Discrete Mathematics and Theoretical Computer Science DMTCS vol. **8**, 2006, 215–234

The first descent in samples of geometric random variables and permutations

Arnold Knopfmacher[1†] and Helmut Prodinger[2‡]

[1] *The John Knopfmacher Centre for Applicable Analysis and Number Theory, School of Mathematics, University of the Witwatersrand, Private Bag 3, Johannesburg, South Africa.*

[2] *Mathematics Department, Stellenbosch University, 7602 Stellenbosch, South Africa.*

received Mar 31, 2004, *revised Oct 17, 2005*, *accepted Jul 16, 2006*.

For words of length n, generated by independent geometric random variables, we study the average initial and end heights of the first strict and weak descents in the word. Higher moments and limiting distributions are also derived. In addition we compute the average initial and end height of the first descent for a random permutation of n letters.

Keywords: geometric random variables, generating function, limiting distribution, descents

1 Introduction

Let X denote a geometrically distributed random variable, i. e. $\mathbb{P}\{X = k\} = pq^{k-1}$ for $k \in \mathbb{N}$ and $q = 1 - p$. The combinatorics of n geometrically distributed independent random variables $X_1, \ldots, X_n$ has attracted recent interest, especially because of applications in computer science. We mention just two areas, the skip list (1; 6; 12; 15) and probabilistic counting (2; 4; 5; 7).

- *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. (The data structure is non-deterministic and is randomly *constructed* using a geometric random variable.) 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.*

- *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).

†This material is based upon work supported by the National Research Foundation under grant number 2053740
‡This material is based upon work supported by the National Research Foundation under grant number 2053748

One of the first combinatorial questions investigated for words $a_1 \dots a_n$, with the letters a_i independently generated according to the geometric distribution, was the number of left-to-right maxima in (13). In (8) the study of left-to-right maxima was continued, but now the parameters studied were the mean value and mean position of the r-th maximum.

More recently the authors (9) studied the number of descents in a string of n geometrically distributed independent random letters. For example in $w = 225114431$ we have 2 strict and 6 weak descents (equality is included in the weak case). We continue the study of descents in this article by considering the initial and end heights and hence the size (magnitude of the first jump down) of the first descent in such a string. In our example word w the initial and end heights of the first descent are respectively 5 and 1 (strict case) and 2 and 2 (weak case). In this paper strict descents are considered in Section 2 and weak descents in Section 3. The limiting behaviour as $q \to 1$ is investigated in Section 4.

Many permutation statistics can be deduced from the corresponding geometric random variable statistic by letting $q \to 1$. However, this does not apply in the case of descent heights, so we consider separately the initial and end heights of the first descent for a random permutation of n letters in Section 5.

2 Strict descents

2.1 The end height of the first strict descent

For the end height of the first strict descent in a sample of geometrically distributed random variables we use the following decomposition. If a random word has at least one strict descent, then it begins with a weakly increasing sequence of values ending in some value $h \geq 2$, which is followed by a letter smaller than h (the first descent). Thereafter the letters of the word can be an arbitrary sequence of geometrically distributed random variables. Alternatively, a random word has no strict descent at all, in which case it is a weakly increasing word. This gives the following decomposition for the set of all words:

$$\{\textit{all words}\} = \bigcup_{h\geq 2} \Big(\{\textit{weakly increasing word ending with } h\}\{\textit{letter} < h\}\{\textit{any word}\}\Big) \\ \cup\{\textit{weakly increasing word}\}.$$

We now consider a probability generating function $F(z,u)$, where z labels the number of random variables, and u marks the end height of the first strict descent. Consider the case of words with at least one strict descent. Firstly, the generating function for weakly increasing words with values at most h is given by the product $\prod_{j=1}^{h} \frac{1}{1-pq^{j-1}z}$. Then the end letter h of this increasing word has generating function $pq^{h-1}z$. The subsequent value (first descent) is marked by u and has any value from 1 to $h-1$, with generating function $\sum_{i=1}^{h-1} pq^{i-1}u^i z$. The arbitrary word that follows the first descent then has generating function $\frac{1}{1-z}$. Finally, words with no descent have generating function $\prod_{j\geq 1} \frac{1}{1-pq^{j-1}z}$.

This leads to

$$\begin{aligned} F(z,u) &= \sum_{h\geq 2}\prod_{j=1}^{h} \frac{1}{1-pq^{j-1}z} pq^{h-1}z\Big(\sum_{i=1}^{h-1} pq^{i-1}u^i z\Big)\frac{1}{1-z} + \prod_{j\geq 1}\frac{1}{1-pq^{j-1}z} \\ &= \frac{1}{1-z}\sum_{h\geq 2}\prod_{j=1}^{h}\frac{1}{1-pq^{j-1}z} pq^{h-1}\frac{puz^2(1-q^{h-1}u^{h-1})}{1-qu} + \prod_{j\geq 1}\frac{1}{1-pq^{j-1}z}. \end{aligned} \tag{2.1}$$

Remark. By definition $F(z,1)=\frac{1}{1-z}$, however, in order to deduce this from the above expression for $F(z,u)$ we require the following proposition whose proof we relegate to the Appendix 6:

Proposition 1

$$\sum_{h\geq 1} pzq^{2h}\prod_{j=1}^{h}\frac{1}{1-pq^jz}=\frac{1}{z}\Big(\prod_{j\geq 1}\frac{1}{1-pq^jz}-1\Big)-q. \tag{2.2}$$

We will also make use of Proposition 1 in order to simplify some of our expressions for the mean values of parameters.

