue of this expression at $z = k$ is

$$\frac{(-1)^n(n-1)!}{k(k-1)\cdots 1\cdot(-1)\cdots(k+1-n)} = \frac{(-1)^{k-1}(n-1)!}{k!(n-1-k)!}.$$

To get asymptotics, we extend the contour of integration and have to consider the residues at the extra poles of

$$\frac{pq^z(p;q)_\infty}{(1-pq^z)(pq^{z+1};q)_\infty}\,\frac{\Gamma(n)\Gamma(-z)}{\Gamma(n-z)}.$$

The poles with largest real part leading to the dominant contribution are at

$$z = -\frac{\log p}{\log q} + \frac{2\pi i k}{\log q}, \quad \text{for} \quad k \in \mathbb{Z}.$$

For $k = 0$ we get the interesting term, and the others define a small fluctuation around this value. We find:

$$\frac{pq^{-\frac{\log p}{\log q}+\frac{2\pi ik}{\log q}}(p;q)_\infty}{L(pq^{1-\frac{\log p}{\log q}-\frac{2\pi ik}{L}};q)_\infty}\frac{\Gamma(n)\Gamma\Big(\frac{\log p}{\log q}+\frac{2\pi ik}{L}\Big)}{\Gamma\Big(n+\frac{\log p}{\log q}+\frac{2\pi ik}{L}\Big)}$$

$$=\frac{(p;q)_\infty}{L(q^{1-\frac{2\pi ik}{L}};q)_\infty}\frac{\Gamma(n)\Gamma\Big(\frac{\log p}{\log q}+\frac{2\pi ik}{L}\Big)}{\Gamma\Big(n+\frac{\log p}{\log q}+\frac{2\pi ik}{L}\Big)}\sim\frac{(p;q)_\infty\Gamma\Big(\frac{\log p}{\log q}+\frac{2\pi ik}{L}\Big)}{L(q;q)_\infty}n^{-\frac{\log p}{\log q}-\frac{2\pi ik}{L}}.$$

The term $k = 0$ leads to

$$\frac{(p;q)_\infty\Gamma\Big(\frac{\log p}{\log q}\Big)}{L(q;q)_\infty}n^{-\frac{\log p}{\log q}}$$

and the other ones to $n^{-\frac{\log p}{\log q}}\Phi(\log_Q n)$, where $\Phi(x)$ is a 1-periodic function with mean zero. Note that $pq^{-t\frac{\log p}{\log q}+\frac{2\pi ik}{\log q}} = 1$, which was used in these computations.

Acknowledgments. Thanks are due to a referee who spotted a fundamental error in an earlier version. This material is based upon work supported by the National Research Foundation of South Africa under grant number 2053748.

REFERENCES

1. G. Andrews, R. Askey, R. Roy: *Special Functions.* Volume **71** of *Encyclopedia Math. Appl.,* Cambridge University Press, 1999.
2. P. Flajolet, R. Sedgewick: *Mellin transforms and asymptotics*: *Finite differences and Rice's integrals.* Theoret. Comput. Sci,, **144** (1995), 101–124.
3. A. Knopfmacher, T. Mansour, S. Wagner: *Records in set partitions.* Electron. J. Combin., **17** (2010), R109.
4. H. Prodinger: *Combinatorics of geometrically distributed random variables*: *Left-to-right maxima.* Discrete Math., **153** (1996), 253–270.
5. H. Prodinger: *Records in geometrically distributed words*: *Sum of positions.* Appl. Anal. Discrete Math., **2** (2008), 234–240.
6. D. Stanton, D. White: *Constructive Combinatorics.* Springer, 1986.

91052 Erlangen,
Germany,
E-mail: olikamilla@gmail.com

(Received August 16, 2010)
(Revised February 3, 2011)

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