

The average position of the first maximum in a sample of geometric random variables

Margaret Archibald, Arnold Knopfmacher

► **To cite this version:**

Margaret Archibald, Arnold Knopfmacher. The average position of the first maximum in a sample of geometric random variables. 2007 Conference on Analysis of Algorithms, AofA 07, 2007, Juan les Pins, France. pp.295-306. hal-01184771

HAL Id: hal-01184771

<https://hal.inria.fr/hal-01184771>

Submitted on 17 Aug 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Analysis of the total costs for variants of the Union-Find algorithm

The average position of the first maximum in a sample of geometric random variables

Margaret Archibald¹ and Arnold Knopfmacher² †

¹ *Laboratory of Foundational Aspects of Computer Science, Department of Mathematics and Applied Mathematics, University of Cape Town, Rondebosch, 7701, South Africa*

² *The John Knopfmacher Centre for Applicable Analysis and Number Theory, Department of Mathematics, University of the Witwatersrand, P. O. Wits, 2050, Johannesburg, South Africa*
margaret.archibald@uct.ac.za, arnold.knopfmacher@wits.ac.za

received 14 Oct 1998, revised 19th January 2008, accepted tomorrow.

We consider samples of n geometric random variables $(\Gamma_1, \Gamma_2, \dots, \Gamma_n)$ where $\mathbb{P}\{\Gamma_j = i\} = pq^{i-1}$, for $1 \leq j \leq n$, with $p + q = 1$. The parameter we study is the position of the first occurrence of the maximum value in a such a sample. We derive a probability generating function for this position with which we compute the first two (factorial) moments. The asymptotic technique known as Rice's method then yields the main terms as well as the Fourier expansions of the fluctuating functions arising in the expected value and the variance.

Keywords: Geometric random variable, Rice's method

1 Introduction

We consider samples of n geometric random variables $(\Gamma_1, \Gamma_2, \dots, \Gamma_n)$ where $\mathbb{P}\{\Gamma_j = i\} = pq^{i-1}$, for $1 \leq j \leq n$, with $p + q = 1$. Such samples have been studied in their own right (see [Pro96] and [KP04] and those papers listed below), as well as with reference to skiplists (for example, see [KP94] and [LP06a]) and probabilistic counting algorithms (see [KP93], among others). In recent years, questions relating to the maximum value of such a sample have attracted quite a lot of attention. In particular, the expectation and distribution of the maximum value, and the probability of a single maximum have been dealt with by various authors in papers [Pro91], [Sol94], [SR90], [ESS93], [BSW94] and [BES95]. Thereafter, Kirschenhofer and Prodinger ([KP96]) studied the expectation and variance of the number of maxima, as well as the probability of exactly m maxima in a random sample, for a fixed $m \geq 1$. Further papers on the topic of how many maxima occur in a geometric sample are listed in [LP06b].

However, the position of the maximum values has not previously been studied. In this paper we study the mean and variance of the position at which the maximum value first occurs. As in the case of the

†This material is based upon work supported by the National Research Foundation under grant number 2053740

previous statistics studied for maxima of geometric samples, the mean and variance do not converge as $n \rightarrow \infty$ but instead exhibit small fluctuations. Our methodology follows that of [KP96], the asymptotic technique that yields the main terms and Fourier expansions of the fluctuating functions very effectively is ‘‘Rice’s method’’, as surveyed in [FS95].

For convenience we denote by position the number of places *before* the maximum first occurs. (To count the first maximum as well in the position, just add 1 to the mean value below; the variance remains unchanged.)

Theorem 1 *The average position E_n of the first (left-most) occurrence of the maximum in a sample of geometric random variables is given by*

$$E_n = n \left(\frac{1}{L} + \frac{1}{1-Q} + \delta_{E1}(\log_Q n) \right) + \frac{1}{L} + \frac{Q}{1-Q} - \delta_{E2}(\log_Q n) + o(1) \tag{1.1}$$

where $Q = \frac{1}{q}$; $L = \log Q$, $\chi_k = 2k\pi i/L$,

$$\delta_{E1}(x) := \frac{1}{L} \sum_{k \neq 0} \chi_k \Gamma(-1 - \chi_k) e^{2k\pi i x} \tag{1.2}$$

and

$$\delta_{E2}(x) := \frac{1}{2L} \sum_{k \neq 0} \chi_k (1 + \chi_k) (\chi_k - 2) \Gamma(-1 - \chi_k) e^{2k\pi i x}. \tag{1.3}$$

We remark that $\delta_{E1}(x)$ and $\delta_{E2}(x)$ are continuous periodic functions of period 1, mean zero and of small amplitude. The same holds for $\delta_v(x)$ below.

