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

Appl. Anal. Discrete Math. **5** (2011), 271–282. doi:10.2298/AADM110817019B

SEPARATION OF THE MAXIMA IN SAMPLES OF GEOMETRIC RANDOM VARIABLES

Charlotte Brennan, Arnold Knopfmacher,
Toufik Mansour, Stephan Wagner

We consider samples of n geometric random variables $\omega_1\,\omega_2\cdots\omega_n$ where $\mathbb{P}\{\omega_j = i\} = pq^{i-1}$, for $1 \le j \le n$, with $p+q=1$. For each fixed integer $d>0$, we study the probability that the distance between the consecutive maxima in these samples is at least d. We derive a probability generating function for such samples and from it we obtain an exact formula for the probability as a double sum. Using Rice's method we obtain asymptotic estimates for these probabilities. As a consequence of these results, we determine the average minimum separation of the maxima, in a sample of n geometric random variables with at least two maxima.

1. INTRODUCTION

We consider samples of n geometric random variables $(\omega_1\,\omega_2\cdots\omega_n)$ where $\mathbb{P}\{\omega_j = i\} = pq^{i-1}$, for $1 \le j \le n$, with $p+q=1$. The combinatorics of n geometrically distributed independent random variables $X_1,\dots,X_n$ has attracted recent interest, especially because of applications to computer science such as skip lists [**3, 11**] and probabilistic counting [**7, 10**].

- *Skip lists* are an alternative to *tries* and *digital search trees.* For each data, a geometric random variable defines the number of pointers that it contributes to the data structure. These pointers are then connected in a specific way that makes access to the data manageable. The analysis leads to parameters that are related to *left-to-right maxima.*

2010 Mathematics Subject Classification. Primary: 60C05. Secondary: 05A15, 05A16.
Keywords and Phrases. Geometric random variables, probability generating functions, maxima, asymptotics.

- *Probabilistic counting* uses hashing and the position of the first digit 1 when reading the binary representation of the hashed value from right to left. This is a geometric random variable with parameter $p = q = \frac{1}{2}$. Thus one has an *urn model*, with urns labelled $1, 2, \ldots$, and the relevant parameter here is the number of non-empty urns (starting with the first urn).

In addition, questions relating to the maximum value of sequences of geometric random variables 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 in various papers, such as [**1**], [**2**], [**4**], [**5**], [**6**] and [**12**]. Thereafter, [**9**] studied the number of maxima, as well as the probability of having exactly m maxima in a random geometric sample, for a fixed $m \geq 1$.

In this paper, we study samples of geometric variables whose maxima are separated by at least $d \geq 1$ values. We obtain the probability generating function and hence asymptotic estimates that a sample has this property. For any d, samples of geometric variables with exactly one maximum are trivially assumed to satisfy the separation condition.

Theorem 1. *The probability generating function $W_d(z)$ of geometric samples with a distance between the maxima at least d is given by*

$$W_d(z) = \sum_{k \geq 1} \frac{pq^{k-1}z}{(1 - z(1 - q^{k-1}))(1 - z(1 - q^{k-1}) - pq^{k-1}z^{d+1}(1 - q^{k-1})^d)}.$$

Proof. Consider a geometric word whose maxima have the value k. We represent this word as follows

$$\boxed{k\text{-}1}^{\,*} \quad k \quad \boxed{k\text{-}1}^{\,*} \quad k \quad \boxed{k\text{-}1}^{\,*} \quad \cdots \quad \boxed{k\text{-}1}^{\,*} \quad k \quad \boxed{k\text{-}1}^{\,*}$$

Here $\boxed{k\text{-}1}$ represents a one letter word consisting of letters from the alphabet $\{1, 2, \ldots, k-1\}$, so the generating function for such a word is $z(1 - q^{k-1})$, and $\boxed{k\text{-}1}^{\,*}$ represents a possibly empty sequence of letters taken from the alphabet $\{1, 2, \ldots, k-1\}$. The generating function for such a sequence is therefore determined as $\frac{1}{1 - z(1 - q^{k-1})}$.