For the mean end height of the first descent we must compute $\frac{\partial}{\partial u}F(z,1)$. Now

$$\frac{\partial}{\partial u}F(z,1)=\frac{z^2}{1-z}\sum_{h\geq 2}q^{h-1}\big(1-hq^{h-1}+(h-1)q^h\big)\prod_{j=1}^{h}\frac{1}{1-pq^{j-1}z}. \tag{2.3}$$

Since the dominant pole is at $z=1$ we have by singularity analysis (3),

$$\begin{aligned}[z^n]\frac{\partial}{\partial u}F(z,1) &\sim \sum_{h\geq 1}q^{h-1}\big(1-(h+1)q^h+hq^{h+1}\big)\prod_{j=1}^{h}\frac{1}{1-pq^j}\\ &=\frac{1}{pq}\sum_{h\geq 1}\big(pq^h-pq^{2h}+(q-1)hpq^{2h}\big)\prod_{j=1}^{h}\frac{1}{1-pq^j}\\ &=\frac{1}{p}+(q-1)\sum_{h\geq 1}hq^{2h-1}\prod_{j=1}^{h}\frac{1}{1-pq^j},\end{aligned}$$

by using (2.2) and one of Euler's partition identities:

$$\sum_{n\geq 0}\frac{x^n}{(q)_n}=\frac{1}{(x)_n}. \tag{2.4}$$

For notational convenience here and below, we make use of the q-series notation $(x)_n := (1-x)(1-xq)\dots(1-xq^{n-1})$. Thus we have shown

Theorem 1 *The expected end height of the first strict descent is asymptotically as $n\to\infty$,*

$$\frac{1}{p}-\sum_{h\geq 1}hpq^{2h-1}\frac{1}{(pq)_h}. \tag{2.5}$$

2.1.1 Higher moments and limiting distributions

From (2.1), the r-th moment about the origin of the end height for $r\geq 1$ is just

$$\frac{p^2z^2}{1-z}\sum_{h\geq 2}\sum_{k\geq 0}\big((k+1)^rq^k-(k+h)^rq^{k+h-1}\big)\prod_{j=1}^{h}\frac{1}{1-pq^{j-1}z}. \tag{2.6}$$

Since the dominant pole is at $z = 1$, singularity analysis implies that this is asymptotic to the constant

$$\nu_r^s(q) := p^2 \sum_{h\ge 2}\sum_{k\ge 0}\big((k+1)^r q^k - (k+h)^r q^{k+h-1}\big)\prod_{j=1}^{h}\frac{1}{1-pq^{j-1}}.$$

For the distribution, if $F(z,u) = \sum_{n\ge 0} p_n(u) z^n$ then again by singularity analysis around $z = 1$

$$p_n(u) \to p_E^s(u) := \sum_{h\ge 2}\prod_{j=1}^{h}\frac{1}{1-pq^{j-1}}pq^{h-1}\frac{pu(1-q^{h-1}u^{h-1})}{1-qu}. \tag{2.7}$$

So the end descent height has a discrete limiting distribution.

2.2 The initial height and size of the first strict descent

We now consider a probability generating function $F(z,u)$, where z labels the number of random variables, and u marks the initial height of the first strict descent.

Let us use the following decomposition again

$\{$*all words*$\}$= $\bigcup_{h\ge 2}\Big(\{$*weakly increasing word ending with* $h\}\{$*letter*$< h\}\{$*any word*$\}\Big)$

$\cup\{$*weakly increasing word*$\}$.

This leads to

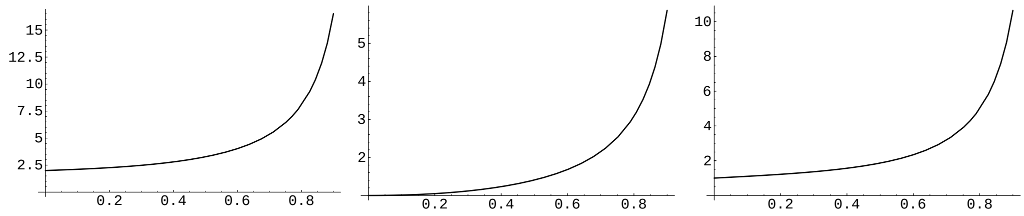
$$\begin{aligned}F(z,u) &= \sum_{h\ge 2}\prod_{j=1}^{h}\frac{1}{1-pq^{j-1}z}pu^hq^{h-1}z\Big(\sum_{i=1}^{h-1}pq^{i-1}z\Big)\frac{1}{1-z} + \prod_{j\ge 1}\frac{1}{1-pq^{j-1}z}\\ &= \frac{pz^2}{1-z}\sum_{h\ge 2}u^hq^{h-1}(1-q^{h-1})\prod_{j=1}^{h}\frac{1}{1-pq^{j-1}z} + \prod_{j\ge 1}\frac{1}{1-pq^{j-1}z}. \end{aligned}\tag{2.8}$$

For the mean initial height of the first descent we must compute $\frac{\partial}{\partial u}F(z,1)$. Now

$$\frac{\partial}{\partial u}F(z,1) = \frac{pz^2}{1-z}\sum_{h\ge 2}hq^{h-1}(1-q^{h-1})\prod_{j=1}^{h}\frac{1}{1-pq^{j-1}z}. \tag{2.9}$$

Since the dominant pole is at $z = 1$ we have

$$\begin{aligned}[z^n]\frac{\partial}{\partial u}F(z,1) &\sim \frac{p}{q}\sum_{h\ge 1}(h+1)q^h(1-q^h)\prod_{j=1}^{h}\frac{1}{1-pq^j}\\ &= \sum_{h\ge 1}hpq^{h-1}\frac{1}{(pq)_h} - \sum_{h\ge 1}hpq^{2h-1}\frac{1}{(pq)_h} + \frac{1}{q}\sum_{h\ge 1}pq^h\frac{1}{(pq)_h} - \frac{1}{q}\sum_{h\ge 1}pq^{2h}\frac{1}{(pq)_h}\\ &= \sum_{h\ge 1}hpq^{h-1}\frac{1}{(pq)_h} - \sum_{h\ge 1}hpq^{2h-1}\frac{1}{(pq)_h} + 1\end{aligned}$$



(a) Mean initial height (b) End height (c) Size of the first strict descent

Fig. 1: *Mean initial height, end height and size of the first strict descent as a function of q.*

by using (2.2) and Euler's partition identity (2.4).