Theorem 2 *The variance of the position of the first occurrence of the maximum in a sample of geometric random variables is given by*

$$Var = n^2 \left(\frac{1+Q}{2L(Q-1)} - \frac{1}{L^2} \right) + n \left(\frac{Q}{(Q-1)^2} - \frac{2}{L^2} + \frac{Q+1}{2L(Q-1)} \right) + \frac{Q}{(Q-1)^2} - \frac{1}{L^2} + o(1). \tag{1.4}$$

There are also negligibly small contributions from the fluctuating terms. These are calculated only for the dominant term of the variance, and are given by

$$n^2([\delta_{E1}^2]_0 + \delta_v(\log_Q n))$$

where

$$[\delta_{E1}^2]_0 = \frac{Q}{2L(1-Q)} + \frac{Q}{(Q-1)^2} + \frac{2}{L} \sum_{l \geq 2} \frac{l(-1)^l}{(Q^l - 1)(l+1)(l-1)} + \frac{\log 2}{L} - \frac{1}{L^2}$$

and

$$\delta_v(x) := \frac{-1}{L} \sum_{k \neq 0} V_k e^{2k\pi i x}$$

with

$$V_k := \frac{\Gamma(-2 - \chi_k)}{L(Q-1)} \left(-L(Q+1) - 4\chi_k(L+Q-1) + \chi_k^2(2-2Q+QL-L) \right) - \sum_{l \geq 1} \frac{l(-1)^l}{(l+1)!} (l - \chi_k) \Gamma(l-1 - \chi_k) \frac{Q^l + 1}{Q^l - 1}.$$

In particular, since the variance is of order n^2 and the mean of order n , we note that the distribution of the position about the mean is not concentrated. Note also that $[\delta_{E_1}^2]_0$ is an extremely small quantity, typically of size 10^{-12} .

2 Method

We start off by expressing all possible patterns of geometric words in a symbolic manner in which the first occurrence of the maximum value k is highlighted. This gives the expression

$$\{1, 2, \dots, k-1\}^* k \{1, 2, \dots, k\}^*.$$

To translate this into a probability generating function, let z mark each letters and let u mark only those letters which lie to the left of the first maximum value k . Then the generating function corresponding to $\{1, 2, \dots, k-1\}$ is

$$pzu + pqzu + pq^2zu + \dots + pq^{k-2}zu = zu(1 - q^{k-1}).$$

Similarly, for the letters which occur right of the first k , $\{1, 2, \dots, k\} \equiv z(1 - q^k)$ and thus

$$\bigcup_{k \geq 1} \{1, 2, \dots, k-1\}^* k \{1, 2, \dots, k\}^* \equiv \sum_{k \geq 1} \frac{1}{1 - zu(1 - q^{k-1})} pq^{k-1} z \frac{1}{1 - z(1 - q^k)}.$$

Hence we have found the bivariate generating function

$$F(z, u) := \sum_{k \geq 1} \frac{pq^{k-1}z}{(1 - zu(1 - q^{k-1}))(1 - z(1 - q^k))}. \quad (2.1)$$

2.1 The expected value

The generating function for the expected position of the first maximum in a sample of n letters is given by $\frac{\partial}{\partial u} F(z, u)|_{u=1}$. Using (2.1) this leads to

$$\frac{\partial}{\partial u} F(z, u)|_{u=1} = \sum_{k \geq 1} \frac{pq^{k-1}(1 - q^{k-1})z^2}{(1 - z(1 - q^k))(1 - z(1 - q^{k-1}))^2}.$$