A distance of at least d between the maxima is represented by a sequence $\boxed{k\text{-}1}^{\,d}\,\boxed{k\text{-}1}^{\,*}$ between each pair of consecutive maxima, with generating function for this sequence

$$z^d(1 - q^{k-1})^d + z^{d+1}(1 - q^{k-1})^{d+1} + \cdots = \frac{z^d(1 - q^{k-1})^d}{1 - z(1 - q^{k-1})}.$$

If there are s maxima, $\boxed{k\text{-}1}^{\,d}\,\boxed{k\text{-}1}^{\,*}$ will occur $s - 1$ times. Thus, finally the generating function including the s maximum values k, together with the first sequence $\boxed{k\text{-}1}^*$ which precedes the first maximum and the last sequence $\boxed{k\text{-}1}^*$ that follows the last maximum, is given by

$$(pq^{k-1}z)^s \left(\frac{z^d(1-q^{k-1})^d}{1-z(1-q^{k-1})}\right)^{s-1} \left(\frac{1}{1-z(1-q^{k-1})}\right)^2.$$

We denote this generating function for a word with s maxima equal to k and with a minimum distance d between the maxima by $f_d(z)$. Summing up over all s, the number of maxima, we have

$$\begin{aligned} F_d(z) := \sum_{s\geq 1} f_d(z) &= \sum_{s\geq 1} \frac{(pq^{k-1}z)^s[z^d(1-q^{k-1})^d]^{s-1}}{[1-z(1-q^{k-1})]^{s+1}} \\ &= \frac{pq^{k-1}z}{[1-z(1-q^{k-1})]^2} \frac{1}{1-\dfrac{pq^{k-1}z^{d+1}(1-q^{k-1})^d}{1-z(1-q^{k-1})}} \\ &= \frac{pq^{k-1}z}{[1-z(1-q^{k-1})][1-z(1-q^{k-1})-pq^{k-1}z^{d+1}(1-q^{k-1})^d]}. \end{aligned}$$

Recall, this result was for a specific value k of the maxima. So, finally we need to sum over all values of k to obtain the desired generating function

$$\begin{aligned} W_d(z) :&= \sum_{k\geq 1} F_d(z) \\ &= \sum_{k\geq 1} \frac{pq^{k-1}z}{(1-z(1-q^{k-1}))(1-z(1-q^{k-1})-pq^{k-1}z^{d+1}(1-q^{k-1})^d)}. \end{aligned} \qquad \square$$

We continue in Sections 2 and 3 to find exact and asymptotic estimates for the coefficients of $W_d(z)$. As a consequence of these results, we determine in Section 4 the average minimum separation of the maxima, in a sample of n geometric random variables with at least two maxima.

2. EXACT FORMULAS FOR THE PROBABILITY

In this section, we find an exact formula for the probability that geometric samples have a distance between the maxima at least d, denoted by $w_d(n) := [z^n]W_d(z)$.

Theorem 2. *The probability that a geometric sample of length n has a distance between the maxima at least d, is given by*

$$w_d(n) = \sum_{j=0}^{\lfloor (n-1)/(d+1)\rfloor} \binom{n-dj}{j+1} p^{j+1} \sum_{s=0}^{n-1-j} (-1)^s \binom{n-1-j}{s} \frac{1}{1-q^{j+1+s}}.$$

Proof. For simplicity let $\alpha = 1 - q^{k-1}$ and $\beta = pq^{k-1}$ then we rewrite $W_d(z)$ as follows

$$\begin{aligned}
W_d(z) &= \sum_{k\geq 1} \frac{\beta z}{(1-z\alpha)(1-z\alpha+z^{d+1}\alpha^d\beta)} \\
&= \sum_{k\geq 1} \frac{\beta z}{(1-z\alpha)^2\Big(1-\frac{z^{d+1}\alpha^d\beta}{1-z\alpha}\Big)} = \sum_{k\geq 1}\sum_{j\geq 0} \frac{\beta z\cdot z^{(d+1)j}\alpha^{dj}\beta^j}{(1-z\alpha)^{j+2}} \\
&= \sum_{k\geq 1}\sum_{j\geq 0}\sum_{i\geq 0} \beta^{j+1}\alpha^{dj+i} z^{(d+1)j+1+i}\binom{j+1+i}{i} \\
&= \sum_{k\geq 1}\sum_{j\geq 0}\sum_{i\geq 0} (pq^{k-1})^{j+1}(1-q^{k-1})^{dj+i} z^{(d+1)j+1+i}\binom{j+1+i}{i}.
\end{aligned}$$

