Ascending runs of sequences of geometrically distributed random variables: a probabilistic analysis

Guy Louchard* and Helmut Prodinger†

March 5, 2001

Abstract

Using a Markov chain approach and a polyomino-like description, we study some asymptotic properties of sequences of ascending runs of geometrically distributed random variables. We analyze the limiting trajectories, the number of runs and the run length distribution, the hitting time to a length k run and the maximum run length.

1 Introduction

We consider geometrically distributed random variables (RV), with the distribution GEOM(p), given by $p(i) := pq^{i-1}, q := 1-p$. Assume that we have a sequence $a_1a_2 \dots a_n$, obtained by independent GEOM(p) RVs. We might refer to such a sequence also as a *word of length n*. An ascending run (run for short) is a maximal (contiguous) increasing subsequence. For example 1 1 3 5 2 4 7 1 3 3 8 has five runs, the second being 1 3 5.

The second author [18] presented several enumerative and asympotic results about ascending runs. In this paper, questions about ascending runs are considered from a more probabilistic point of view. In particular we try to continue the approach the first author used in [15]. That accounts to consider the runs either as a stochastic process or as a polyomino. In this way, the derivation of several asymptotic distributions of RVs and processes such as asymptotic Markov chain and limiting trajectories, number of runs and run length distribution, hitting time to a length k run and maximum run length can be achieved. Our results are mainly based on q–series.

A general point of interest is the limit $q \to 1$, as the model turns into the model of random permutations. Consequently, one obtains results that depend only on the order statistics as corollaries. See [7] for a more elaborate description.
The function $\prod_{n\geq 0}(1+pzq^n)$ appears frequently in our analysis. It is known as a q–exponential function (see [9]), and its limit for $q \to 1$ is e^z.

2 Associated Markov Chain and limiting trajectories

In this section, we first describe the runs process as a Markov chain. We then study the run length distribution. We finally analyze the limiting trajectories and the number of runs distribution.

*Université Libre de Bruxelles, Département d'Informatique, CP 212, Boulevard du Triomphe, B-1050 Bruxelles, Belgium, email:louchard@ulb.ac.be

†University of the Witwatersrand, The John Knopfmacher Centre for Applicable Analysis and Number Theory, Mathematics Department, P.O. Wits, 2050 Johannesburg, South Africa, e-mail: helmut@gauss.cam.wits.ac.za

2.1 Stochastic analysis

It is immediately checked that we can see the runs as a Markov chain, the states of which are given by the possible values i of the GEOM(p) RVs starting a run, together with the run length. The following notations will be used in the sequel.

$$\begin{aligned}
L(i) &:= \text{length of an ascending run starting with value } i,\\
P(i) &:= \sum_{j\le i} p(j) = 1-q^j,\\
\varphi(i,l,j) &:= \Pr[L(i)=l, i_{l+1}=j]\\
&= \sum_{i<i_2<\ldots<i_l\ge j} p(i_2)\cdots p(i_l)p(j),\\
\varphi(i,l) &:= \Pr[L(i)=l] = \sum_j \varphi(i,l,j)\\
&= \sum_{i<i_2<\ldots<i_l} p(i_2)\cdots p(i_l)P(i_l); \quad \text{note that}\\
\sum_l \varphi(i,l) &= 1,\\
\Pi[[i,l],[j,\lambda]] &:= \Pr[\text{next run starts at } j,\ \text{has length } \lambda,\\
&\quad\ \text{given that the previous run was starting at } i \text{ and had length } l]\\
&= \varphi(i,l,j)\varphi(j,\lambda)/\varphi(i,l).
\end{aligned}$$

Two other transition matrices are given by

$$\begin{aligned}
\Pi[[i,l],j] &:= \Pr[\text{next run starts at value} j,\ \text{given that the previous run was starting at } i \text{ and had length } l]\\
&= \sum_\lambda \Pi[[i,l],[j,\lambda]]\\
&= \varphi(i,l,j)/\varphi(i,l), \qquad (1)
\end{aligned}$$

hence

$$\begin{aligned}
\Pi[[i,l],[j,\lambda]] &= \Pi[[i,l],j]\varphi(j,\lambda), \qquad (2)\\
\Pi[i,j] &:= \Pr[\text{next run starts at } j,\ \text{given that the previous run was starting at } i]\\
&= \sum_l \varphi(i,j,l). \qquad (3)
\end{aligned}$$

Note that $\varphi(i,l)$ can be computed directly. We will use one of the Euler identities, with the standard notation $(a;x)_k = \prod_{r=1}^k (1-ax^{r-1})$ (see [2]):

$$\prod_{k=0}^\infty (1+q^k z) = \sum_{l=0}^\infty \frac{z^l q^{\binom{l}{2}}}{(q;q)_l}. \qquad (4)$$

This formula is useful for the enumeration of ascending runs because of the following

$$\Pr[X_1 < X_2 < \ldots < X_n] = [z^n]\prod_{k=0}^\infty (1+pq^k z) = \frac{p^n q^{\binom{n}{2}}}{(q;q)_n}.$$

Therefore

$$\sum_{i<i_2<\ldots<i_l} pq^{i_2-1}\ldots pq^{i_l-1} = \left(\frac{p}{q}\right)^{l-1} \sum_{i<i_2<\ldots<i_l} q^{i_2+\ldots+i_l}$$

$$
\begin{aligned}
&= \left(\frac{p}{q}\right)^{l-1} q^{(i+1)(l-1)} \sum_{0\le i_2<\ldots<i_l} q^{i_2+\ldots+i_l} \\
&= p^{l-1}q^{i(l-1)}[t^{l-1}]\prod_{n\ge 0}(1+tq^n) \\
&= p^{l-1}q^{i(l-1)}\frac{q^{\binom{l-1}{2}}}{(q;q)_{l-1}}
\end{aligned}
$$

and consequently

$$\varphi(i,l) = p^{l-1}q^{i(l-1)}\frac{q^{\binom{l-1}{2}}}{(q;q)_{l-1}} - p^l q^{il}\frac{q^{\binom{l}{2}}}{(q;q)_l}. \tag{5}$$

Seeing that answer, one can even think of a combinatorial reasoning for this formula. Indeed, the first term is the probability for an increasing sequence of length $l-1$, subject to all the elements being larger than i, and the second term is the probability for an increasing sequence of length l, under the same restriction.

The stationary measure of (2) is given by $\pi(i,l)$, which satisfies

$$\sum_{i,l}\pi(i,l)\Pi[[i,l],[j,\lambda]] = \pi(j,\lambda),$$

or, by (2),

$$\sum_{i,l}\pi(i,l)\Pi[[i,l],j]\varphi(j,\lambda) = \pi(j,\lambda).$$

Set

$$\psi(j) := \sum_{i,l}\pi(i,l)\Pi[[i,l],j] \tag{6}$$

and

$$\pi(i) := \sum_{l}\pi(i,l),$$

hence

$$\sum_i \pi(i) = 1. \tag{7}$$

On one side, we have

$$\pi(j,\lambda) = \psi(j)\varphi(j,\lambda),$$

hence

$$\psi(j) \equiv \pi(j)$$

and

$$\pi(j,\lambda) = \pi(j)\varphi(j,\lambda).$$

On the other side, (6) leads to

$$
\begin{aligned}
\pi(j) &= \sum_{i,l}\pi(i)\varphi(i,l)\Pi[[i,l],j] \\
&= \sum_i \pi(i)\sum_l \varphi(i,j,l) \\
&= \sum_i \pi(i)\Pi[i,j] \quad \text{by (3).}
\end{aligned}
$$

So, with (7), $\pi(i)$ is the stationary measure of $\Pi(i,j)$, which is, of course, probabilistically obvious.

As in [15], we need to compute the following averages and second moments

$$\begin{aligned}\overline{L}(i) &= \sum_l \varphi(i,l)l,\\ s^2(i) &= \sum_l \varphi(i,l)l^2,\\ \overline{L} &= \sum_i \pi(i)\overline{L}(i).\end{aligned}$$

The quantity $\overline{L}$ should evaluate to $1+q$, according to [18], where the average length of the mth run was considered, with the limit $1+q$ for $m \to \infty$. Also, we must compute

$$s^2 = \sum_i \pi(i)s^2(i),$$

which is the stationary second moment of the run length, and

$$\begin{aligned}S &:= \sum_{k,i,l,j,\lambda} \pi(i)\varphi(i,l)(l-\overline{L})\Pi^k[[i,l],[j,\lambda]](\lambda-\overline{L})\\ &= \sum_{k,i,l} \pi(i)\varphi(i,l)(l-\overline{L}) \sum_{w_1,w_2,\ldots,w_{k-1}} \Pi[[i,l],w_1]\Pi[w_1,w_2]\cdots\Pi[w_{k-2},w_{k-1}]\sum_j \Pi[w_{k-1},j](\overline{L}(j)-\overline{L}),\end{aligned}$$

which could lead to the asymptotic variance of the total length $X(m)$ (number of GEOM(p) RVs) of m runs. However, we found that approach computationally difficult.