We decompose this into partial fractions so that the coefficients of z^n can be obtained:

$$\begin{aligned} [z^n] \frac{\partial}{\partial u} F(z, u)|_{u=1} &= [z^n] \sum_{k \geq 1} \left[\frac{q^{1-k} - 1}{p(1 - z(1 - q^k))} - \frac{q^{1-k} - 2 + q}{p(1 - z(1 - q^{k-1}))} - \frac{1}{(1 - z(1 - q^{k-1}))^2} \right] \\ &= \sum_{k \geq 1} \left[\frac{q^{1-k} - 1}{p} (1 - q^k)^n - \frac{q^{1-k} - 2 + q}{p} (1 - q^{k-1})^n - (n+1)(1 - q^{k-1})^n \right] \\ &= \frac{1}{p} \sum_{k \geq 2} [(q^{1-k} - 1)(1 - q^k)^n - (q^{1-k} + q + n - 1)(1 - q^{k-1})^n], \end{aligned}$$

since the $k = 1$ term is zero. The binomial theorem is now used to expand certain factors, bearing in mind that in order for the sums on k to converge, we need to remove the terms of the binomial expansion where $i = 0$ or $i = 1$ and treat them separately. Thus

$$[z^n] \frac{\partial}{\partial u} F(z, u)|_{u=1} = \frac{1}{p} \sum_{k \geq 2} (q^{1-k} - 1) \left[\sum_{i=2}^n \binom{n}{i} (-q^k)^i + 1 - nq^k \right] - \frac{1}{p} \sum_{k \geq 2} (q^{1-k} + q + n - 1) \left[\sum_{i=2}^n \binom{n}{i} (-q^{k-1})^i + 1 - nq^{k-1} \right].$$

Interchanging the order of summation and performing the inner sums over k leads to

$$[z^n] \frac{\partial}{\partial u} F(z, u)|_{u=1} = \sum_{i=2}^n \binom{n}{i} (-1)^i \frac{q^{i-1}(q^i - 1 + nq^i - nq)}{(1 - q^{i-1})(1 - q^i)} + \frac{qn(n-1)}{p}$$

which we express in the form (recall $Q = \frac{1}{q}$)

$$[z^n] \frac{\partial}{\partial u} F(z, u)|_{u=1} = - \underbrace{\sum_{i=2}^n \binom{n}{i} (-1)^i \frac{1}{Q^{i-1} - 1}}_{\alpha} - n \underbrace{\sum_{i=2}^n \binom{n}{i} (-1)^i \frac{1}{Q^i - 1}}_{\beta} + \frac{n(n-1)}{Q-1}. \quad (2.2)$$

The terms α and β are typical expressions to which ‘Rice’s method’ can be applied, due to the presence of the binomial coefficient inside an alternating sum. First we deal with α , and the function involved for Rice’s method is $f(z) = \frac{1}{Q^{z-1}-1}$. Rice’s method also uses the kernel function

$$[n; z] := \frac{(-1)^{n-1} n!}{z(z-1) \cdots (z-n)} = \frac{\Gamma(n+1)\Gamma(-z)}{\Gamma(n+1-z)}.$$

Now we consider the function $[n; z]f(z)$ which has a double pole at $z = 1$ and simple poles at $z = 0$ and $z = \chi_k + 1$ for $k \neq 0$ where $k \in \mathbb{Z}$. The first two contribute to the main term (asymptotically as $n \rightarrow \infty$) and the others contribute to the fluctuations. The residue of the double pole at $z = 1$ can be obtained from the expansions (set $\varepsilon = z - 1$)

$$f(z) \sim \frac{1}{\varepsilon L(1 + \frac{\varepsilon L}{2})} \quad \text{and} \quad [n; z-1] \sim n \left(\frac{1}{\varepsilon} - 1 + H_n \right),$$

where $H_m = \sum_{i=1}^m \frac{1}{i}$ is a harmonic number. After taking the coefficient of ε^{-1} in the product of these, we get the asymptotic contribution as $n \rightarrow \infty$ to be

$$\frac{-\gamma n}{L} + \frac{n}{L} + \frac{1}{2L} - \frac{n}{L} \log n + \frac{n}{2}$$

(where \log denotes \log_e). For the simple pole at $z = 0$, we have $f(z) \sim \frac{1}{Q^{-1}-1}$ and $[n; z] \sim \frac{-1}{z}$, which leads to a contribution of $\frac{Q}{1-Q}$. For the sum labelled β , the function for Rice’s method is $f(z) = \frac{1}{Q^z-1}$ but the double pole is now at $z = 0$, so expanding we get:

$$f(z) \sim \frac{1}{zL} \left(1 - \frac{zL}{2} \right) \quad \text{and} \quad [n; z] \sim \frac{-1}{z} (1 + zH_n),$$

which together give an asymptotic contribution of

$$-\frac{n}{2} + \frac{n}{L} \log n + \frac{\gamma n}{L} + \frac{1}{2L}.$$

We also have a simple pole at $z = 1$ which contributes $-\frac{n^2}{Q-1}$. Adding these four terms to the last term in (2.2) simplifies to the main terms in (1.1).