Therefore we have shown

Theorem 2 *The expected initial height of the first strict descent is asymptotically as $n \to \infty$,*

$$\sum_{h\geq 1} hpq^{h-1}\frac{1}{(pq)_h} - \sum_{h\geq 1} hpq^{2h-1}\frac{1}{(pq)_h} + 1. \tag{2.10}$$

The mean size of the first descent is obtained by subtracting the mean end height from the mean initial height of the first descent. From Theorems 1 and 2 we obtain

Theorem 3 *The expected size of the first strict descent is asymptotically as $n \to \infty$,*

$$\sum_{h\geq 1} hpq^{h-1}\frac{1}{(pq)_h} - \frac{q}{p}. \tag{2.11}$$

Figure 1 illustrates the dependence on q of the mean initial height 1(a), end height 1(b) and size of the first strict descent 1(c). As $q \to 0$, the initial height, end height and size tend, respectively to 2, 1 and 1. These parameters then grow as q increases. The growth as $q \to 1$ will be quantified in Section 4.

2.2.1 Higher moments and limiting distributions

The r-th moment about the origin of the initial height is just

$$\frac{pz^2}{1-z}\sum_{h\geq 2} h^r q^{h-1}(1-q^{h-1})\prod_{j=1}^{h}\frac{1}{1-pq^{j-1}z}. \tag{2.12}$$

Since the dominant pole is at $z = 1$, singularity analysis implies that this is asymptotic to the constant

$$\mu_r^s(q) := \frac{p}{q}\sum_{h\geq 1}(h+1)^r q^h(1-q^h)\prod_{j=1}^{h}\frac{1}{1-pq^j}.$$

For the distribution, if $F(z,u) = \sum_{n\geq 0} p_n(u)z^n$ then again by singularity analysis around $z = 1$

$$p_n(u) \to p_I^s(u) := \sum_{h\geq 2} pu^h q^{h-1}(1-q^{h-1})\prod_{j=1}^{h}\frac{1}{1-pq^{j-1}}. \tag{2.13}$$

So the initial descent height also has a discrete limiting distribution.

3 Weak descents

3.1 The end height of the first weak descent

We now consider a probability generating function $F(z,u)$, where z labels the number of random variables, and u marks the end height of the first weak descent.

We use the following decomposition

$$\{\textit{all words}\} = \bigcup_{h\ge1}\Big(\{\textit{strictly increasing word ending with } h\}\{\textit{letter} \le h\}\{\textit{any word}\}\Big)$$
$$\cup\{\textit{strictly increasing word}\}.$$

This leads to

$$\begin{aligned}F(z,u)&=\sum_{h\ge1}\prod_{j=1}^{h-1}(1+pq^{j-1}z)pq^{h-1}z\Big(\sum_{i=1}^{h}pq^{i-1}u^iz\Big)\frac{1}{1-z}+\prod_{j\ge1}(1+pq^{j-1}z)\\&=\frac{1}{1-z}\sum_{h\ge1}\prod_{j=1}^{h-1}(1+pq^{j-1}z)pq^{h-1}\frac{puz^2(1-q^hu^h)}{1-qu}+\prod_{j\ge1}(1+pq^{j-1}z).\end{aligned}\tag{3.1}$$

Once again to verify that $F(z,u)$ gives $\frac{1}{1-z}$ when $u=1$ requires an identity (also proved in the Appendix 6):

Proposition 2

$$\sum_{h\ge0}pzq^{2h+1}\prod_{j=0}^{h-1}(1+pq^jz)=\frac{1}{z}\Big(\prod_{j\ge0}(1+pq^jz)-1\Big)-1.\tag{3.2}$$

For the mean end height of the first weak descent we must compute $\frac{\partial}{\partial u}F(z,1)$. Now

$$\frac{\partial}{\partial u}F(z,1)=\frac{z^2}{1-z}\sum_{h\ge1}q^{h-1}\big(1-(h+1)q^h+hq^{h+1}\big)\prod_{j=1}^{h-1}(1+pq^{j-1}z).\tag{3.3}$$

Since the dominant pole is at $z=1$ we have

$$\begin{aligned}[z^n]\frac{\partial}{\partial u}F(z,1)&\sim\sum_{h\ge1}q^{h-1}\big(1-(h+1)q^h+hq^{h+1}\big)\prod_{j=1}^{h-1}(1+pq^{j-1})\\&=\sum_{h\ge0}\big(q^h-q^{2h+1}\big)\prod_{j=0}^{h-1}(1+pq^j)+(q-1)\sum_{h\ge1}hq^{2h-1}\prod_{j=0}^{h-2}(1+pq^j)\\&=\frac{1}{p}+(q-1)\sum_{h\ge0}(h+1)q^{2h+1}\prod_{j=0}^{h-1}(1+pq^j),\end{aligned}$$

by using (3.2). Hence

Theorem 4 *The expected end height of the first weak descent is asymptotically as $n \to \infty$,*

$$\frac{1}{p} - \sum_{h \geq 0} (h+1) p q^{2h+1} (-p)_h. \tag{3.4}$$

3.1.1 Higher moments and limiting distributions

The r-th moment about the origin of the weak end height for $r \geq 1$ is

$$\frac{p^2 z^2}{1-z} \sum_{h \geq 1} \sum_{k \geq 0} \big((k+1)^r q^k - (k+h+1)^r q^{k+h}\big) \prod_{j=1}^{h-1} (1 + p q^{j-1} z). \tag{3.5}$$

Since the dominant pole is at $z = 1$, singularity analysis implies that this is asymptotic to the constant

$$\nu_r^w(q) := p^2 \sum_{h \geq 1} \sum_{k \geq 0} \big((k+1)^r q^k - (k+h+1)^r q^{k+h}\big) \prod_{j=1}^{h-1} (1 + p q^{j-1}).$$

For the distribution, if $F(z,u) = \sum_{n \geq 0} p_n(u) z^n$ then again by singularity analysis around $z = 1$

$$p_n(u) \to p_E^w(u) := \sum_{h \geq 1} \prod_{j=1}^{h-1} (1 + p q^{j-1}) p q^{h-1} \frac{pu(1 - q^h u^h)}{1 - qu}. \tag{3.6}$$

So the weak end descent height has a discrete limiting distribution.