2.2 Some generating functions

In this subsection, we consider first some asymptotics for probability transition matrices, stationary distribution and run lengths. Then we analyze the asymptotics of m successive runs.

2.2.1 One ascending run

We build a corresponding polyomino as follows: we choose a starting integer i, and construct an ascending sequence of l integers, $i = i_1 < \ldots < i_l$, followed by a last integer j with $j \le i_l$. We mark each integer by z; the total size of the polyomino $\xi := \sum_{k=1}^{l} i_k + j$ is marked by x, the last increasing integer i_l is marked by η, and j is marked by θ.

The trivariate Generating Function (GF) $\Gamma(\theta,x,z|i)$ that comprises all such polyominos, labelled as described, can be computed as follows: set

$$\begin{aligned}g_1 &:= zx^i\eta^i, \text{ and, for } k \ge 1,\\ g_{k+1}(\eta,\theta,x,z) &= \sum_{\nu=1}^{\infty}[\eta^\nu]g_k(\eta,0,x,z)z\left[\sum_{r=\nu+1}^{\infty}(x\eta)^r + \sum_{j=1}^{\nu}(x\theta)^j\right]\\ &= g_k(1,0,x,z)xz\frac{\theta}{1-x\theta} + \sum_{\nu=1}^{\infty}[\eta^\nu]g_k(\eta,0,x,z)xz\left[(x\eta)^\nu\frac{\eta}{1-x\eta} - (x\theta)^\nu\frac{\theta}{1-x\theta}\right];\end{aligned}$$

this sequence describes the effect of "adding a new slice." The GF g_k corresponds to a polyomino of k integers (each integer marked by z), with total size marked by x, where the last integer is marked either by η, if it is still increasing, or by θ, if it is smaller than or equal to the previous integer. Hence

$$\begin{aligned}\Psi(\eta,\theta,x,z|i) &:= \sum_{k=1}^{\infty} g_k(\eta,\theta,x,z)\\ &= g_1 + \Psi(1,0,x,z|i)\frac{xz\theta}{1-x\theta} + \Psi(x\eta,0,x,z|i)\frac{xz\eta}{1-x\eta} - \Psi(x\theta,0,x,z|i)\frac{xz\theta}{1-x\theta}.\end{aligned}$$

This is a functional equation that we are going to solve by iteration. For that we set

$$A_2(\eta,x,z|i) := \sum_{k=0}^{\infty} \frac{x^{\binom{k+1}{2}} z^{k+1}\eta^k x^{(k+1)i}\eta^i}{(x\eta;x)_k},$$

and derive, by iteration of the above functional equation,

$$\begin{aligned}\Psi(\eta,\theta,x,z|i) &:= \Psi(1,0,x,z|i)\frac{xz\theta}{1-x\theta} + A_2(\eta,x,z|i) - \Psi(x\theta,0,x,z|i)\frac{xz\theta}{1-x\theta} \\ &= \Psi(1,0,x,z|i)\frac{xz\theta}{1-x\theta} + A_2(\eta,x,z|i) - A_2(x\theta,x,z|i)\frac{xz\theta}{1-x\theta}.\end{aligned}$$

Setting $\eta = 1$ and $\theta = 0$ this leads to

$$\Psi(1,0,x,z|i) = A_2(1,x,z|i)$$

so that we find eventually

$$\Psi(\eta,\theta,x,z|i) = [A_2(1,x,z|i) - A_2(x\theta,x,z|i)]\frac{xz\theta}{1-x\theta} + A_2(\eta,x,z|i).$$

Finally,

$$\begin{aligned}\Gamma(\theta,x,z|i) &= \Psi(0,\theta,x,z|i) \\ &= [A_2(1,x,z|i) - A_2(x\theta,x,z|i)]\frac{xz\theta}{1-x\theta}.\end{aligned}$$

The generating function Γ "contains" the quantities of interest, as described in the next proposition:

Proposition 2.1

$$\begin{aligned}\varphi(i,l,j) &= [z^{l+1}\theta^j]\Gamma(\theta,q,z|i)(p/q)^l/q^i, \\ \varphi(i,l) &= [z^{l+1}]\Gamma(1,q,z|i)(p/q)^l/q^i \\ &= q^{\binom{l-1}{2}}p^{l-1}q^{(l-1)i}\left[\frac{1}{(q;q)_{l-1}} - \frac{pq^{l-1+i}}{(q;q)_l}\right], \\ \Pi[i,j] &= [\theta^j]\Gamma(\theta,q,p/q|i)/(pq^{i-1}).\end{aligned}$$

□

Of course, we recover (5). It is an easy check to derive that $\sum_j \Pi[i,j] = 1$, which follows from $\sum_j \Pi[i,j] = \Gamma(1,q,p/q|i)/(pq^{i-1})$ and

$$A_2(x,x,z|i) = \frac{1-x}{zx}\Big[A_2(1,x,z|i) - zx^i\Big].$$

To obtain $\pi(i)$, we start from the stationary equation:

$$\pi(j) = \sum_{k=0}^{\infty} q^{\binom{k}{2}}p^k \sum_i \pi(i)q^{ki}[\theta^j]\frac{p\theta}{1-q\theta}\left[\frac{1}{(q;q)_k} - \frac{(q\theta)^{k+i}}{(q^2\theta;q)_k}\right].$$

We compute the GF of $\pi(i)$, $\Theta(\xi) := \sum_{j=1}^{\infty}\pi(j)\xi^j$:

$$\Theta(\xi) = \sum_{k=0}^{\infty} q^{\binom{k}{2}}p^k \left\{\Theta(q^k)\frac{1}{(q;q)_k} - \Theta(\xi q^{k+1})\frac{(q\xi)^k}{(q^2\xi;q)_k}\right\}\frac{p\xi}{1-q\xi}. \tag{8}$$

Since we find this relation only manageable for ξ being a power of q, we define $G(j) := \Theta(q^j)$ and rewrite the functional equation in terms of $G(j)$:

$$
\begin{aligned}
G(j) &= \sum_{k=0}^{\infty} q^{\binom{k}{2}} p^k \left\{ G(k)\frac{1}{(q;q)_k} - G(k+1+j)\frac{q^{(j+1)k}}{(q^{j+2};q)_k} \right\} \frac{pq^j}{1-q^{j+1}} \qquad (9)\\
&= \frac{pq^j}{1-q^{j+1}} \sum_{k=0}^{\infty} q^{\binom{k}{2}} p^k G(k)\frac{1}{(q;q)_k} - \sum_{k=0}^{\infty} q^{\binom{k}{2}} p^k G(k+1+j)\frac{pq^{(j+1)k+j}(q;q)_j}{(q;q)_{k+j+1}}\\
&= \frac{pq^j}{1-q^{j+1}} \sum_{k=0}^{\infty} q^{\binom{k}{2}} p^k G(k)\frac{1}{(q;q)_k} - \frac{(q;q)_j}{p^j q^{\binom{j}{2}}} \sum_{k=j+1}^{\infty} q^{\binom{k}{2}} p^k G(k)\frac{1}{(q;q)_k}.
\end{aligned}
$$

Denote $C := \sum_{k=0}^{\infty} q^{\binom{k}{2}} p^k G(k)/(q;q)_k$, then we get from that the recurrence equation:

$$
G(l) = \frac{pq^l}{1-q^{l+1}}C - \frac{(q;q)_l}{q^{\binom{l}{2}}p^l}\left[C - \sum_{k=0}^{l} G(k)\frac{q^{\binom{k}{2}}p^k}{(q;q)_k}\right]. \qquad (10)
$$

Note that $G(0) = \Theta(1) = \sum_j \pi(j) = 1$, so that $l = 1$ leads to

$$
G(1) = \frac{q}{1+q}C - C + 1 + G(1),
$$

and therefore $C = 1 + q$. An equivalent formula to the recursion (10) is

$$
\frac{q^{\binom{l}{2}}p^l G(l)}{(q;q)_l} = (1+q)\frac{p^{l+1}q^{\binom{l+1}{2}}}{(q;q)_{l+1}} - (1+q) + \sum_{0\le j\le l} G(j)\frac{q^{\binom{j}{2}}p^j}{(q;q)_j}. \qquad (11)
$$

Now define the generating function

$$
\psi(z) := \sum_{l\ge 0} \frac{G(l)q^{\binom{l}{2}}p^l z^l}{(q;q)_l}.
$$

Multiply (11) by z^l and sum:

$$
\psi(z) = \frac{1+q}{z}\sum_{l\ge 1}\frac{p^l q^{\binom{l}{2}}}{(q;q)_l}z^l - \frac{1+q}{1-z} + \frac{1}{1-z}\psi(z).
$$