This last expression allows us to extract the coefficient of z^n, where $i = n-1-(d+1)j$ as follows

$$\begin{aligned}
w_d(n) = [z^n]W_d(z) &= \sum_{j=0}^{\lfloor (n-1)/(d+1)\rfloor} \sum_{k\geq 1} (pq^{k-1})^{j+1}(1-q^{k-1})^{n-1-j}\binom{n-dj}{j+1} \\
&= \sum_{j=0}^{\lfloor (n-1)/(d+1)\rfloor} \sum_{k\geq 1}\sum_{s=0}^{n-1-j} \binom{n-dj}{j+1} p^{j+1}(q^{k-1})^{j+1}(-1)^s\binom{n-1-j}{s}(q^{k-1})^s \\
&= \sum_{j=0}^{\lfloor (n-1)/(d+1)\rfloor} \sum_{k\geq 1}\sum_{s=0}^{n-1-j} (-1)^s\binom{n-dj}{j+1} p^{j+1}\binom{n-1-j}{s}(q^{j+1+s})^{k-1} \\
&= \sum_{j=0}^{\lfloor (n-1)/(d+1)\rfloor} \sum_{s=0}^{n-1-j} (-1)^s\binom{n-dj}{j+1}\binom{n-1-j}{s}\frac{p^{j+1}}{1-q^{j+1+s}} \\
&= \sum_{j=0}^{\lfloor (n-1)/(d+1)\rfloor} \binom{n-dj}{j+1} p^{j+1}\sum_{s=0}^{n-1-j} (-1)^s\binom{n-1-j}{s}\frac{1}{1-q^{j+1+s}}. \qquad \square
\end{aligned}$$

3. ASYMPTOTICS FOR $w_d(n)$

The series $\sum_{s=0}^{n-1-j}(-1)^s\binom{n-1-j}{s}\frac{1}{1-q^{j+1+s}}$ is an alternating sum containing a binomial coefficient. It is a perfect candidate for "Rice's method" in [**8**], for which we use the lemma below.

Lemma 3. *Let $\mathcal{C}$ be a curve surrounding the points $0, 1, \ldots, n$ in the complex plane and let $f(z)$ be analytic inside $\mathcal{C}$. Then*

$$\sum_{k=0}^{n}(-1)^k\binom{n}{k}f(k) = -\frac{1}{2\pi i}\int_{\mathcal{C}}[n;z]f(z)\mathrm{d}z,$$

where

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

We apply this lemma with $f(k) = \dfrac{1}{1-q^{a+k}}$ to obtain the following formula:

Lemma 4. *For $n > 0$, $0 < q < 1$ and $a > 0$, the identity*

$$\sum_{k=0}^{n}(-1)^k\binom{n}{k}\frac{1}{1-q^{a+k}} = \frac{1}{\log(1/q)}\sum_{k\in\mathbb{Z}}\frac{n!\Gamma(a+2k\pi i/\log(1/q))}{\Gamma(n+1+a+2k\pi i/\log(1/q))}$$

holds.

Proof. The technique for obtaining identities of this type can be found in [**8**, Theorem 2]. First, rewrite the sum as a contour integral:

$$\sum_{k=0}^{n}(-1)^k\binom{n}{k}\frac{1}{1-q^{a+k}} = \frac{1}{2\pi i}\int_{\mathcal{C}}\frac{(-1)^n n!}{z(z-1)\cdots(z-n)}\cdot\frac{1}{1-q^{a+z}}\,\mathrm{d}z,$$

where $\mathcal{C}$ is the rectangle formed by the four lines $\operatorname{Re} z = -\frac{a}{2}$, $\operatorname{Re} z = r > n$, $\operatorname{Im} z = r$, $\operatorname{Im} z = -r$. On the latter three, the integrand is $O(r^{-2})$, so that their contribution vanishes if we let $r \to \infty$. Hence we have

$$\sum_{k=0}^{n}(-1)^k\binom{n}{k}\frac{1}{1-q^{a+k}} = -\frac{1}{2\pi i}\int_{-a/2-i\infty}^{-a/2+i\infty}\frac{(-1)^n n!}{z(z-1)\cdots(z-n)}\cdot\frac{1}{1-q^{a+z}}\,\mathrm{d}z.$$