2.2 The fluctuations in the expected value.

The fluctuations in the asymptotic estimates for the sums α and β arise from the simple poles at $z = \chi_k + 1$ and $z = \chi_k$ ($k \neq 0$) and are given by

$$-\frac{n}{L} \sum_{k \neq 0} \Gamma(-1 - \chi_k) e^{2k\pi i \log_Q n} + \frac{1}{2L} \sum_{k \neq 0} \chi_k (\chi_k + 1) \Gamma(-1 - \chi_k) e^{2k\pi i \log_Q n},$$

and

$$-\frac{n}{L} \sum_{k \neq 0} \Gamma(-\chi_k) e^{2k\pi i \log_Q n} - \frac{1}{2L} \sum_{k \neq 0} \chi_k (1 - \chi_k) \Gamma(-\chi_k) e^{2k\pi i \log_Q n}$$

respectively. We combine these together to produce the functions $\delta_{E1}(x)$ and $\delta_{E2}(x)$ given in (1.2) and (1.3).

2.3 The variance

Using the same generating function as in (2.1), we need to compute the second partial derivative with respect to u to find the variance. Firstly,

$$\frac{\partial^2}{\partial u^2} F(z, u) = \sum_{k \geq 1} \frac{2pq^{k-1} z^3 (1 - q^{k-1})^2}{(1 - z(1 - q^k))(1 - zu(1 - q^{k-1}))^3}.$$

Next substituting $u = 1$ and using partial fractions we obtain

$$\begin{aligned} \frac{\partial^2}{\partial u^2} F(z, 1) = 2 \sum_{k \geq 1} \left[\frac{(q^{k-1} - 1)^2}{p^2 q^{2k-2} (1 - z(1 - q^k))} - \frac{1}{(1 - z(1 - q^{k-1}))^3} - \frac{q^{1-k} (1 - 3q^{k-1} + 2q^k)}{p(1 - z(1 - q^{k-1}))^2} \right. \\ \left. - \frac{q^{2-2k} (1 + 3q^{2k-2} - 3q^{k-1} + q^k - 3q^{2k-1} + q^{2k})}{p^2 (1 - z(1 - q^{k-1}))} \right]. \end{aligned}$$

Extracting the coefficient of z^n in this and then simplifying leads to

$$\begin{aligned}
 [z^n] \frac{\partial^2}{\partial u^2} F(z, 1) &= 2 \sum_{k \geq 1} \left[\frac{(q^{k-1} - 1)^2}{p^2 q^{2k-2}} (1 - q^k)^n - \frac{1 + 3q^{2k-2} - 3Q^{k-1} + q^k - 3q^{2k-1} + q^{2k}}{p^2 q^{2k-2}} (1 - q^{k-1})^n \right. \\
 &\quad \left. - \binom{n+2}{n} (1 - q^{k-1})^n - \frac{1 - 3q^{k-1} + 2q^k}{pq^{k-1}} (n+1)(1 - q^{k-1})^n \right] \\
 &= \underbrace{\frac{2}{(Q-1)^2} \sum_{i=3}^n \binom{n}{i} (-1)^i \frac{Q^i - 1}{1 - Q^{i-2}}}_{(a)} + \underbrace{\frac{4Q}{(Q-1)^2} \sum_{i=3}^n \binom{n}{i} (-1)^i \frac{Q^i - 1}{Q^{i-1} - 1}}_{(b)} \quad (2.3) \\
 &\quad + \underbrace{\frac{2n}{Q-1} \sum_{i=3}^n \binom{n}{i} (-1)^i \frac{Q^i}{1 - Q^{i-1}}}_{(c)} + \underbrace{\left(\frac{n(3Q-1)}{Q-1} - n^2 \right) \sum_{i=3}^n \binom{n}{i} (-1)^i \frac{Q^i}{Q^i - 1}}_{(d)} \\
 &\quad + \underbrace{\frac{2Q^2}{(Q-1)^2} - \frac{2nQ(Q+1)}{(Q-1)^2} - \frac{n^2Q(3Q^2 - 7Q - 6)}{2(Q-1)^2(Q+1)} + \frac{n^3Q(2Q^2 - 2Q - 1)}{(Q-1)^2(Q+1)} - \frac{n^4Q^2}{2(Q^2 - 1)}}_{(e)}.
 \end{aligned}$$