From this and Euler's partition identity we get

$$
\psi(z) = -\frac{1-z}{z^2}(1+q)\prod_{n\ge 0}(1+pzq^n) + \frac{1+q}{z^2}.
$$

Therefore ($m \ge 0$)

$$
\begin{aligned}
[z^m]\psi(z) &= -(1+q)[z^{m+2}]\prod_{n\ge 0}(1+pzq^n) + (1+q)[z^{m+1}]\prod_{n\ge 0}(1+pzq^n)\\
&= -(1+q)\frac{p^{m+2}q^{\binom{m+2}{2}}}{(q;q)_{m+2}} + (1+q)\frac{p^{m+1}q^{\binom{m+1}{2}}}{(q;q)_{m+1}}\\
&= (1+q)\frac{p^{m+1}q^{\binom{m+1}{2}}}{(q;q)_{m+2}}\left[-pq^{m+1} + 1 - q^{m+2}\right]\\
&= (1+q)\frac{p^{m+1}q^{\binom{m+1}{2}}}{(q;q)_{m+2}}\left[1 - q^{m+1}\right].
\end{aligned}
$$

From this we can compute $G(l)$:

$$
\begin{aligned}
G(l) &= (1+q)\frac{p^{l+1}q^{\binom{l+1}{2}}}{(q;q)_{l+2}}\left[1-q^{l+1}\right]\frac{(q;q)_l}{q^{\binom{l}{2}}p^l} \\
&= (1+q)\frac{pq^l}{(q;q)_{l+2}}(q;q)_{l+1} \\
&= (1+q)\frac{pq^l}{1-q^{l+2}}. \qquad (12)
\end{aligned}
$$

Also we obtain the identity

$$
C = 1+q = \sum_{k=0}^{\infty} q^{\binom{k+1}{2}}p^k(1-q^2)\frac{1}{(1-q^{k+2})(q;q)_k}.
$$

which can be directly proved as follows: It is equivalent to

$$
\sum_{k\geq 0} q^{\binom{k+1}{2}}p^k\frac{1}{(1-q^{k+2})(q;q)_k} = \frac{1}{1-q}.
$$

We compute:

$$
\begin{aligned}
& \sum_{k\geq 0} q^{\binom{k+1}{2}}p^k\frac{1}{(1-q^{k+2})(q;q)_k} \\
=\ & \sum_{k\geq 0} q^{\binom{k+1}{2}}p^k\frac{1-q^{k+1}}{(q;q)_{k+2}} \\
=\ & \sum_{k\geq 0} q^{\binom{k+1}{2}}p^k\frac{1}{(q;q)_{k+2}} - \sum_{k\geq 0} q^{\binom{k+1}{2}}p^k\frac{q^{k+1}}{(q;q)_{k+2}} \\
=\ & \sum_{k\geq 2} q^{\binom{k-1}{2}}p^{k-2}\frac{1}{(q;q)_k} - \sum_{k\geq 2} q^{\binom{k-1}{2}}p^{k-2}\frac{q^{k-1}}{(q;q)_k} \\
=\ & \frac{q}{p^2}\sum_{k\geq 2} q^{\binom{k}{2}}\frac{(p/q)^k}{(q;q)_k} - \frac{1}{p^2}\sum_{k\geq 2} q^{\binom{k}{2}}\frac{p^k}{(q;q)_k} \\
=\ & \frac{q}{p^2}\left[\prod_{n\geq 0}(1+\frac{p}{q}\,q^n) - 1 - \frac{p/q}{1-q}\right] - \frac{1}{p^2}\left[\prod_{n\geq 0}(1+p\,q^n) - 1 - \frac{p}{1-q}\right] \\
=\ & \frac{q}{p^2}\left[(1+\frac{p}{q})\prod_{n\geq 0}(1+p\,q^n) - 1 - \frac{1}{q}\right] - \frac{1}{p^2}\left[\prod_{n\geq 0}(1+p\,q^n) - 2\right] \\
=\ & \frac{q}{p^2}\left[\frac{1}{q}\prod_{n\geq 0}(1+p\,q^n) - 1 - \frac{1}{q}\right] - \frac{1}{p^2}\left[\prod_{n\geq 0}(1+p\,q^n) - 2\right] \\
=\ & \frac{q}{p^2}\left(-1-\frac{1}{q}\right) + \frac{2}{p^2} = \frac{1}{p^2}(-q-1+2) = \frac{1}{p}.
\end{aligned}
$$

Now we can compute the probabilities $\pi(j)$:

Proposition 2.2

$$
\pi(j) = (1-q^2)q^{2(j-1)}.
$$

Proof We have just proved that for arbitrary k,

$$G(k) = \Theta(q^k) = \sum_{j\geq 1} \pi(j) q^{kj} = \frac{(1-q^2)q^k}{1-q^{k+2}}.$$

Write simply z instead of q^k, then

$$\Theta(z) = \sum_{j\geq 1} \pi(j) z^j = \frac{(1-q^2)z}{1-q^2 z}.$$

Indeed we know (see for instance [19, p.210]), that if f and g are holomorphic in a region Ω, and if $f(z) = g(z)$ for all z in some set which has a limit point in Ω, then $f(z) = g(z)$ for all $z \in \Omega$. Therefore

$$\begin{aligned}
\pi(j) &= [z^j]\frac{(1-q^2)z}{1-q^2 z} \\
&= (1-q^2)[z^{j-1}]\frac{1}{1-q^2 z} \\
&= (1-q^2)q^{2(j-1)}.
\end{aligned}$$

■

When $q \to 1$, we obtain

$$\pi(j) \sim 2pe^{-2pj},$$

which shows the asymptotic exponential property of the stationary distribution.

We also derive the asymptotic distribution $\overline{F}(l) := \Pr[L = l]$ of the run length L:

$$\begin{aligned}
\overline{F}(l) := \sum_i \pi(i)\varphi(i,l) &= q^{(l-1)(l-2)/2} p^{l-1} \left[\frac{\Theta(q^{l-1})}{(q;q)_{l-1}} - \frac{pq^{l-1}\Theta(q^l)}{(q;q)_l} \right] \\
&= \frac{q^{l(l-1)/2} p^{l-1}(1-q^2)}{(q;q)_l} \left(\frac{1-q^l}{1-q^{l+1}} - \frac{pq^l}{1-q^{l+2}} \right),
\end{aligned} \tag{13}$$

and, as $\Theta(1) = 1$, $\sum_l \overline{F}(l) = 1$ by (8).
When $q \to 1$, we obtain

$$\frac{p^k}{(q;q)_k} \sim \frac{1}{k!}$$

and

$$\overline{F}(l) \sim \frac{2(l^2+l-1)}{(l+2)!}, \tag{14}$$

which sums to 1.

By partial summation and (5), we obtain

$$\overline{L}(i) = \sum_l \varphi(i,l) l = \sum_{l=0}^{\infty} \frac{(pq^i)^l q^{\binom{l}{2}}}{(q;q)_l},$$

and, by (4), we derive

$$\overline{L}(i) = \prod_{k=0}^{\infty} (1 + q^k pq^i).$$

Finally

$$\overline{L} = \sum_i \pi(i)\overline{L}(i) = \sum_{k=0}^{\infty} q^{\binom{k}{2}} p^k \Theta(q^k) \frac{1}{(q;q)_k} \equiv C = 1 + q,$$

which conforms to [18]. Similarly,

$$s^2(i) = \sum_l \varphi(i,l) l^2 = \sum_{l=0}^{\infty} \frac{(2l+1)(pq^i)^l q^{\binom{l}{2}}}{(q;q)_l},$$

but, multiplying (4) by z and differentiating w.r.t. z, we get

$$\prod_{k=0}^{\infty}(1+q^k z)\left[1 + z\sum_{l=0}^{\infty}\frac{q^l}{1+q^l z}\right] = \sum_{l=0}^{\infty}\frac{(l+1)z^l q^{\binom{l}{2}}}{(q;q)_l}.$$

This gives another expression:

$$s^2(i) = \prod_{k=0}^{\infty}(1+q^k pq^i)\left[1 + 2pq^i\sum_{l=0}^{\infty}\frac{q^l}{1+q^l pq^i}\right].$$

We can also compute

$$\begin{aligned} s^2 &= \sum_i \pi(i)s^2(i) = \sum_{k=0}^{\infty}(2k+1)q^{\binom{k}{2}}p^k\Theta(q^k)\frac{1}{(q;q)_k} \\ &= \sum_{k=0}^{\infty}(2k+1)q^{\binom{k+1}{2}}p^k(1-q^2)\frac{1}{(1-q^{k+2})(q;q)_k}. \end{aligned}$$

When $q \to 1$, we obtain

$$s^2 \sim \sum_{k=0}^{\infty}(2k+1)\frac{2}{(k+2)k!} = 4e - 6 = 4.873127314\ldots.$$