Now we consider the integral along another rectangle $\mathcal{C}'$ that is formed by the lines $\operatorname{Re} z = -\frac{a}{2}$, $\operatorname{Re} z = -r$, $\operatorname{Im} z = r$, $\operatorname{Im} z = -r$, where $r = \frac{(2\ell+1)\pi i}{\log(1/q)}$ ($\ell \in \mathbb{Z}$) is chosen so as to avoid the poles of $\frac{1}{1-q^{a+z}}$, which are given by $u_k = -\Big(a + \frac{2k\pi i}{\log(1/q)}\Big)$, $k \in \mathbb{Z}$. Then the same argument shows that the contribution of three sides of the rectangle vanishes as $r \to \infty$, which implies that

$$\begin{aligned}\sum_{k=0}^{n}(-1)^k\binom{n}{k}\frac{1}{1-q^{a+k}} &= -\frac{1}{2\pi i}\int_{-a/2-i\infty}^{-a/2+i\infty}\frac{(-1)^n n!}{z(z-1)\cdots(z-n)}\cdot\frac{1}{1-q^{a+z}}\,\mathrm{d}z\\ &= -\sum_{k\in\mathbb{Z}}\operatorname*{Res}_{z=u_k}\frac{(-1)^n n!}{z(z-1)\cdots(z-n)}\cdot\frac{1}{1-q^{a+z}}\\ &= \frac{1}{\log 1/q}\sum_{k\in\mathbb{Z}}\frac{n!\Gamma(-u_k)}{\Gamma(n+1-u_k)}.\end{aligned}$$

□

In the following, we use the abbreviation $Q = 1/q$ and $\chi_k = \frac{2k\pi i}{\log Q}$. We apply the previous lemma with $n-j-1 > 0$ in the place of n and $a = j+1 > 0$ to obtain the double sum

$$w_d(n) = \frac{1}{\log Q} \sum_{j=0}^{\lfloor (n-1)/(d+1) \rfloor} \binom{n-dj}{j+1} (n-j-1)! p^{j+1} \sum_{k \in \mathbb{Z}} \frac{\Gamma(j+1+\chi_k)}{\Gamma(n+1+\chi_k)}.$$

The inner sum is uniformly convergent, so we may interchange the order of summation:

$$w_d(n) = \frac{1}{\log Q} \sum_{k \in \mathbb{Z}} \sum_{j=0}^{\lfloor (n-1)/(d+1) \rfloor} \binom{n-dj}{j+1} (n-j-1)! p^{j+1} \cdot \frac{\Gamma(j+1+\chi_k)}{\Gamma(n+1+\chi_k)}.$$

Assume first that $d = o(n)$, and note that

$$\begin{aligned}\binom{n-dj}{j+1}(n-j-1)! &= \frac{n!}{(j+1)!} \cdot \frac{(n-dj)(n-dj-1)\cdots(n-(d+1)j)}{n(n-1)\cdots(n-j)} \\ &= \frac{n!}{(j+1)!} \prod_{r=0}^{j} \left(1 - \frac{dj}{n-r}\right).\end{aligned}$$

We would like to replace the last product by its asymptotic expansion. To this end, we estimate the sum over all $j \geq \frac{n}{3d}$: first of all, we have the inequality

$$\begin{aligned}\frac{n!}{j!} \left| \frac{\Gamma(j+1+\chi_k)}{\Gamma(n+1+\chi_k)} \right| &= \prod_{r=j+1}^{n} \frac{1}{|1+\chi_k/r|} \leq \frac{1}{|1+\chi_k/(j+1)||1+\chi_k/(j+2)|} \\ &\leq \begin{cases} 1 & k = 0, \\ \frac{(j+1)(j+2)}{|\chi_k|^2} & \text{otherwise} \end{cases}\end{aligned}$$

if $j < n-1$, which is the case for all nonzero summands in our sum. Therefore,

$$\frac{n!}{j!} \sum_{k \in \mathbb{Z}} \frac{\Gamma(j+1+\chi_k)}{\Gamma(n+1+\chi_k)} \ll 1 + \sum_{k \geq 1} \frac{j^2}{k^2} \ll j^2.$$