Now Rice's method can again be used to approximate the alternating sums. First we deal with part (a) by considering the poles of the function $[n; z]f(z)$ with $f(z) = \frac{Q^z - 1}{1 - Q^{z-2}}$. This function has a double pole at $z = 2$ and a simple pole at $z = 1$. For the double pole we expand to two terms about $\varepsilon = 0$ where $\varepsilon = z - 2$. This gives

$$f(z) = \frac{Q^{\varepsilon+2} - 1}{1 - Q^\varepsilon} \sim \frac{Q^2(1 + \varepsilon L + \frac{1}{2}\varepsilon^2 L^2) - 1}{-\varepsilon L(1 + \frac{1}{2}\varepsilon L)} \sim \frac{-1 + Q^2 + \frac{1}{2}\varepsilon L - Q^2 \frac{1}{2}\varepsilon L + Q^2 \varepsilon L}{-\varepsilon L}$$

and

$$[n; z - 2] \sim \frac{-n(n-1)}{4} \left(\frac{2}{\varepsilon} - 3 + 2H_{n-2} \right).$$

Since $H_{n-2} \sim \gamma + \log n - \frac{3}{2n} - \frac{13}{12n^2}$ as $n \rightarrow \infty$, the asymptotic residue is

$$\begin{aligned}
 [\varepsilon^{-1}] &\frac{2}{(Q-1)^2} \frac{n(n-1)}{4\varepsilon L} \left(\frac{2}{\varepsilon} - 3 + 2\gamma + 2\log n - \frac{3}{n} - \frac{13}{6n^2} \right) \left(-1 + Q^2 + \frac{1}{2}\varepsilon L - Q^2 \frac{1}{2}\varepsilon L + Q^2 \varepsilon L \right) \\
 &= \frac{1}{2L(Q-1)^2} \left(2n^2 \log n(Q^2 - 1) + n^2(Q^2 L + L + 3 - 3Q^2 + 2\gamma Q^2 - 2\gamma) - 2n \log n(Q^2 - 1) \right. \\
 &\quad \left. - n(Q^2 L + L + 6 - 6Q^2 + 2\gamma Q^2 - 2\gamma) + \frac{5}{6}(Q^2 - 1) \right).
 \end{aligned}$$

For the simple pole we let $\varepsilon = z - 1$ and get a residue of $\frac{2nQ}{(Q-1)^2}$.

Now we look at the alternating sum labelled (b) which has a double pole at $z = 1$ and a simple pole at $z = 2$. By expanding around $\varepsilon := z - 1$ we find that the residue for the double pole is:

$$n \log n \frac{4Q}{L(Q-1)} + \frac{4nQ}{L(Q-1)^2} \left((Q-1)(\gamma - 1) + QL - \frac{L(Q-1)}{2} \right) - \frac{2Q}{L(Q-1)}.$$

For the simple pole the residue is

$$-\frac{2n^2Q(Q+1)}{(Q-1)^2} + \frac{2nQ(Q+1)}{(Q-1)^2}.$$

For the sum in (c) the double pole is also at $z = 1$ but this time there are two simple poles, at $z = 0$ and $z = 2$. For the double pole, we compute a residue of

$$-\frac{2Qn^2 \log n}{L(Q-1)} - \frac{2n^2Q}{L(Q-1)} \left(\gamma - 1 + \frac{L}{2} \right) + \frac{nQ}{L(Q-1)} + \frac{Q}{6L(Q-1)}.$$