A more explicit expression for s^2 is computed as follows:

$$s^2 = (1+q)\sum_{k\ge 0}(2k+1)q^{\binom{k+1}{2}}p^{k+1}\frac{1-q^{k+1}}{(q;q)_{k+2}}.$$

Define

$$F(z) := (1+q)\sum_{k\ge 0}q^{\binom{k+1}{2}}p^{k+1}\frac{1-q^{k+1}}{(q;q)_{k+2}}z^{2k+1},$$

then $s^2 = F'(1)$. Now

$$\begin{aligned} F(z) &:= (1+q)\sum_{k\ge 0}q^{\binom{k+1}{2}}p^{k+1}\frac{1-q^{k+1}}{(q;q)_{k+2}}z^{2k+1} \\ &= (1+q)\sum_{k\ge 0}q^{\binom{k+1}{2}}p^{k+1}\frac{1}{(q;q)_{k+2}}z^{2k+1} - (1+q)\sum_{k\ge 0}q^{\binom{k+2}{2}}p^{k+1}\frac{1}{(q;q)_{k+2}}z^{2k+1} \\ &= (1+q)\sum_{k\ge 2}q^{\binom{k-1}{2}}p^{k-1}\frac{1}{(q;q)_k}z^{2k-3} - (1+q)\sum_{k\ge 2}q^{\binom{k}{2}}p^{k-1}\frac{1}{(q;q)_k}z^{2k-3} \\ &= \frac{q(1+q)}{pz^3}\sum_{k\ge 2}q^{\binom{k}{2}}(p/q)^k\frac{1}{(q;q)_k}z^{2k} - \frac{(1+q)}{pz^3}\sum_{k\ge 2}q^{\binom{k}{2}}p^k\frac{1}{(q;q)_k}z^{2k} \\ &= \frac{q(1+q)}{pz^3}\left[\prod_{n\ge 0}(1+\frac{p}{q}z^2q^n) - 1 - \frac{1}{q}z^2\right] - \frac{(1+q)}{pz^3}\left[\prod_{n\ge 0}(1+pz^2q^n) - 1 - z^2\right] \\ &= \frac{q(1+q)}{pz^3}\left[\frac{q+pz^2}{q}\prod_{n\ge 0}(1+pz^2q^n) - 1 - \frac{1}{q}z^2\right] - \frac{(1+q)}{pz^3}\left[\prod_{n\ge 0}(1+pz^2q^n) - 1 - z^2\right] \\ &= \frac{1+q}{z^3} - \frac{1-z^2}{z^3}(1+q)\prod_{n\ge 0}(1+pz^2q^n). \end{aligned}$$

Hence

$$s^2 = -3(1+q) + 2(1+q)\prod_{n\geq 0}(1+pq^n).$$

The variance of the stationary run length is given by

$$\mathrm{VAR}(L) = s^2 - \overline{L}^2.$$

and, when $q \to 1$, we obtain $4e - 10 = 0.873127314\ldots$.

For further use, let us define the run length distribution function and the truncated mean run length, starting from i:

$$\Lambda(i,l) := \sum_{w\leq l}\varphi(i,w)$$

and

$$\overline{L}(i,l) := \sum_{w<l} w\varphi(i,w).$$

Again by partial summation, we obtain

$$\begin{aligned}
\Lambda(i,l) &= 1 - C(i,l), \quad \text{with} && (15)\\
C(i,l) &= \frac{q^{\binom{l}{2}}p^l q^{li}}{(q;q)_l}, && (16)\\
\overline{L}(i,l) &= \overline{L}(i) - \left[(l-1)C(i,l-1) + \sum_{k>l-2}\frac{q^{\binom{k}{2}}p^k q^{ki}}{(q;q)_k}\right]. && (17)
\end{aligned}$$

2.2.2 m ascending runs

Let the total length $X(m)$ (number of GEOM(p) RVs) of m runs be given by

$$X(m) = \sum_{k=1}^{m} L(i_k).$$

This corresponds to the length of a word where each letter is a GEOM(p) RV.

By the ergodic theorem, we know that

$$\begin{aligned}
\mathbb{E}(X(m)) &\sim m\overline{L},\\
\mathrm{VAR}(X(m)) &\sim m[s^2 - \overline{L}^2] + 2mS = m\sigma^2, \text{ say.}
\end{aligned}$$

(see for instance Billingsley [4, p. 172].

If we could compute S explicitly, this would give σ^2. We have been unable to do this, so we turn to another method related to singularity analysis as used in [14, 16]. Note that, here, we use Probability GFs instead of ordinary GFs related to polyominos. Let

$$\Lambda_m(z) := \sum_{n\geq 0}[\text{Probability that a word of length } n \text{ has (exactly) } m \text{ ascending runs}]z^n.$$

It appears already in [18] (and is easy to see) that

$$F_m(z) := \sum_{n=0}^{\infty}\Pr[X(m)=n]z^n = \frac{z-1}{z}[\Lambda_0(z) + \cdots + \Lambda_m(z)] + \frac{1}{z},$$

and so

$$\mathrm{VAR}[X(m)] = F_m''(1) + F_m'(1) - (F_m'(1))^2.$$

Set

$$R(w,z) = \sum_{m\ge 0} \Lambda_m(z) w^m.$$

In [18], the double GF was not correctly computed (which had, fortunately, no consequences). However, the correct formula is

$$R(w,z) = \frac{1-w}{1-w\prod_{n\ge 0}(1+pz(1-w)q^n)}$$

and we will briefly present a derivation of it. We start from the correct formula

$$[z^n]\Lambda_m(z) = [z^n]\sum_{k=0}^{m}(-1)^k\binom{n+1}{k}P^{m-k},$$

where $P = P(z)$ is a shorthand notation for $\prod_{k\ge 0}(1+pzq^k)$. Multiplying it by w^m, and summing over $m \ge 0$ yields:

$$\begin{aligned}
&[z^n]\sum_{m\ge 0} w^m \Lambda_m(z)\\
=\ &[z^n]\sum_{0\le k\le m} w^m(-1)^k\binom{n+1}{k}P^{m-k}\\
=\ &[z^n]\frac{1}{1-wP(z)}\sum_{0\le k} w^k(-1)^k\binom{n+1}{k}\\
=\ &[z^n]\frac{1}{1-wP(z)}(1-w)^{n+1}\\
=\ &[z^n]\frac{1-w}{1-wP(z)}(1-w)^{n}\\
=\ &[z^n]\frac{1-w}{1-wP(z(1-w))}.
\end{aligned}$$

Hence

$$\sum_{m,n} w^m z^n \Lambda_m(z) = \frac{1-w}{1-wP(z(1-w))}.$$

The forthcoming book [12] has generating functions of a similar flavour, credited to Désarmenien. In the limit $q \to 1$ we get

$$R(w,z) = \frac{1-w}{1-we^{z(1-w)}},$$

which is the generating function for Eulerian numbers $A(n,k)$. Note the Knuth's version [11] of the Eulerian numbers differs slightly. Hence the double GF $R(w,z)$ provides q–Eulerian numbers which differ from those investigated by Carlitz [5].

Now we can continue the computation which will lead us to the variance. Note that

$$\begin{aligned}
\left(\frac{z-1}{z}\Lambda_m(z)\right)'\bigg|_{z=1} &= \Lambda_m(1),\\
\left(\frac{z-1}{z}\Lambda_m(z)\right)''\bigg|_{z=1} &= -2\Lambda_m(1) + 2\Lambda'_m(1).
\end{aligned}$$

Now

$$F_m(z) = \frac{z-1}{z}[w^m]\frac{1}{1-w}R(w,z) + \frac{1}{z}.$$

(We should also recall the already proved formula $F_m(1) \sim 1+q$, which we rederived here by different means.) Differentiating w.r.t. z, and plugging in $z=1$, we get by singularity analysis[1]

$$F'_m(1) \sim [w^m]\bigg(\frac{1+q}{(1-w)^2} - \frac{q(1+q)^2}{1+q+q^2}\frac{1}{1-w} + \dots\bigg) - 1,$$

and so

$$F'_m(1) \sim (1+q)(m+1) - \frac{q(1+q)^2}{1+q+q^2} - 1 + \dots,$$

or

$$F'_m(1) \sim (1+q)m - \frac{q^2}{1+q+q^2}.$$

Also

$$R'(w,1) = \frac{1-w}{\Big(1-w\prod_{n\ge0}(1+p(1-w)q^n)\Big)^2}\cdot w\prod_{n\ge0}(1+p(1-w)q^n)\cdot\sum_{n\ge0}\frac{p(1-w)q^n}{1+p(1-w)q^n},$$

and

$$\begin{aligned}
\frac{1}{1-w}R'(w,1) &\sim (q+1)^2(1-w)^{-3} - \frac{(q+1)\left(q^3+4q^2+2q+1\right)}{q^2+q+1}(1-w)^{-2}\\
&+ \frac{\left(2q^3+3q+1\right)q^2(q+1)^2}{(q^2+1)(q^2+q+1)^2}(1-w)^{-1},\\
[w^m]\frac{1}{1-w}R'(w,1) &\sim (q+1)^2\frac{(m+2)(m+1)}{2} - \frac{(q+1)\left(q^3+4q^2+2q+1\right)}{q^2+q+1}(m+1)\\
&+ \frac{\left(2q^3+3q+1\right)q^2(q+1)^2}{(q^2+1)(q^2+q+1)^2}\\
&= \frac12(q+1)^2m^2 + \frac{(q+1)\left(q^3-2q^2+2q+1\right)}{2(q^2+q+1)}m - \frac{(q+1)q^2(-1+q)^2}{(q^2+1)(q^2+q+1)^2}.
\end{aligned}$$