Hence the contribution of all terms with $j \geq \frac{n}{3d}$ is

$$\begin{aligned}\text{(3.1)}\quad & \frac{1}{\log Q} \sum_{k \in \mathbb{Z}} \sum_{j=\lceil n/3d \rceil}^{\lfloor (n-1)/(d+1) \rfloor} \binom{n-dj}{j+1} (n-j-1)! p^{j+1} \cdot \frac{\Gamma(j+1+\chi_k)}{\Gamma(n+1+\chi_k)} \\ & \qquad \ll \sum_{j \geq n/3d} j p^{j+1} \ll \frac{n}{d} p^{n/(3d)}.\end{aligned}$$

For $j < \frac{n}{3d}$, we can use the following expansion:

$$\prod_{r=0}^{j} \left(1 - \frac{dj}{n-r}\right) = \exp\left(\sum_{r=0}^{j} \log\left(1 - \frac{dj}{n-r}\right)\right) = \exp\left(-\sum_{r=0}^{j} \frac{dj}{n-r} + O\left(\frac{d^2 j^3}{n^2}\right)\right)$$

$$= \exp\left(-\frac{dj(j+1)}{n} + O\left(\frac{d^2j^3}{n^2}\right)\right) = 1 - \frac{dj(j+1)}{n} + O\left(\frac{d^2j^4}{n^2}\right).$$

This gives us

$$\frac{1}{\log Q}\sum_{k\in\mathbb{Z}}\sum_{j<n/3d}\binom{n-dj}{j+1}(n-j-1)!p^{j+1}\cdot\frac{\Gamma(j+1+\chi_k)}{\Gamma(n+1+\chi_k)}$$
$$= \frac{1}{\log Q}\sum_{k\in\mathbb{Z}}\frac{n!}{\Gamma(n+1+\chi_k)}\sum_{j<n/3d}\frac{p^{j+1}}{(j+1)!}\Gamma(j+1+\chi_k)$$
$$\left(1 - \frac{dj(j+1)}{n} + O\left(\frac{d^2j^4}{n^2}\right)\right).$$

We extend the range of the inner summation to the entire interval $[0,\infty)$ at the expense of another exponentially small error term (as before in (3.1)) and use the identities

$$\sum_{j\geq 0}\frac{p^{j+1}}{(j+1)!}\Gamma(j+1+\chi_k) = \begin{cases}\log Q & k=0,\\ (q^{-\chi_k}-1)\Gamma(\chi_k)=0 & \text{otherwise}\end{cases}$$

and

$$\sum_{j\geq 1}\frac{p^{j+1}}{(j-1)!}\Gamma(j+1+\chi_k) = p^2q^{-2-\chi_k}\Gamma(2+\chi_k) = p^2q^{-2}\Gamma(2+\chi_k).$$

Putting everything together, this yields

$$w_d(n) = 1 - \frac{dp^2}{nq^2\log Q}\sum_{k\in\mathbb{Z}}\frac{n!\Gamma(2+\chi_k)}{\Gamma(n+1+\chi_k)} + O\left(\frac{d^2}{n^2}\right).$$

It remains to deal with the sum over k: one has

$$\sum_{k\in\mathbb{Z}}\frac{n!\Gamma(2+\chi_k)}{\Gamma(n+1+\chi_k)} = \sum_{k\in\mathbb{Z}}\Gamma(2+\chi_k)e^{-\chi_k\log n} + O\left(\frac{1}{n}\right)$$

by means of Stirling's approximation, cf. [**9**]. Hence we obtain the following theorem.