The simple poles contribute $\frac{-2nQ}{(Q-1)^2}$ and $\frac{n^3Q^2}{(Q-1)^2} - \frac{n^2Q^2}{(Q-1)^2}$ respectively. The double pole for (d) is at $z = 0$, and it contributes a residue of

$$n^2 \log n \frac{1}{L} + n^2 \left(\frac{\gamma}{L} + \frac{1}{2} \right) + n \log n \frac{1-3Q}{L(Q-1)} + n \left(\frac{1}{2L} + \left(\gamma + \frac{L}{2} \right) \frac{1-3Q}{L(Q-1)} \right) - \frac{1}{12L} + \frac{1-3Q}{2L(Q-1)}.$$

The simple poles at $z = 1$ and $z = 2$ contribute

$$-\frac{n^3Q}{Q-1} + \frac{n^2Q(3Q-1)}{(Q-1)^2} \quad \text{and} \quad \frac{n^4Q^2}{2(Q^2-1)} + \frac{n^3Q^2(1-2Q)}{(Q-1)^2(Q+1)} + \frac{n^2Q^2(3Q-1)}{2(Q-1)^2(Q+1)}$$

respectively. The higher order terms in all the above sums cancel out with the terms in (e). The cancellations leave us with a second factorial moment of

$$n^2 \frac{2L+3-4Q+Q^2}{2L(Q-1)^2} + n \frac{8QL-2L-5Q^2+4Q+1}{2L(Q-1)^2} + \frac{2Q^2L-3Q^2-1+4Q}{L(Q-1)^2}.$$

To calculate the dominant terms of the variance we must add to this the expected value given in (1.1) and subtract the square of the expected value. In squaring the expected value, a further possible contribution to the constant term of the variance will arise from the product of the terms of order n and of order $1/n$ in the mean. However, computations show that the coefficient of the term of order $1/n$ in the mean is actually zero in this case. Squaring the expectation yields

$$n^2 \frac{2L+3-4Q+Q^2}{2L(Q-1)^2} + n \frac{8QL-2L-5Q^2+4Q+1}{2L(Q-1)^2} + \frac{2Q^2L-3Q^2-1+4Q}{L(Q-1)^2} \\ + \left(n \left(\frac{1}{L} + \frac{1}{1-Q} \right) + \frac{1}{L} + \frac{Q}{1-Q} \right)^2 - \left(n \left(\frac{1}{L} + \frac{1}{1-Q} \right) + \frac{1}{L} + \frac{Q}{1-Q} \right)^2,$$

which simplifies to the expression given in (1.4).

But we also get a further nonzero contribution to the n^2 term when we square the fluctuations of the expected value. To calculate this additional term, we make use of a paper by Prodinger ([Pro04]) which explains how to calculate this 'zereth' Fourier coefficient of the square. We consider the function

$$\delta_{E1}(x) = \frac{1}{L} \sum_{k \neq 0} \chi_k \Gamma(-1 - \chi_k) e^{2k\pi i x}. \quad (2.4)$$

The same function appears on page 5 of [Pro04] (where it is called $\sigma(x)$), but in that paper it is assumed that $Q = 2$ whereas we deal with the more general geometric case. We find two expressions for

$$I_1 := \frac{1}{2\pi i} \int_{\frac{1}{2}-i\infty}^{\frac{1}{2}+i\infty} F(z) dz, \quad \text{where} \quad F(z) := \frac{-L}{e^{Lz}-1} z^2 \Gamma(-1-z) \Gamma(-1+z).$$

Shifting left and collecting residues gives one form for I_1 , and summing the negative residues right of the line $\Re = \frac{1}{2}$ gives another. Equating these two expressions leaves us with

$$\sum_{k \neq 0} \chi_k(-\chi_k) \Gamma(-1-\chi_k) \Gamma(-1+\chi_k) = \frac{L}{2(1-Q)} + \frac{QL^2}{(Q-1)^2} + 2L \sum_{l \geq 2} \frac{l(-1)^l}{(Q^l-1)(l+1)(l-1)} + L \log 2 - \frac{L}{2} - 1.$$

This is exactly what we need, since squaring the function in (1.2) will give us the left-hand side, except for a factor of L^2 . Thus dividing through by L^2 gives the result:

$$[\delta_{E1}^2]_0 = \frac{Q}{2L(1-Q)} + \frac{Q}{(Q-1)^2} + \frac{2}{L} \sum_{l \geq 2} \frac{l(-1)^l}{(Q^l-1)(l+1)(l-1)} + \frac{\log 2}{L} - \frac{1}{L^2}.$$

This is the ‘zeroth’ coefficient of the Fourier series – which when multiplied by n^2 contributes to the main term in the variance.