Hence

$$\begin{aligned}
F''_m(1) &\sim 2\bigg(\frac12(q+1)^2m^2 + \frac{(q+1)\left(q^3-2q^2+2q+1\right)}{2(q^2+q+1)}m - \frac{(q+1)q^2(-1+q)^2}{(q^2+1)(q^2+q+1)^2}\bigg)\\
&- 2\bigg((1+q)(m+1) - \frac{q(1+q)^2}{1+q+q^2}\bigg)\\
&+ 2,
\end{aligned}$$

or

$$F''_m(1) \sim (1+q)^2m^2 + \frac{(1+q)(q^3-4q^2-1)}{1+q+q^2}m + ts$$

Finally,

$$\mathrm{VAR}(X(m)) \sim \frac{q(q^3+1)}{q^2+q+1}m - \frac{(q^2-4q+1)(q+1)q^2}{(q^2+1)(q^2+q+1)^2}.$$

So

$$\sigma^2 = \frac{q(q^3+1)}{q^2+q+1}.$$

As already mentioned, the limit $q\to1$ leads us to random permutations, and

[1]The dominant singularity here is $w=1$, and estimates for the modulus of the next (larger) one would give as effective error bounds, of exponential decay.

$$\mathrm{VAR}(X(m)) \sim \frac{2}{3}m + \frac{2}{9}.$$

That formula was already proved by Knuth, see [11, Ex.5.1.3.10].

Although it is not needed in the sequel, we provide an alternative formula for $R(w,z)$:

$$\frac{1}{w} - \frac{1-w}{wR(w,z)} = \prod_{n\geq 0}(1+pz(1-w)q^n) = (1+pz(1-w))\prod_{n\geq 0}(1+p(zq)(1-w)q^n)$$

and so

$$\frac{1}{w} - \frac{1-w}{wR(w,z)} = (1+pz(1-w))\left[\frac{1}{w} - \frac{1-w}{wR(w,zq)}\right]$$

or

$$-\frac{1-w}{wR(w,z)} = \frac{pz(1-w)}{w} - \frac{(1-w)(1+pz(1-w))}{wR(w,zq)}$$

or

$$\frac{1}{R(w,z)} = \frac{1+pz(1-w)}{R(w,zq)} - pz.$$

We cannot iterate that since it starts with 1, however if we set $S(z) = \frac{1}{R(w,z)} - 1$, then

$$S(z) = (1+pz(1-w))S(zq) - pzw,$$

and

$$S(z) = -pzw\sum_{k\geq 0}(pz(w-1);q)_k q^k,$$

and finally

$$R(w,z) = \frac{1}{1 - pzw\sum_{k\geq 0}(pz(w-1);q)_k q^k}.$$

2.3 Limiting trajectories

Let us denote by M the number of runs. Let us first fix M to m and let us consider the total length $X(m)$ (the number of GEOM(p) RVs, i.e. the length of the word) of m runs:

$$X(m) = \sum_{k=1}^{m} L(i_k)$$

where $i_1, \dots, i_m$ are the successive starting values of the runs. We can check that our Markov chain is φ-mixing (see for instance [4, p. 168]) and we apply the Functional Central Limit Theorem ([4, p. 174, Thm. 20.1]). We obtain the following result, where $B(t)$ is the standard Brownian Motion (BM).

Theorem 2.3

$$\frac{X(\lfloor Nt \rfloor) - \overline{L}Nt}{\sigma\sqrt{N}} \Rightarrow B(t), \quad N \to \infty,\ t \in [0,1], \tag{18}$$

Let us now condition on $X(m) = n$. M becomes a RV. A realization of X for fixed n is given by Theorem 2.3, where we stop at a random time m such that $X(m) = n$. Proceeding as in [13] it is easy to check that this amounts to fix $N = n/\overline{L}$ in the denominator of (18), and we obtain the following result.

Theorem 2.4 *Conditioned on $X(m) = n$,*

$$\frac{X(\lfloor mt \rfloor) - \overline{L}mt}{\sigma\sqrt{n/\overline{L}}} \Rightarrow B(t).$$

Moreover, from [17], the RV M:= number of runs is such that

$$M \sim \mathcal{N}(n\mu_1, n\sigma_1^2), \tag{19}$$

where

$$\begin{aligned} \mathcal{N} &:= \text{a Gaussian (normal) RV}, \\ \mu_1 &= 1/\overline{L} = \frac{1}{1+q}, \\ \sigma_1^2 &= \sigma^2/\overline{L}^3. \end{aligned}$$

This is actually an application of the generalized Renewal Theorem: the hitting time of the BM $BM(t)$ of (18) to the barrier $n(1-t)/(\sigma\sqrt{n/\overline{L}})$ is easily seen to lead asymptotically to a Gaussian RV.

When $q \to 1$, we obtain $\mu_1 \sim 1/2$, $\sigma_1^2 \sim 1/12$.

Let us remark that we can also use Bender's techniques, as given by Theorems 1, 3 and (3.2) in [3]: We rediscover the asymptotic Gaussian distribution of M. Indeed, let $h(w,z) := \frac{1}{R(w,z)}$. Set

$$\begin{aligned} r_1 &:= -h_w/h_z, \\ r_2 &:= -(r_1^2 h_{zz} + 2r_1 h_{zw} + h_w + h_{ww})/h_z. \end{aligned}$$

For $w = 1$, we see that the dominant singularity of $R(1,z)$ is given by $z^* = 1$. Set $w = 1$, $z = 1$. After some algebra, this gives

$$h_w^* = -1/(1+q),\ h_z^* = -1,\ h_{ww}^* = 2q/(1+q+q^2),\ h_{zz}^* = 0,\ h_{zw} = -p/(1+q).$$

Bender's results lead to

$$\begin{aligned} \mu_1 &:= -r_1/z^* = \frac{1}{1+q} \\ \sigma_1^2 &:= \mu_1^2 - r_2/z^* = \frac{q(q^3+1)}{(q^2+q+1)(1+q)^3}. \end{aligned}$$

which conforms to (19).

We have done a simulation of $U = 2000$ sequences of $n = 10000$ GEOM(0.25) RVs (this simulation will be extensively used in Section 3), leading to 11429877 runs. The observed and limiting run length distributions are given in Figure 1. (The limiting distribution is given by (13)). The fit is quite good. A typical trajectory for $X(i)$ for $n = 10000$, $p = 0.25$ has given $M = 5732$ runs. The unnormalized trajectory is given in Figure 2, which shows a "filament silhouette." The normalized trajectory for $\frac{X(i)-\overline{L}i}{\sigma\sqrt{M}}$ and $\frac{X(i-1)-\overline{L}i}{\sigma\sqrt{M}}$, showing the run length is given in Figure 3. Of course, both trajectories are asymptotically equivalent for $n \to \infty$. A zoom on $i = [100\ldots120]$ is given in Figure 4. We have also checked the asymptotic Gaussian property of M as given by (19). This is shown in Figure 5. The asymptotic variance σ_1^2 is given by $0.08604522890\cdots$ and the observed variance by $0.0865\cdots$

3 Hitting time and maximum run length distribution

In this section, we study the hitting time to a length k run and the maximum run length distribution. We also analyze the limiting case $q \to 1$. All vectors and matrices will be written in bold.