3.2 The initial height and size of the first weak descent

We now consider a probability generating function $F(z,u)$, where z labels the number of random variables, and u marks the initial height of the first weak descent.

We use again the decomposition

$\{$*all words*$\} = \cup_{h \geq 1} \Big(\{$*strictly increasing word ending with* $h\}\{$*letter*$\leq h\}\{$*any word*$\} \Big)$
$\cup \{$*strictly increasing word*$\}$.

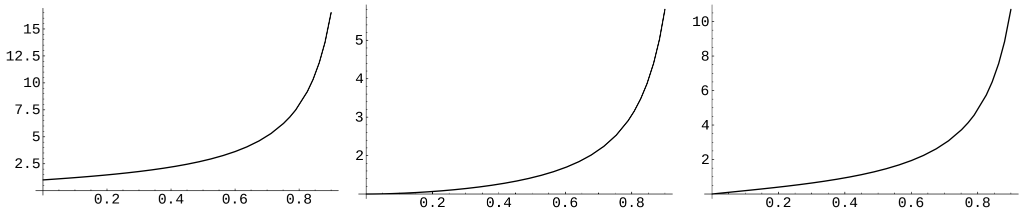
This leads to

$$F(z,u) = \sum_{h \geq 1} \prod_{j=1}^{h-1} (1 + p q^{j-1} z) p q^{h-1} u^h z \Big(\sum_{i=1}^{h} p q^{i-1} z \Big) \frac{1}{1-z} + \prod_{j \geq 1} (1 + p q^{j-1} z)$$

$$= \frac{z^2}{1-z} \sum_{h \geq 1} p q^{h-1} u^h (1 - q^h) \prod_{j=1}^{h-1} (1 + p q^{j-1} z) + \prod_{j \geq 1} (1 + p q^{j-1} z). \tag{3.7}$$

Now

$$\frac{\partial}{\partial u} F(z,1) = \frac{z^2}{1-z} \sum_{h \geq 1} h p q^{h-1} (1 - q^h) \prod_{j=1}^{h-1} (1 + p q^{j-1} z). \tag{3.8}$$



(a) Mean initial height (b) End height (c) Size of the first weak descent

Fig. 2: *Mean initial height, end height and size of the first weak descent as a function of q.*

Since the dominant pole is at $z = 1$ we have

$$[z^n]\frac{\partial}{\partial u}F(z,1) \sim \sum_{h\ge 0}(h+1)pq^h(1-q^{h+1})\prod_{j=0}^{h-1}(1+pq^j)$$
$$= \sum_{h\ge 0}(h+1)pq^h(-p)_h - \sum_{h\ge 0}(h+1)pq^{2h+1}(-p)_h.$$

Hence

Theorem 5 *The expected initial height of the first weak descent is asymptotically as $n \to \infty$,*

$$\sum_{h\ge 0}(h+1)pq^h(-p)_h - \sum_{h\ge 0}(h+1)pq^{2h+1}(-p)_h. \tag{3.9}$$

From Theorems 4 and 5 we obtain

Theorem 6 *The expected size of the first weak descent is asymptotically as $n \to \infty$,*

$$\sum_{h\ge 0}(h+1)pq^h(-p)_h - \frac{1}{p}. \tag{3.10}$$

Figure 2 illustrates the dependence on q of the mean initial height 2(a), end height 2(b) and size of the first weak descent 2(c).

3.2.1 Higher moments and limiting distributions

The r-th moment about the origin of the initial weak height is

$$\frac{pz^2}{1-z}\sum_{h\ge 1}h^r q^{h-1}(1-q^h)\prod_{j=1}^{h-1}(1+pq^{j-1}z) \tag{3.11}$$

and singularity analysis implies that this is asymptotic to the constant

$$\mu_r^w(q) := \frac{p}{q}\sum_{h\ge 1}h^r q^h(1-q^h)\prod_{j=1}^{h-1}(1+pq^{j-1}).$$

For the distribution, if $F(z,u)=\sum_{n\geq 0}p_n(u)z^n$ then again by singularity analysis around $z=1$

$$p_n(u)\to p_I^w(u):=\sum_{h\geq 1}pu^hq^{h-1}(1-q^h)\prod_{j=1}^{h-1}(1+pq^{j-1}). \tag{3.12}$$

So once again the initial weak descent height has a discrete limiting distribution.

4 The behaviour for $q\to 1$

The figures indicate that the quantities in Theorems 1, 2 and 3, as well as in Theorems 4, 5 and 6 tend to infinity as $q\to 1$. We will quantify this observation in this section.