2.4 As $Q \rightarrow 1$

It is well known that as $q \rightarrow 1$ samples of geometric variables tend in behaviour to that of a permutation of n numbers. It is interesting to let $Q \rightarrow 1$ in our results and compare this to the permutations case. For permutations, the average number of places before the maximum and the second moment are

$$\frac{1}{n} \sum_{k=1}^n (k-1) = \frac{n-1}{2} \quad \text{and} \quad \frac{1}{n} \sum_{k=1}^n (k-1)^2 = \frac{n^2}{3} - \frac{n}{2} + \frac{1}{6}$$

respectively. Thus the variance is $\frac{n^2}{12} - \frac{1}{12}$, which is also obtained by taking the limit as $Q \rightarrow 1$ of the main term in the variance in Theorem 2.

2.5 The fluctuations in the variance.

It is possible to compute the fluctuations in the variance explicitly; we now do this in the case of the fluctuations involving the largest term n^2 of the variance. To do this we need to look at two possible sources, namely fluctuations occurring in the second factorial moment calculations and those fluctuations from the expected value which, when squared, will involve an n^2 term.

We start with the latter. (A simpler example of this type was dealt with in [Pro04] and called $[\sigma_0^2]_k$.) When the expectation is squared, so are its fluctuations and as we saw above the complex powers in some of these terms cancel and contribute to $[\delta_{E1}^2]_0$. However, most of the pairs of complex powers do not cancel and lead to terms of the form $\chi_j \chi_k \Gamma(-1-\chi_k) \Gamma(-1-\chi_j)$ for $j \neq 0$ and $j \neq k$. Consider the function

$$F(z) := \frac{L}{e^{Lz}-1} z(-z+\chi_k) \Gamma(-\chi_k-1+z) \Gamma(-1-z).$$

The residue at χ_j is

$$-\frac{4j(k-j)\pi^2}{L^2}\Gamma(-1-\chi_j)\Gamma(-1-\chi_{k-j}) = \chi_j\chi_{k-j}\Gamma(-1-\chi_j)\Gamma(-1-\chi_{k-j}).$$

The residue at both $z = 0$ and $z = \chi_k$ is $\chi_k\Gamma(-1-\chi_k)$. Again we write I_1 in two different ways and then compare them. For the first, we shift the line of integration to the left and add the above residues, giving

$$I_1 = \frac{1}{2\pi i} \int_{(-\frac{1}{2})} F(z)dz + 2\chi_k\Gamma(-1-\chi_k) + \sum_{j \neq 0; j \neq k} \chi_j\chi_{k-j}\Gamma(-1-\chi_k)\Gamma(-1-\chi_{k-j}).$$

We then rewrite $\frac{1}{e^{Lz}-1}$ as $-1 - \frac{1}{e^{-Lz}-1}$ and change the variable z to $z + \chi_k$. This means that

$$I_1 = LI_2 - \frac{1}{2\pi i} \int_{(-\frac{1}{2})} F(-z)dz + 2\chi_k\Gamma(-1-\chi_k) + \sum_{j \neq 0; j \neq k} \chi_j\chi_{k-j}\Gamma(-1-\chi_k)\Gamma(-1-\chi_{k-j}),$$

where the integral on the right hand side is also equal to I_1 and

$$I_2 := \frac{1}{2\pi i} \int_{(-\frac{1}{2})} z(z + \chi_k)\Gamma(-1+z)\Gamma(-1-z-\chi_k).$$

By shifting the line of integration to the right, the residues at $z = 1$ and $z = -\chi_k + l; l \geq 1$ give us

$$I_2 = -(1 + \chi_k)\Gamma(-2 - \chi_k) - \sum_{l \geq 1} \frac{(-1)^l l}{(l+1)!} (l - \chi_k)\Gamma(l - 1 - \chi_k).$$

For the other version of I_1 , we shift the line of integration to the right and take negative residues at $z = \chi_k + 1$ and $z = l, l \geq 1$ so that

$$I_1 = \frac{-L}{1-Q}(1 + \chi_k)\Gamma(-2 - \chi_k) + \sum_{l \geq 1} \frac{Ll(-1)^l}{(l+1)!(Q^l-1)} (l - \chi_k)\Gamma(l - 1 - \chi_k).$$