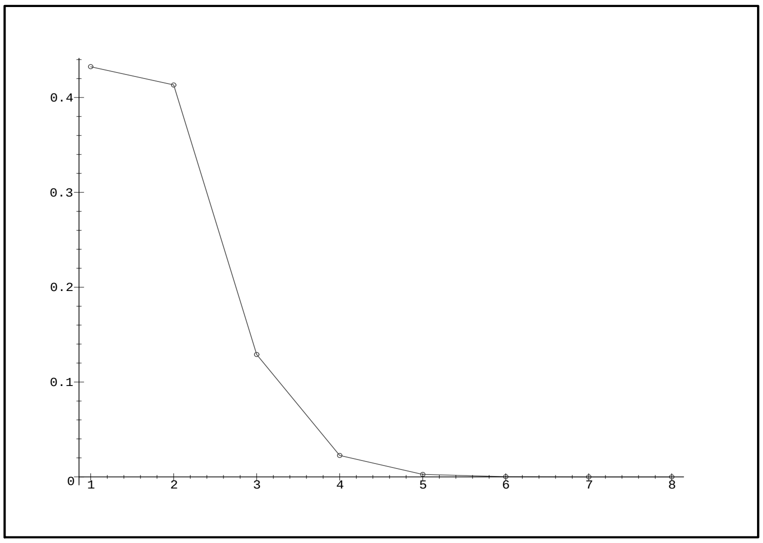

Figure 1: Run length distribution (observed = circle, asymptotic = line)

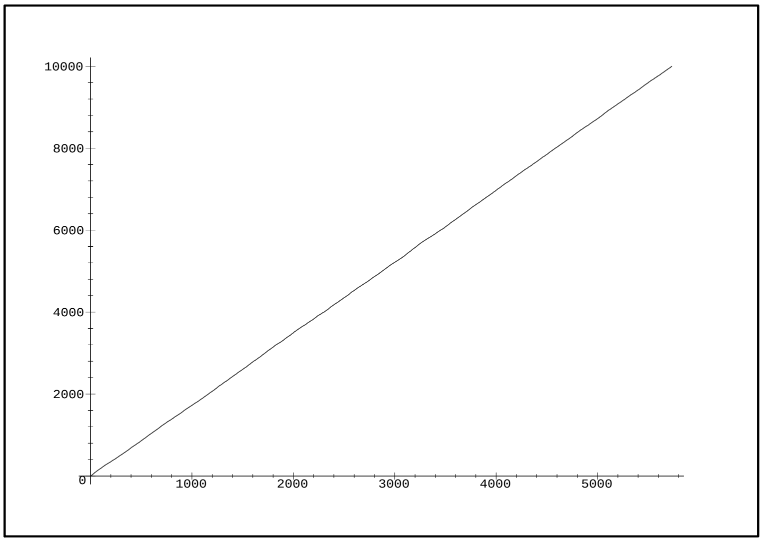

Figure 2: Unnormalized $X(\cdot)$

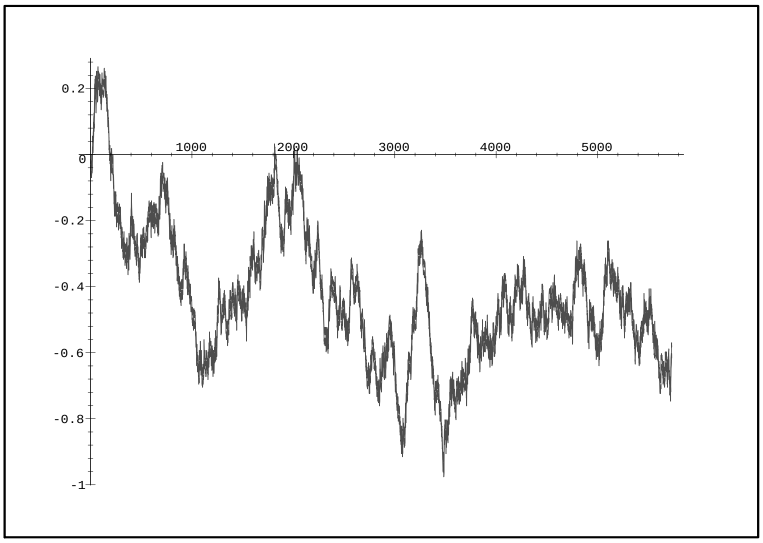

Figure 3: Normalized $X(\cdot)$, with run length

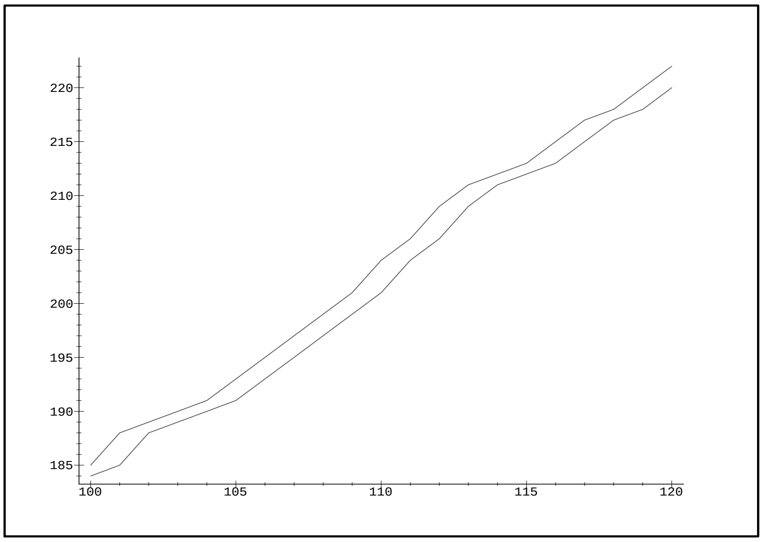

Figure 4: Unnormalized $X(\cdot)$, with run length (zoom)

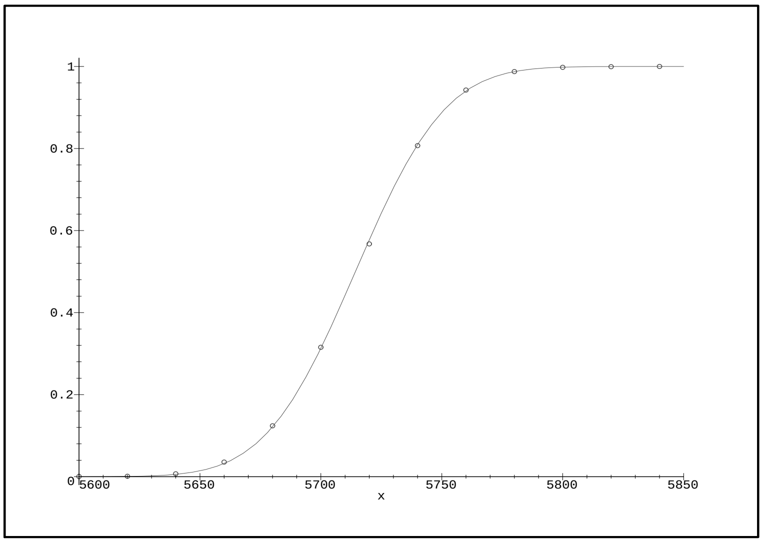

Figure 5: Observed and limiting M distribution function

3.1 GEOM(p) RVs with fixed $p > 0$

Let us define $T_k :=$ the time (counted in terms of number of GEOM(p) RVs, i.e. the length of the word) necessary to obtain a run of length k. We set $h_i := \mathbb{E}_i(T_k)$, (when possible, we drop the k for ease of notation). With (16) and (17), we derive

$$\begin{aligned} h_i &= [\overline{L}(i,k) + kC(i,k-1)] + \sum_{l<k} \sum_u \varphi(i,l,u) h_u \\ &= [\overline{L}(i,k) + kC(i,k-1)] + \sum_u \Pi(i,u) h_u - \sum_{l \geq k} \sum_u \varphi(i,l,u) h_u \\ &= [\overline{L}(i,k) + kC(i,k-1)] + \sum_u \Pi(i,u) h_u - C(i,k-1) \sum_{l \geq k} \sum_u \frac{\varphi(i,l,u)}{C(i,k-1)} h_u. \end{aligned} \tag{20}$$

By (16), we observe that $C(i,k-1)) = \mathcal{O}(\varepsilon)$, $\varepsilon := q^{\binom{k}{2}} p^k$, $k \to \infty$, uniformly on i, and by standard properties of Markov chains (see Keilson [10], Aldous [1]), we know that the hitting time to k is such that (we write D_i for $D_i(k)$):

$$\begin{aligned} h_i &= \frac{D_i}{\varepsilon} + \psi(i) + \mathcal{O}(\varepsilon) \text{ (Actually a Laurent series exists for } \varepsilon \text{ sufficiently small)}, \\ \Pr_i[T_k \geq x] &\sim e^{-x/h_i}, \ x \to \infty. \end{aligned} \tag{21}$$

We will soon check that $\mathbf{D}$ is independent of i. First we note that

$$\overline{L}(i,k) + kC(i,k-1)] = \overline{L}(i) + \mathcal{O}(\varepsilon).$$

Equation (20) leads to

$$\begin{aligned} D_i + \varepsilon\psi(i) &= \varepsilon\overline{L}(i) + \left[\sum_u \Pi(i,u)[D_u + \varepsilon\psi(u)]\right] \\ &- C(i,k-1) \sum_{l \geq k} \sum_u \frac{\varphi(i,l,u)}{C(i,k-1)} [D_u + \varepsilon\psi(u)] + \mathcal{O}(\varepsilon^2) \\ &= \mathbf{\Pi D}(i) + \mathcal{O}(\varepsilon). \end{aligned} \tag{22}$$