We will make use of Heine's formula

$$\sum_{m\geq 0}\frac{(a)_m(b)_mt^m}{(q)_m(c)_m}=\frac{(b)_\infty(at)_\infty}{(c)_\infty(t)_\infty}\sum_{m\geq 0}\frac{(c/b)_m(t)_m}{(q)_m(at)_m}(b)^m. \tag{4.1}$$

We start our study with a sum that arises in Theorems 2 and 3,

$$R_1:=\sum_{h\geq 1}hpq^{h-1}\frac{1}{(pq)_h}=\frac{p}{q}\frac{d}{dz}\sum_{h\geq 0}(zq)^h\frac{1}{(pq)_h}\bigg|_{z=1}.$$

Now we apply Heine's formula (4.1), with $t=zq$, $a=0$, $b=q$, and $c=pq$:

$$\begin{aligned}R_2&:=\sum_{h\geq 0}(zq)^h\frac{1}{(pq)_h}\\&=\frac{(q)_\infty}{(pq)_\infty(zq)_\infty}\sum_{m\geq 0}\frac{(p)_m(zq)_m}{(q)_m}q^m.\end{aligned}$$

A further application of Heine's formula, with $t=q$, $a=zq$, $b=p$, and $c=0$ turns this into

$$\begin{aligned}R_2&=\frac{(q)_\infty}{(pq)_\infty(zq)_\infty}\frac{(p)_\infty(zq^2)_\infty}{(q)_\infty}\sum_{m\geq 0}\frac{(q)_m}{(q)_m(zq^2)_m}p^m\\&=\frac{q}{1-zq}\sum_{m\geq 0}\frac{p^m}{(zq^2)_m}\\&=\frac{q}{p}\sum_{m\geq 1}\frac{p^m}{(zq)_m}.\end{aligned}$$

Thus

$$R_1=\frac{d}{dz}\sum_{m\geq 1}\frac{p^m}{(zq)_m}\bigg|_{z=1}=\sum_{m\geq 1}\frac{p^m}{(q)_m}\sum_{i=1}^m\frac{q^i}{1-q^i}.$$

At this stage, we can perform the limit for $q \to 1$ and find

$$\lim_{q\to1} R_1 \cdot (1-q) = \sum_{m\ge1} \frac{1}{m!} \sum_{i=1}^{m} \frac{1}{i} = \sum_{m\ge1} \frac{H_m}{m!} = 2.165382215\ldots.$$

The next computation of a sum that arises in Theorems 1 and 2, is similar:

$$R_3 := \sum_{h\ge1} hpq^{2h-1} \frac{1}{(pq)_h} = \frac{p}{q} \frac{d}{dz} \sum_{h\ge0} (zq^2)^h \frac{1}{(pq)_h} \bigg|_{z=1}.$$

We apply Heine's formula, with $t = zq^2$, $a = 0$, $b = q$, and $c = pq$:

$$\begin{aligned} R_4 &:= \sum_{h\ge0} (zq^2)^h \frac{1}{(pq)_h} \\ &= \frac{(q)_\infty}{(pq)_\infty (zq^2)_\infty} \sum_{m\ge0} \frac{(p)_m (zq^2)_m}{(q)_m} q^m. \end{aligned}$$

A further application of Heine's formula, with $t = q$, $a = zq^2$, $b = p$, and $c = 0$ turns this into

$$\begin{aligned} R_4 &= \frac{(q)_\infty}{(pq)_\infty (zq^2)_\infty} \frac{(p)_\infty (zq^3)_\infty}{(q)_\infty} \sum_{m\ge0} \frac{(q)_m}{(q)_m (zq^3)_m} p^m \\ &= \frac{q}{1-zq^2} \sum_{m\ge0} \frac{p^m}{(zq^3)_m} \\ &= \frac{q}{p} \sum_{m\ge1} \frac{p^m}{(zq^2)_m}. \end{aligned}$$

Thus

$$R_3 = \frac{d}{dz} \sum_{m\ge1} \frac{p^m}{(zq^2)_m} \bigg|_{z=1} = \sum_{m\ge1} \frac{p^m}{(q^2)_m} \sum_{i=2}^{m+1} \frac{q^i}{1-q^i}.$$

We can perform the limit for $q \to 1$ and find

$$\lim_{q\to1} R_3 \cdot (1-q) = \sum_{m\ge1} \frac{1}{(m+1)!} \sum_{i=2}^{m+1} \frac{1}{i} = \sum_{m\ge2} \frac{H_m - 1}{m!} = \sum_{m\ge1} \frac{H_m}{m!} - e + 1,$$

From this we infer e. g., that the quantity in Theorem 2 behaves for $q \to 1$ as

$$\sum_{h\ge1} hpq^{h-1} \frac{1}{(pq)_h} - \sum_{h\ge1} hpq^{2h-1} \frac{1}{(pq)_h} + 1 \sim \frac{e-1}{1-q}.$$

Next, we consider a sum arising in Theorems 5 and 6,

$$R_5 := \sum_{h\ge0} (h+1)pq^h (-p)_h = \frac{d}{dz} pz \sum_{h\ge0} (zq)^h (-p)_h \bigg|_{z=1}.$$

We use Heine's formula; $t = zq$, $a = -p$, $b = q$, $c = 0$:

$$\begin{aligned} R_6 &:= \sum_{h\ge 0} (zq)^h (-p)_h \\ &= \frac{(q)_\infty(-pqz)_\infty}{(zq)_\infty} \sum_{m\ge 0} \frac{(zq)_m}{(q)_m(-pqz)_m} q^m. \end{aligned}$$

We use Heine's formula again; $t = q$, $a = zq$, $b = 0$, $c = -pqz$. For this, $(c/b)_m b^m$ has to be interpreted as $(b-c)(b-c)\dots(b-cq^{m-1})$, after which b can be replaced by zero;

$$\begin{aligned} R_6 &= \frac{(q)_\infty(-pqz)_\infty}{(zq)_\infty} \sum_{m\ge 0} \frac{(zq)_m}{(q)_m(-pqz)_m} q^m \\ &= \frac{(q)_\infty(-pqz)_\infty}{(zq)_\infty} \frac{(zq^2)_\infty}{(-pqz)_\infty(q)_\infty} \sum_{m\ge 0} \frac{(-1)^m q^{\binom{m}{2}} (-pqz)^m (q)_m}{(q)_m (zq^2)_m} \\ &= \frac{1}{1-zq} \sum_{m\ge 0} \frac{q^{\binom{m}{2}} (pqz)^m}{(zq^2)_m} \\ &= \frac{1}{pz} \sum_{m\ge 1} \frac{q^{\binom{m}{2}} (pz)^m}{(zq)_m}. \end{aligned}$$