Theorem 5. *If $d=o(n)$, then the probability $w_d(n)$ has the asymptotic expansion*

$$w_d(n) = 1 - \frac{dp^2}{nq^2\log Q}\psi(\log_Q n) + O\left(\frac{d^2}{n^2}\right),$$

where ψ is the 1-periodic function given by the Fourier series

$$\psi(x) = \sum_{k\in\mathbb{Z}}\Gamma(2-2k\pi i/\log Q)e^{2k\pi ix}.$$

If, on the other hand, $d \sim \alpha n$ for some $\alpha > 0$, then the sum over j becomes a finite sum, which simplifies matters:

Theorem 6. *If $d \sim \alpha n$, then*

$$w_d(n) \sim \frac{1}{\log Q} \sum_{j<1/\alpha} (1 - j\alpha)^{j+1} \frac{p^{j+1}}{(j+1)!} \psi_j(\log_Q n),$$

where ψ_j is the 1-periodic function

$$\psi_j(x) = \sum_{k\in\mathbb{Z}} \Gamma(j + 1 - 2k\pi i/\log Q)e^{2k\pi ix}.$$

Remark. In the case $d = o(n)$, the fluctuations (i.e. all terms with $k \neq 0$) arising from $\psi(\log_Q n)/n \to 0$ as $n \to \infty$. Whereas, if $d \sim \alpha n$, for any $0 < \alpha \leq 1$, the amplitude of the corresponding fluctuating functions of Theorem 6 remains fixed as $n \to \infty$. These two cases are illustrated in Figure 1 in the case $p = 1/2$.

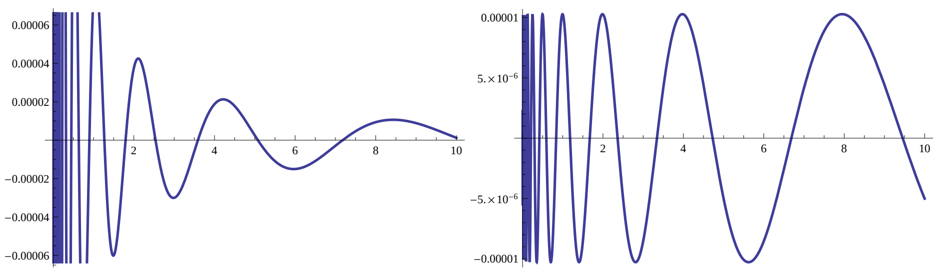

Figure 1. The fluctuating functions of Theorems 5 and 6 for $d = 1$ and $d = n/3$, respectively.

4. THE AVERAGE MINIMUM SEPARATION OF THE MAXIMA

Our asymptotic estimates for $w_d(n)$ allow us to compute the mean value of the minimum separation of the maxima in samples of n geometric random variables. This is given by $m(n) := \sum_{d=1}^{n-2} w_d(n)$. Now

$$\begin{aligned}\sum_{d=1}^{n-2} w_d(n) &= \frac{1}{\log Q} \sum_{d=1}^{n-2} \sum_{j=0}^{\lfloor (n-1)/(d+1)\rfloor} \binom{n-dj}{j+1} (n-j-1)!p^{j+1} \sum_{k\in\mathbb{Z}} \frac{\Gamma(j+1+\chi_k)}{\Gamma(n+1+\chi_k)} \\ &= \frac{1}{\log Q} \sum_{j=1}^{\lfloor (n-1)/2\rfloor} \sum_{d=1}^{\lfloor (n-j-1)/j\rfloor} \binom{n-dj}{j+1} (n-j-1)!p^{j+1} \sum_{k\in\mathbb{Z}} \frac{\Gamma(j+1+\chi_k)}{\Gamma(n+1+\chi_k)} \\ &\quad + \frac{pn}{\log Q} \sum_{k\in\mathbb{Z}} \frac{n!\Gamma(1+\chi_k)}{\Gamma(n+1+\chi_k)} + O(1).\end{aligned}$$

We start with the sum over d: first of all, rewrite the sum as

$$\sum_{d=1}^{\lfloor (n-j-1)/j \rfloor} \binom{n-dj}{j+1} (n-j-1)! = \frac{n!}{(j+1)!} \sum_{d=1}^{\lfloor (n-j-1)/j \rfloor} \prod_{r=0}^{j} \left(1 - \frac{dj}{n-r}\right).$$

We replace the product by a simpler function:

$$\begin{aligned}\prod_{r=0}^{j} \left(1 - \frac{dj}{n-r}\right) &= \left(1 - \frac{dj}{n}\right)^{j+1} \prod_{r=0}^{j} \left(1 - \frac{djr}{(n-r)(n-dj)}\right) \\ &= \left(1 - \frac{dj}{n}\right)^{j+1} \left(1 + O\left(\sum_{r=0}^{j} \frac{djr}{(n-r)(n-dj)}\right)\right) \\ &= \left(1 - \frac{dj}{n}\right)^{j+1} + O\left(\frac{dj^3}{n^2}\left(1 - \frac{dj}{n}\right)^{j}\right).\end{aligned}$$