Identification of ε powers in (22) leads to $[\mathbf{I}-\mathbf{\Pi}]\mathbf{D}=0$, which confirms that $\mathbf{D}$ is independent of i. Now we premultiply (20) by $\boldsymbol{\pi}$. This leads to

$$\boldsymbol{\pi}\mathbf{h}=\overline{L}+\boldsymbol{\pi}\mathbf{h}-\sum_{i=1}^{\infty}\pi(i)\left[\frac{q^{\binom{k-1}{2}}p^{k-1}q^{(k-1)i}}{(q;q)_{k-1}}\right]\frac{D}{\varepsilon}+\mathcal{O}(\varepsilon),$$

or

$$\begin{aligned}0 &= \overline{L}-\frac{q^{\binom{k-1}{2}}p^{k-1}G(k-1)}{(q;q)_{k-1}}\frac{D}{\varepsilon}+\mathcal{O}(\varepsilon)\\ &= \overline{L}-\frac{q^{\binom{k}{2}}p^{k}(1-q^{2})}{p(q;q)_{k-1}(1-q^{k+1})}\frac{D}{\varepsilon}+\mathcal{O}(\varepsilon) \qquad (23)\end{aligned}$$

by (12). Therefore, we obtain $D=(q;q)_{k-1}(1-q^{k+1})$. We proceed now as in [14]. Let $\mathcal{M}(n):=$ maximum run length based on n GEOM(p) RVs . We know that

$$\Pr[\mathcal{M}(n)<k]=\Pr[T_k\geq n+1].$$

We start our asymptotic analysis by first using (21):

$$\Pr[\mathcal{M}(n)<k]\sim\exp[-\exp[\log n+k\log(p)+k(k-1)/2\log(q)-\log D]],\quad n\to\infty$$

Set now $k:=j+1$, $D(j):=(q;q)_j(1-q^{j+2})$. We derive the following theorem.

Theorem 3.1 *Let*

$$\eta:=j(j+1)/2\log(q^{-1})+j\log(p^{-1})-\log n-\log p+\log(D(j)).$$

Then, with integer j and $\eta=\mathcal{O}(1)$,

$$\Pr[\mathcal{M}(n)\leq j]\sim G_1(\eta),\quad n\to\infty$$

where $G_1(\eta):=\exp[-\exp[-\eta]]$. We also have

$$\Pr[\mathcal{M}(n)=j]\sim f_1(\eta)$$

where $f_1(\eta):=G_1(\eta)-G_1(\eta-1)$.

$G_1(.)$ is actually the extreme-value distribution function. The observed and limiting distribution function (DF) are given in Figure 6 (observed = circle, asymptotic = line).

Let us make two remarks:

- as $j\to\infty$, which is a consequence of $n\to\infty$, we have $D(j)\to(q;q)_\infty$. But $n=10000$ is not large enough to entail this limit: in Figure 6, we have kept the true $D(j)$ value .
- Let $\psi_1(n):=\log[np/D(j)]$. Then with $F(j):=j(j+1)/2\log(q^{-1})+j\log(p^{-1}),\eta=F(j)-\psi_1(n)$, we see that, asymptotically, at most two values of j support the probability measure of $\mathcal{M}(t)$: either for some integer j, $\eta=\mathcal{O}(1)$. Then j and $j+1$ support the probability measure, or no such value of j exists, then $F(j_1-1)\ll\psi_1(n)\ll F(j_1)$ for some j_1, and all probability measure is entirely supported by j_1. We have already encountered this property in diagonally-convex animals (see [14]). The linear case for $F(j)$ is found, for instance, in column-convex animals [14], Carlitz compositions [16], and runs of geometrically distributed RVs [15].

Let us now consider the maximum run length of type u (i.e. starting with value u). We derive

$$\begin{aligned}h_i(u) &= \overline{L}(i)+\sum_j\Pi(i,j)h_j(u),\quad i\neq u,\\ h_u(u) &= [\overline{L}(u,k)+kC(u,k-1)]+\sum_j\Pi(u,j)h_j(u)-C(u,k-1)\sum_{l\geq k}\sum_j\frac{\varphi(u,l,j)}{C(u,k-1)}h_j(u).\end{aligned}$$

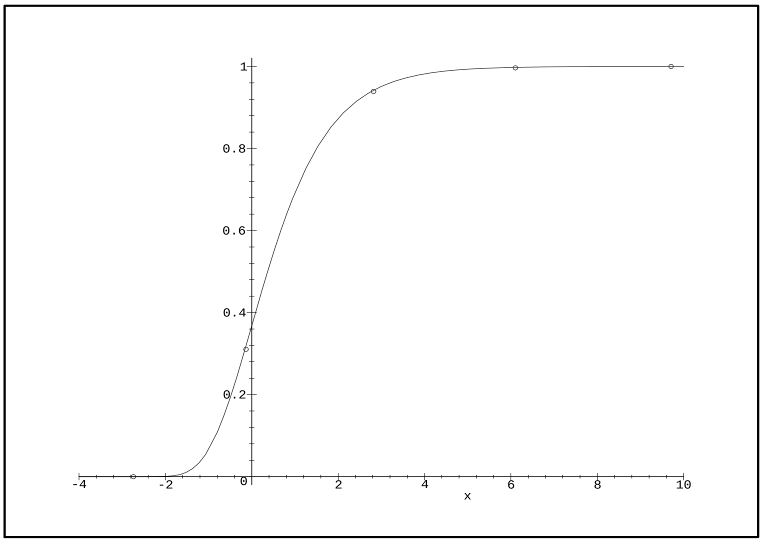

Figure 6: Observed and limiting $\mathcal{M}$ distribution function

Proceeding as above, with $\varepsilon := q^{\binom{k}{2}} p^k q^{(k+1)(u-1)}$, we obtain

$$h_i(u) \sim \frac{D(u)}{\varepsilon} + \mathcal{O}(1),$$

with $D(j,u) = (q;q)_j$ (independently of u). All previous results are easily adapted. We now obtain

$$\eta := j(j+1)/2 \log(q^{-1}) + j \log(p^{-1}) + (j+2)(u-1)\log(q^{-1}) - \log n - \log p + \log(D(j,u)).$$

Note that, again, $D(j,u) \to (q;q)_\infty$. For $u = 1$, the limiting distribution is given by Theorem 3.1. This means that the maximum run length is asymptotically only made of runs of type 1. For $u = 2$, the simulation is given in Figure 7

3.2 GEOM(p) RV, $q \to 1$

When $q \to 1$, the picture is quite different: we use now

$$\varepsilon = \frac{p^k}{(q;q)_{k-1}(1-q^{k+1})} \sim \frac{1}{(k+1)(k-1)!},$$

and (23) leads to $D = 1$. Setting again $k = j+1$, we derive

$$\Pr[\mathcal{M}(n) \le j] \sim \exp[-\exp[\log n - \log(j!) - \log(j+2)]].$$

We must first obtain $\psi(n)$ such that

$$\log n - \log(\Gamma(\psi+1)) - \log(\psi+2) = 0. \tag{24}$$

Using the notation: $Ln := \log n, Ln^{(2)} := \log Ln$, $Ln^{(i+1)} := \log Ln^{(i)}$, and Stirling's formula: $\Gamma(z+1) \sim e^{-z} z^z \sqrt{2\pi z}$, we see that the dominant term of $\psi(n)$ is given by $Ln/Ln^{(2)}$. Proceeding by bootstrapping, this leads to

$$\psi(n) \sim \frac{Ln}{Ln^{(2)}} \left(1 + \frac{Ln^{(3)}}{Ln^{(2)}} + \frac{1}{Ln^{(2)}} + \left[\frac{Ln^{(3)}}{Ln^{(2)}}\right]^2 + \frac{Ln^{(3)}}{(Ln^{(2)})^2} + \cdots \right). \tag{25}$$

This leads to the following theorem:

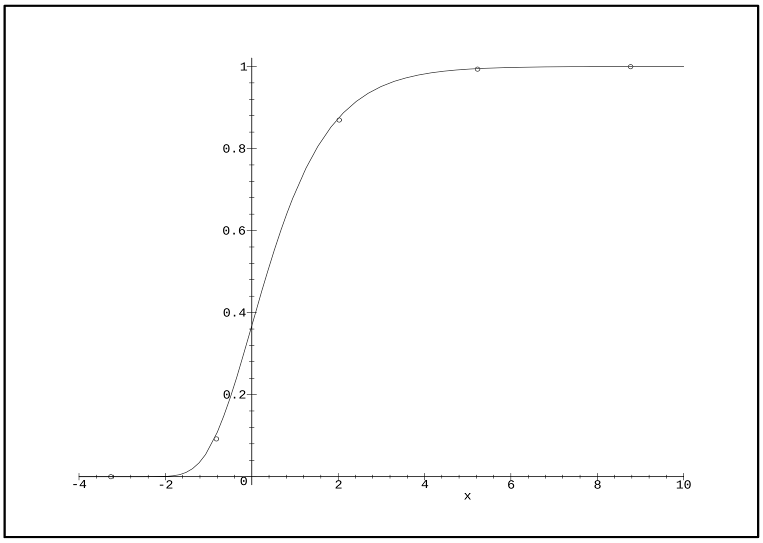

Figure 7: Observed and limiting $\mathcal{M}$ distribution function, u=2

Theorem 3.2 *Let*

$$\eta := (j - \psi(n))Ln^{(2)}.$$

Then, with integer j and $\eta = \mathcal{O}(1)$,

$$\Pr[\mathcal{M}(n) \leq j] \sim G_1(\eta), \quad n \to \infty$$

where $G_1(\eta) := \exp[-\exp[-\eta]]$. We also have

$$\Pr[\mathcal{M}(n) = j] \sim f_1(\eta)$$

where $f_1(\eta) := G_1(\eta) - G_1(\eta - Ln^{(2)})$.