Hence

$$R_5 = \frac{d}{dz} \sum_{m\ge 1} \frac{q^{\binom{m}{2}}(pz)^m}{(zq)_m}\bigg|_{z=1} = \sum_{m\ge 1} \frac{q^{\binom{m}{2}} p^m}{(q)_m} \left[m + \sum_{i=1}^{m} \frac{q^i}{1-q^i} \right].$$

Thus

$$\lim_{q\to 1} R_5 \cdot (1-q) = \sum_{m\ge 1} \frac{H_m}{m!} = 2.165382215.$$

Finally, a similar calculation for a sum found in Theorems 4 and 5:

$$R_7 := \sum_{h\ge 0} (h+1) p q^{2h+1} (-p)_h = \frac{d}{dz} pqz \sum_{h\ge 0} (zq^2)^h (-p)_h \Big|_{z=1}.$$

We use Heine's formula; $t = zq^2$, $a = -p$, $b = q$, $c = 0$:

$$\begin{aligned} R_8 &:= \sum_{h\ge 0} (zq^2)^h (-p)_h \\ &= \frac{(q)_\infty(-pqz^2)_\infty}{(zq^2)_\infty} \sum_{m\ge 0} \frac{(zq^2)_m}{(q)_m(-pqz^2)_m} q^m. \end{aligned}$$

We use Heine's formula again; $t = q$, $a = zq^2$, $b = 0$, $c = -pqz^2$:

$$\begin{aligned}
R_8 &= \frac{(q)_\infty(-pqz^2)_\infty}{(zq^2)_\infty} \frac{(zq^3)_\infty}{(-pqz^2)_\infty(q)_\infty} \sum_{m\ge 0} \frac{(-1)^m q^{\binom{m}{2}}(-pq^2z)^m (q)_m}{(q)_m (zq^3)_m} \\
&= \frac{1}{1-zq^2} \sum_{m\ge 0} \frac{q^{\binom{m}{2}}(pq^2z)^m}{(zq^3)_m} \\
&= \frac{1}{pqz} \sum_{m\ge 1} \frac{q^{\binom{m}{2}}(pqz)^m}{(zq^2)_m}.
\end{aligned}$$

Hence

$$R_7 = \frac{d}{dz} \sum_{m\ge 1} \frac{q^{\binom{m}{2}}(pqz)^m}{(zq^2)_m}\bigg|_{z=1} = \sum_{m\ge 1} \frac{q^{\binom{m}{2}}(pq)^m}{(q^2)_m}\left[m + \sum_{i=2}^{m+1} \frac{q^i}{1-q^i}\right].$$

Thus

$$\lim_{q\to 1} R_7 \cdot (1-q) = \sum_{m\ge 1} \frac{H_{m+1}-1}{(m+1)!} = \sum_{m\ge 1} \frac{H_m}{m!} - e + 1.$$

5 The height of the first descent in permutations

5.1 Initial height of first descent in permutations

Various permutation statistics, such as the number of descents (9) can be deduced from the corresponding geometic random variable statistic by letting $q \to 1$. However, as $q \to 1$, the values of the geometric sample do not correspond to those of the set $\{1, 2, \ldots, n\}$, so one cannot deduce the initial height of the first descent in a permutation of the set $\{1, 2, \ldots, n\}$ from the corresponding geometic random variable statistic. We therefore consider this question separately.

Suppose that this initial descent height is k. If this occurs in position $j + 1$ then we have $k - 1$ choices for the element that follows k in the permutation and $\binom{k-2}{j}$ choices for the elements that form the strictly increasing sequence of $j - 1$ values that preceed k. There are $(n - j - 2)!$ arrangements of the remaining elements to complete the permutation. Hence the total number of permutations with initial descent height k is

$$\sum_{j=0}^{k-2} (k-1)\binom{k-2}{j}(n-j-2)!\,. \tag{5.1}$$

As all permutations of n-letters other than $\{1, 2, \ldots, n\}$ have a descent, if we sum (5.1) over k we obtain the identity

$$\sum_{k=2}^{n} (k-1) \sum_{j=0}^{k-2} \binom{k-2}{j}(n-j-2)! = n! - 1. \tag{5.2}$$

A direct proof is as follows

$$
\begin{aligned}
A &:= \sum_{k=2}^{n}(k-1)\sum_{j=0}^{k-2}\binom{k-2}{j}(n-j-2)! \\
&= \sum_{j=0}^{n-2}\frac{(n-j-2)!}{j!}\sum_{k=j+2}^{n}\frac{(k-1)!}{(k-2-j)!} \\
&= \sum_{j=0}^{n-2}(n-j-2)!(j+1)\sum_{k=j+2}^{n}\binom{k-1}{j+1} \\
&= \sum_{j=0}^{n-2}(n-j-2)!(j+1)\binom{n}{j+2} \\
&= n!\sum_{j=0}^{n-2}\Big[\frac{1}{(j+1)!}-\frac{1}{(j+2)!}\Big] = n!-1.
\end{aligned}
$$

The average initial height is then given by the expression

$$
\frac{h(n)}{n!} := \frac{1}{n!}\sum_{k=2}^{n}k(k-1)\sum_{j=0}^{k-2}\binom{k-2}{j}(n-j-2)!\,. \tag{5.3}
$$