The estimate is uniform in j. Let us also remark that the O-term is not necessarily smaller than the first term (if j is too large, this is no longer the case), but this is not important for the rest of the argument in view of the exponential term p^{j+1}. Now the Euler-Maclaurin sum formula yields

$$\sum_{d=1}^{\lfloor (n-j-1)/j \rfloor} \left(1 - \frac{dj}{n}\right)^{j+1} = \int_0^{n/j} \left(1 - \frac{jt}{n}\right)^{j+1} \mathrm{d}t + O(1) = \frac{n}{j(j+2)} + O(1)$$

and similarly

$$\begin{aligned}\sum_{d=1}^{\lfloor (n-j-1)/j \rfloor} d\left(1 - \frac{dj}{n}\right)^{j} &= \frac{n}{j} \sum_{d=1}^{\lfloor (n-j-1)/j \rfloor} \left(1 - \frac{dj}{n}\right)^{j} - \left(1 - \frac{dj}{n}\right)^{j+1} \\ &= \frac{n}{j}\left(\frac{n}{j(j+1)(j+2)} + O(1)\right).\end{aligned}$$

Hence we obtain

$$\sum_{d=1}^{\lfloor (n-j-1)/j \rfloor} \binom{n-dj}{j+1} (n-j-1)! = \frac{n!}{(j+1)!} \left(\frac{n}{j(j+2)} + O\left(\frac{j^2}{n}\right) + O(1)\right).$$

The remaining steps (summation over all j, replacing $n!/\Gamma(n+1+\chi_k)$ by $\exp(-\chi_k \log n)$) are analogous to the previous section. We end up with the following theorem:

Theorem 7. *The mean value $m(n)$ of the minimum separation between maxima satisfies*

$$m(n) = n\phi(\log_Q n) + O(1),$$

where the 1*-periodic function* ϕ *is given by the Fourier series*

$$\phi(x) = \frac{p}{\log Q} \sum_{k \in \mathbb{Z}} \left(\Gamma(1 - 2k\pi i/\log Q) + \sum_{j \geq 1} \frac{p^j \Gamma(j + 1 - 2k\pi i/\log Q)}{j(j+1)(j+2)j!} \right) e^{2k\pi i x}.$$

The constant term in the Fourier series is

$$\frac{p}{\log Q} \left(1 + \sum_{j \geq 1} \frac{p^j}{j(j+1)(j+2)} \right) = \frac{7p - 2}{4 \log Q} + \frac{q^2}{2p}.$$

The coefficients of the remaining terms can be written as hypergeometric functions.

The fluctuations arising from $n\phi(\log_Q n)$ and $\phi(\log_Q n)$ are illustrated below for $p = 1/2$.

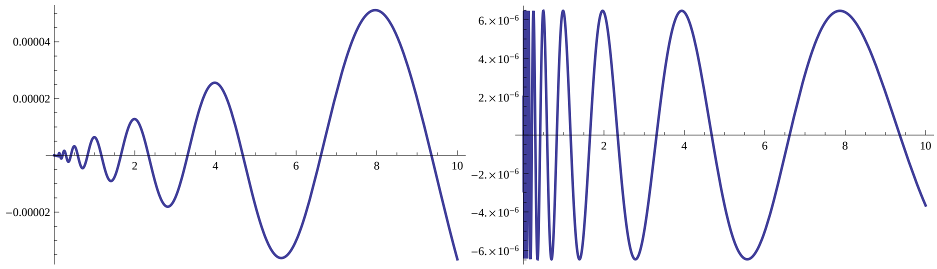

Figure 2. The fluctuations from $n\phi(\log_Q n)$ and $\phi(\log_Q n)$.