We remark that η depends linearly on j, but, due to the factor $Ln^{(2)}$, again at most two values of j support the probability measure. If we could write the the asymptotic moments of $\mathcal{M}(n)$ as Harmonic sums, they could be analyzed with Mellin transforms: see Flajolet et al. [8] and Flajolet [6]. We could for instance derive

$$\bar{\mathbb{E}}[\mathcal{M}(n)] \sim \psi(n) + \frac{1}{2} + \frac{\gamma}{Ln^{(2)}} \tag{26}$$

We have done a simulation of $U = 2000$ sequences of $n = 10000$ GEOM(10^{-8}) RVs. For the number of runs and the run length, all RV conform to the limiting distributions. By (23), we must look for the solution of

$$\frac{q^{\binom{k}{2}} p^k (1-q^2)}{p(q;q)_{k-1}(1-q^{k+1})(1+q)} \cdot n = 1$$

with $k = \psi + 1$. The solution of (24) is given by $\psi = 6.27\ldots$. Of course, n is not large enough to allow the use of our asymptotic expansion (25). The first 3 terms of (25) lead to $7.50\ldots$. The next term leads to $8.04\ldots$! So we have computed a numerical value of ψ by using (26) and the observed mean of $\mathcal{M}(n)$ given by $7.051\ldots$. This gives $\psi = 6.236\ldots$. The observed and limiting distribution function are given in Figure 8 (observed = circle, asymptotic = line, the picture is shifted by 4 units in order to clearly display all observed points). The fit is quite good.

In Figure 9 we show the observed and limiting run length distribution. The limiting one, given by (14), corresponds to the random permutation.

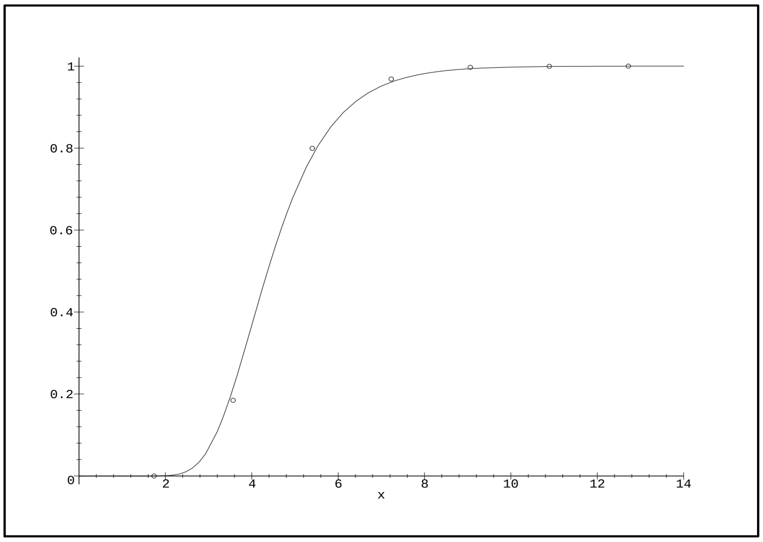

Figure 8: Observed and limiting $\mathcal{M}$ distribution function, $p = 10^{-8}$

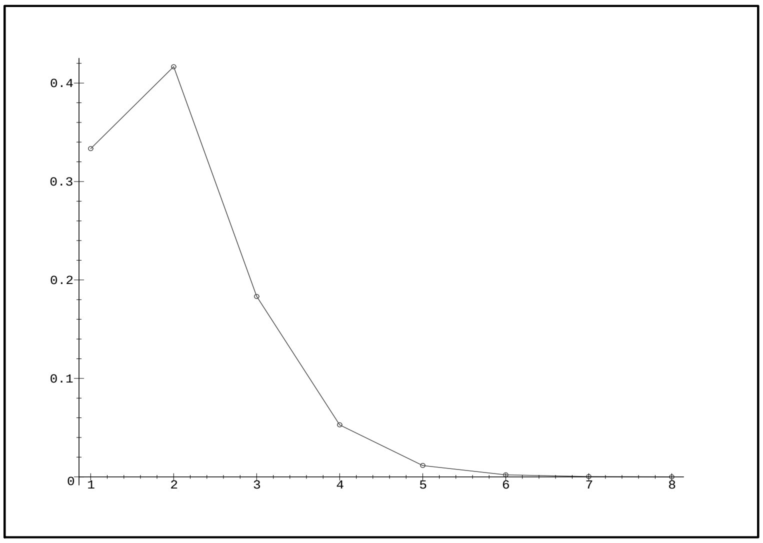

Figure 9: Run length distribution (observed = circle, asymptotic = line) , $p = 10^{-8}$

4 Conclusion

Using a description of the runs either as a stochastic process or as a polyomino, we have obtained several asymptotic probability distributions. We intend to pursue this approach on other runs properties such as weakly ascending runs, alternating runs (see, for instance, Knuth [11, Sec.5.1.3]).

References

[1] D. Aldous. *Probability Approximations via the Poisson Clumping Heuristic.* Springer-Verlag, 1989.

[2] G. Andrews. *The Theory of Partitions*, volume 2 of *Encyclopedia of Mathematics and its Applications.* Addison–Wesley, 1976.

[3] E.A. Bender. Central and local limit theorems applied to asymptotics enumeration. *Journal of Combinatorial Theory,* Series A, 15:91–111, 1973.

[4] P. Billingsley. *Convergence of Probability Measures.* Wiley, 1968.

[5] L. Carlitz. A combinatorial property of q–Eulerian numbers. *American Mathematical Monthly*, 82:51–54, 1975.

[6] P. Flajolet. Approximate counting : a detailed analysis. *BIT*, 25:113–134, 1985.

[7] P. Flajolet. On permutation enumeration via regular languages. *private communication*, 2000.

[8] P. Flajolet, X. Gourdon, and P. Dumas. Mellin transforms and asymptotics: Harmonic sums. *Theoretical Computer Science*, 144:3–58, 1995.

[9] G. Gasper and M. Rahman. Basic hypergeometric series. In *Encyclopedia of Mathematics and Its Applications.* Cambridge University Press, 1990.

[10] J. Keilson. *Markov Chain Models-Rarity and Exponentiality.* Springer-Verlag, 1979.

[11] D. E. Knuth. *The Art of Computer Programming*, volume 3: Sorting and Searching. Addison-Wesley, 1973. Second edition, 1998.

[12] M. Lothaire. *Algebraic Combinatorics on Words.* 2001.

[13] G. Louchard. Probabilistic analysis of some (un)directed animals. *Theoretical Computer Science*, 159(1):65–79, 1996.

[14] G. Louchard. Probabilistic analysis of column-convex and directed diagonally-convex animals. II: Trajectories and shapes. *Random Structures and Algorithms*, 15:1–23, 1999.

[15] G. Louchard. Runs of geometrically distributed random variables: a probabilistic analysis. *To appear in the JCAM special issue on the Analysis of Algorithms*, 2001. http://www.ulb.ac.be/di/mcs/louchard/run.ps.

[16] G. Louchard and H. Prodinger. Probabilistic analysis of Carlitz compositions. *Submitted.* http://www.ulb.ac.be/di/mcs/louchard/GL388.ps and http://www.wits.ac.za/helmut/abstract/abs_155.htm.

[17] G. Louchard and W. Szpankowski. Average profile and limiting distribution for a phrase size in the Lempel-Ziv parsing algorithm. *IEEE Transactions on Information Theory*, 41:478–488, 1995.

[18] H. Prodinger. Combinatorics of geometrically distributed random variables: length of ascending runs. *LNCS*, 1776:473–482, 2000.

[19] W. Rudin. *Real and Complex Analysis.* McGraw-Hill, 1966.