Theorem 7 *We have*

$$
\frac{h(n)}{n!} = (n+1)\Big[\sum_{j=0}^{n-1}\frac{1}{j!}-2+\frac{2}{(n+1)!}\Big] \sim (e-2)(n+1),
$$

as $n \to \infty$*, where* $e-2 = 0.7182818284\ldots$*.*

Proof: We have

$$
\begin{aligned}
B &:= \frac{1}{n!}\sum_{k=2}^{n} k(k-1)\sum_{j=0}^{k-2}\binom{k-2}{j}(n-j-2)! \\
&= \frac{1}{n!}\sum_{j=0}^{n-2}\frac{(n-j-2)!}{j!}\sum_{k=j+2}^{n}\frac{k!}{(k-2-j)!} \\
&= \frac{1}{n!}\sum_{j=0}^{n-2}(n-j-2)!(j+1)(j+2)\sum_{k=j+2}^{n}\binom{k}{j+2} \\
&= \frac{1}{n!}\sum_{j=0}^{n-2}(n-j-2)!(j+1)(j+2)\binom{n+1}{j+3} \\
&= (n+1)\sum_{j=0}^{n-2}\frac{(j+1)(j+2)}{(j+3)!} \\
&= (n+1)\sum_{j=0}^{n-2}\Big[\frac{1}{(j+1)!}-\frac{2}{(j+2)!}+\frac{2}{(j+3)!}\Big] \\
&= (n+1)\Big[\sum_{j=0}^{n-1}\frac{1}{j!}-2+\frac{2}{(n+1)!}\Big] \sim (n+1)(e-2).
\end{aligned}
$$

$\square$

Remark. We note that $h(n)$ satisfies the recurrence
$h(1) = 0$, $h(n) = (n+1)h(n-1) + n(n-1)$ for $n \geq 2$. In addition one easily sees that$(n+1)!(e-2) - h(n) = n + \delta_n$, where $0 < \delta_n \leq \delta_1 < 1$. This gives the formula $h(n) = \lfloor (e-2)(n+1)! - n \rfloor$ for $n \geq 1$.

The $h(n)$ sequence

$$0, 2, 14, 82, 512, 3614, 28954, 260642, 2606492, 28671502, 344058134, 4472755874, \ldots$$

is not in Sloane (16).

5.2 End height of first descent in permutations

As in the previous section, one cannot deduce the end height of the first descent in a permutation of the set $\{1, 2, \ldots, n\}$ from the corresponding geometic random variable statistic, as $q \to 1$.

Suppose that this end descent height is m and the initial descent height is $k > m$. If k occurs in position $j+1$ then we have $n-m$ choices for the value of k and $\binom{k-2}{j}$ choices for the elements that form the strictly increasing sequence of $j-1$ values that preceed k. There are $(n-j-2)!$ arrangements of the remaining elements to complete the permutation. Hence the total number of permutations with end descent height m is

$$\sum_{k=m+1}^{n}\sum_{j=0}^{k-2}\binom{k-2}{j}(n-j-2)!\,. \tag{5.4}$$

As all permutations of n-letters other than $\{1,2,\ldots,n\}$ have a descent, if we sum (5.4) over m we obtain the identity

$$\sum_{m=1}^{n-1}\sum_{k=m+1}^{n}\sum_{j=0}^{k-2}\binom{k-2}{j}(n-j-2)! = n!-1. \tag{5.5}$$

A direct proof is

$$\begin{aligned} C &:= \sum_{m=1}^{n-1}\sum_{k=m+1}^{n}\sum_{j=0}^{k-2}\binom{k-2}{j}(n-j-2)! \\ &= \sum_{k=2}^{n}\sum_{m=1}^{k-1}\sum_{j=0}^{k-2}\binom{k-2}{j}(n-j-2)! \\ &= \sum_{k=2}^{n}(k-1)\sum_{j=0}^{k-2}\binom{k-2}{j}(n-j-2)! = A. \end{aligned}$$

The average end height is then given by the expression

$$\frac{g(n)}{n!} := \frac{1}{n!}\sum_{m=1}^{n-1} m \sum_{k=m+1}^{n}\sum_{j=0}^{k-2}\binom{k-2}{j}(n-j-2)!\,. \tag{5.6}$$

Theorem 8 *We have*

$$\frac{g(n)}{n!} = \frac{h(n)}{2n!} \sim (e/2-1)(n+1), \tag{5.7}$$

as $n\to\infty$, *where* $e/2-1 = 0.3591409142\ldots$.

Proof: We have

$$\begin{aligned} D &:= \frac{1}{n!}\sum_{m=1}^{n-1} m \sum_{k=m+1}^{n}\sum_{j=0}^{k-2}\binom{k-2}{j}(n-j-2)! \\ &= \frac{1}{n!}\sum_{k=2}^{n}\sum_{m=1}^{k-1} m \sum_{j=0}^{k-2}\binom{k-2}{j}(n-j-2)! \\ &= \frac{1}{2n!}\sum_{k=2}^{n}k(k-1)\sum_{j=0}^{k-2}\binom{k-2}{j}(n-j-2)! = \frac{B}{2}. \end{aligned}$$

□

From Theorems 7 and 8 the mean size of the first descent is also asymptotic to $(e/2-1)(n+1)$ as $n \to \infty$.

The $g(n)$ sequence for $n \geq 1$ is

$$0, 1, 7, 41, 256, 1807, 14477, 130321, 1303246, 14335751, 172029067, 2236377937, \ldots$$

is $A080047$ in Sloane (16). It is described there as the operation count to create all permutations of n distinct elements using Algorithm L (lexicographic permutation generation) from Knuth (11, chapter 7.2.1.2). The term $g(n-1)$ for $n \geq 2$ gives the number of times l has to be repeatedly decreased in step L3. Knuth (10) also studies the lengths of the runs in permutations.

Sloane also has the recurrence $g(1) = 0$, $g(n) = (n+1)g(n-1) + n(n-1)/2$ for $n \geq 2$ and the formula $g(n) = \lfloor (e/2-1)(n+1)! - n/2 \rfloor$ for $n \geq 1$.

References

[1] L. Devroye. A limit theory for random skip lists. *Advances in Applied Probability*, 2:597–609, 1992.

[2] P. Flajolet and G. N. Martin. Probabilistic counting algorithms for data base applications. *Journal of Computer and System Sciences*, 31:182–209, 1985.