Let $p_1(n)$ denote the probability that a sample of length n has exactly one maximum value. It is known that

$$p_1(n) = \frac{p}{\log Q} \sum_{k \in \mathbb{Z}} \Gamma(1 - 2k\pi i/\log Q) e^{2k\pi i x} + O\left(\frac{1}{n}\right),$$

see [**9**]. This means that a large contribution to the mean $m(n)$, of $np_1(n)$, comes from these geometric samples of length n with exactly one maximum value. It is more meaningful to exclude this case and to consider instead the *conditional mean value* of the minimum separation of the maxima, for samples of length n with *at least two maxima*. This is given by

$$m_2(n) := \frac{m(n) - np_1(n)}{1 - p_1(n)}, \tag{3.2}$$

where $1 - p_1(n)$ is the probability that a sample of length n has at least two maxima. Then (3.2), together with Theorem 7 lead to the following result.

Theorem 8. *The average minimum separation between maxima, for samples of* n *geometric variables with at least two maxima satisfies*

$$m_2(n) = n\tilde{\phi}(\log_Q n) + O(1). \tag{3.3}$$

The 1*-periodic function* $\tilde{\phi}$ *is given by*

$$\tilde{\phi}(x) = \frac{\displaystyle\sum_{k\in\mathbb{Z}}\sum_{j\geq 1}\frac{p^j\Gamma(j+1-2k\pi i/\log Q)}{j(j+1)(j+2)j!}e^{2k\pi ix}}{\displaystyle p^{-1}\log Q - \sum_{k\in\mathbb{Z}}\Gamma(1-2k\pi i/\log Q)e^{2k\pi ix}}.$$

If p is not too close to 1, the fluctuations are quite tiny and can essentially be ignored. In particular, in the special case $p = 1/2$, we have from Theorem 8 that $m_2(n) \approx \frac{n}{4}$. By contrast, $m(n) \approx n\Big(\frac{1}{4} + \frac{3}{8\log 2}\Big) \approx 0.791011n$ for $p = 1/2$. However, of this, $\frac{n}{2\log 2} \approx 0.721348n$ is in fact the contribution from samples with only one maximum.

REFERENCES

1. Y. Baryshnikov, B. Eisenberg, G. Stengle: *A necessary and sufficient condition for the existence of the limiting probability of a tie for first place.* Statist. Probab. Lett., **23** (1995), 203–209.
2. J. J. A. M. Brands, F. W. Steudel, R. J. G. Wilms: *On the number of maxima in a discrete sample.* Statist. Probab. Lett., **20** (1994), 209–217.
3. L. Devroye: *A limit theory for random skip lists.* Adv. in Appl. Probab., **2** (1992), 597–609.
4. E3436 Problem. *Amer. Math. Monthly*, **98** (1991), 366.
5. E3436 Solution. *Amer. Math. Monthly*, **101** (1994), 78–80.
6. B. Eisenberg, G. Stengle, G. Strang: *The asymptotic probability of a tie for first place.* Ann. Appl. Probab., **3** (1993), 731–745.
7. P. Flajolet, G. N. Martin: *Probabilistic counting algorithms for data base applications.* J. Comput. System Sci., **31** (1985), 182–209.
8. P. Flajolet, R. Sedgewick: *Mellin transforms and asymptotics: Finite differences and Rice's integrals. Theoret. Comput. Sci.*, **144** (1995), 101–124.
9. P. Kirschenhofer, H. Prodinger: *The number of winners in a discrete geometrically distributed sample.* Ann. Appl. Probab., **6** (1996), 687–694.
10. P. Kirschenhofer, H. Prodinger, W. Szpankowski: *Analysis of a splitting process arising in probabilistic counting and other related algorithms.* Random Structures Algorithms, **9** (1996), 379–401.
11. W. Pugh: *Skip lists: A probabilistic alternative to balanced trees.* ACM Commun., **33** (1990), 668–676.
12. W. Szpankowski, V. Rego: *Yet another application of a binomial recurrence. Order statistics.* Computing, **43** (4) (1990), 401–410.

(Received February 15, 2011)
(Revised July 11, 2011)

The John Knopfmacher Centre
for Applicable Analysis and Number Theory,
School of Mathematics,
University of the Witwatersrand,
Private Bag 3, Wits 2050,
Johannesburg,
South Africa
E-mails: Charlotte.Brennan@wits.ac.za
Arnold.Knopfmacher@wits.ac.za

Department of Mathematics,
University of Haifa,
31905 Haifa,
Israel
E-mail: toufik@math.haifa.ac.il

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