Equating these two forms for I_1 leads to the contribution to the order n^2 fluctuations from the squaring of the expected value, namely

$$\frac{1}{L}(1 + \chi_k)\Gamma(-2 - \chi_k)\frac{Q+1}{Q-1} + \frac{1}{L} \sum_{l \geq 1} \frac{(-1)^l l}{(l+1)!} (l - \chi_k)\Gamma(l - 1 - \chi_k)\frac{Q^l+1}{Q^l-1} - \frac{2\chi_k}{L^2}\Gamma(-1 - \chi_k). \quad (2.5)$$

The other contributions come from the terms in (2.3) which produce an order n^2 fluctuation – i.e., (a), (c), and part of (d). The residues from (a) and (c) are

$$\frac{2n^2(Q+1)}{L(1-Q)} \sum_{k \neq 0} \Gamma(-2 - \chi_k)e^{2k\pi i \log_Q n} \quad \text{and} \quad \frac{2n^2Q}{L(1-Q)} \sum_{k \neq 0} \Gamma(-1 - \chi_k)e^{2k\pi i \log_Q n}. \quad (2.6)$$

The quadratic part of (d) gives

$$\frac{-n^2}{L} \sum_{k \neq 0} \Gamma(-\chi_k)e^{2k\pi i \log_Q n}. \quad (2.7)$$

We add together (2.5), (2.6), and (2.7) to obtain $n^2\delta_v(x)$, as given in Theorem 2.

References

- [BES95] Y. Baryshnikov, B. Eisenberg, and G. Stengle. A necessary and sufficient condition for the existence of the limiting probability of a tie for first place. *Statistics and Probability Letters*, 23:203–209, 1995.
- [BSW94] J.J.A.M. Brands, F.W. Steutel, and R.J.G. Wilms. On the number of maxima in a discrete sample. *Statistics and Probability Letters*, 20:209–217, 1994.
- [ESS93] B. Eisenberg, G. Stengle, and G. Strang. The asymptotic probability of a tie for first place. *Annals of Applied Probability*, 3:731–745, 1993.
- [FS95] P. Flajolet and R. Sedgewick. Mellin transforms and asymptotics: Finite differences and Rice’s integrals. *Theoretical Computer Science*, 144 (1-2):101–124, 1995.
- [KP93] P. Kirschenhofer and H. Prodinger. A result in order statistics related to probabilistic counting. *Computing*, 51:15–27, 1993.
- [KP94] P. Kirschenhofer and H. Prodinger. The path length of random skip lists. *Acta Informatica*, 31:775–792, 1994.
- [KP96] P. Kirschenhofer and H. Prodinger. The number of winners in a discrete geometrically distributed sample. *Annals in Applied Probability*, 6:687–694, 1996.
- [KP04] A. Knopfmacher and H. Prodinger. The number of descents in samples of geometric random variables. *Mathematics and Computer Science III: Algorithms, Trees, Combinatorics and Probabilities. Series: Trends in Mathematics (Mathematics, Computer Science). Edited by Drmota, M.; Flajolet, P.; Gardy, D.; Gittenberger, B., XV:339–350, 2004.*
- [LP06a] G. Louchard and H. Prodinger. Analysis of a new skip list variant. *DMTCS proc. (MathInfo conference, Nancy, September 2006)*, AG:365–374, 2006.
- [LP06b] G. Louchard and H. Prodinger. The number of elements close to near-records in geometric samples. *Quaestiones Mathematicae*, 29:447–470, 2006.
- [Pro91] E3436 Problem. *American Mathematical Monthly*, 98:366, 1991.
- [Pro96] H. Prodinger. Combinatorics of geometrically distributed random variables: Left-to-right maxima. *Discrete Mathematics*, 153:253–270, 1996.
- [Pro04] H. Prodinger. Periodic oscillations in the analysis of algorithms. *Journal of the Iranian Statistical Society*, 3:251–270, 2004.
- [Sol94] E3436 Solution. *American Mathematical Monthly*, 101:78–80, 1994.
- [SR90] W. Szpankowski and V. Rego. Yet another application of a binomial recurrence. Order statistics. *Computing*, 43(4):401–410, 1990.