[3] P. Flajolet and A. M. Odlyzko. Singularity analysis of generating functions. *SIAM J. Discrete Math.*, 3:216-240, 1990.

[4] P. Kirschenhofer and H. Prodinger. On the analysis of probabilistic counting. In E. Hlawka and R. F. Tichy, editors, *Number–theoretic Analysis*, volume 1452 of *Lecture Notes in Mathematics*, pages 117–120, 1990.

[5] P. Kirschenhofer and H. Prodinger. A result in order statistics related to probabilistic counting. *Computing*, 51:15–27, 1993.

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

[7] P. Kirschenhofer, H. Prodinger, and W. Szpankowski. Analysis of a splitting process arising in probabilistic counting and other related algorithms. *Random Structures and Algorithms*, 9:379–401, 1996.

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

[9] A. Knopfmacher and H. Prodinger. Descents in samples of geometric random variables. Mathematics and Computer Science: Algorithms, Trees, Combinatorics and Probabilities, Drmota, Flajolet, Gardy and Gittenberger (Eds.), 339–350, 2004.

[10] D. E. Knuth. *The art of computer programming*, volume 3: Sorting and Searching. Addison-Wesley, 1973. Second edition, 1998.

[11] D. E. Knuth. *The art of computer programming*, volume 4: Chapter 7. Preliminary version available at *http://www-cs-faculty.stanford.edu/∼knuth/news.html*.

[12] T. Papadakis, I. Munro, and P. Poblete. Average search and update costs in skip lists. *BIT*, 32:316–332, 1992.

[13] H. Prodinger. Combinatorics of geometrically distributed random variables: Left-to-right maxima. *Discrete Mathematics*, 153:253–270, 1996.

[14] H. Prodinger. Combinatorics of geometrically distributed random variables: Length of ascending runs. Proceedings of LATIN'2000, *Lecture Notes in Comp. Sci.* 1776, Gonnet, Panario and Viola (Eds.), 473–482, 2000.

[15] W. Pugh. Skip lists: a probabilistic alternative to balanced trees. *Communications of the ACM*, 33:668–676, 1990.

[16] N. J. A. Sloane and S. Plouffe, *The encyclopaedia of integer sequences*, Academic Press, 1995. Online edition available at *http://www.research.att.com/∼njas/sequences/*.

6 Appendix

Proof of Proposition 1:

We will make use of Heine's formula (4.1). We wish to evaluate

$$X := \sum_{h\ge1} pzq^{2h} \prod_{j=1}^{h} \frac{1}{1-pq^jz} = pz\biggl(\sum_{n\ge0} q^{2n}\frac{1}{(pqz)_n} - 1\biggr).$$

Set

$$\xi := \sum_{n\ge0} q^{2n}\frac{1}{(pqz)_n}$$

and use Heine's formula with parameters $t=q^2$, $a=0$, $b=q$, $c=pqz$:

$$\begin{aligned}\xi &= \frac{(q)_\infty}{(pqz)_\infty(q^2)_\infty}\sum_{m\ge0}\frac{(pz)_m(q^2)_m}{(q)_m}q^m\\ &= \frac{1}{(pqz)_\infty}\sum_{m\ge0}\frac{(pz)_m(q)_{m+1}}{(q)_m}q^m\\ &= \frac{1}{(pqz)_\infty}\sum_{m\ge0}(pz)_m q^m - \frac{q}{(pqz)_\infty}\sum_{m\ge0}(pz)_m q^{2m}. \qquad (6.1)\end{aligned}$$

Now compute the first sum, again with Heine's formula: $t=q$, $a=q$, $b=pz$, $c=0$: (and then Euler's partition identity (2.4))

$$\begin{aligned}\sum_{m\ge0}(pz)_m q^m &= \frac{(pz)_\infty(q^2)_\infty}{(q)_\infty}\sum_{m\ge0}\frac{(q)_m}{(q)_m(q^2)_m}(pz)^m\\ &= \frac{(pz)_\infty}{pz}\biggl(\sum_{m\ge0}\frac{1}{(q)_m}(pz)^m - 1\biggr)\\ &= \frac{(pz)_\infty}{pz}\biggl(\frac{1}{(pz)_\infty} - 1\biggr)\\ &= \frac{1}{pz} - \frac{(pz)_\infty}{pz}. \qquad (6.2)\end{aligned}$$

Now compute the second sum, again with Heine's formula: $t = q^2$, $a = q$, $b = pz$, $c = 0$: (and then (2.4))

$$\begin{aligned}
\sum_{m\ge 0}(pz)_m q^{2m} &= \frac{(pz)_\infty (q^3)_\infty}{(q^2)_\infty}\sum_{m\ge 0}\frac{(q^2)_m}{(q)_m (q^3)_m}(pz)^m \\
&= (pz)_\infty \sum_{m\ge 0}\frac{(q^2)_m}{(q)_m (q^2)_{m+1}}(pz)^m \\
&= (pz)_\infty \sum_{m\ge 0}\frac{1-q^{m+1}}{(q)_{m+2}}(pz)^m \\
&= \frac{(pz)_\infty}{(pz)^2}\sum_{m\ge 2}\frac{(pz)^m}{(q)_m} - \frac{q(pz)_\infty}{(pqz)^2}\sum_{m\ge 2}\frac{(pqz)^m}{(q)_m} \\
&= \frac{(pz)_\infty}{(pz)^2}\sum_{m\ge 0}\frac{(pz)^m}{(q)_m} - \frac{(pz)_\infty}{(pz)^2} - \frac{(pz)_\infty}{p^2 z} - \frac{q(pz)_\infty}{(pqz)^2}\sum_{m\ge 0}\frac{(pqz)^m}{(q)_m} + \frac{q(pz)_\infty}{(pqz)^2} + \frac{(pz)_\infty}{p^2 z} \\
&= \frac{1}{(pz)^2} - \frac{(pz)_\infty}{(pz)^2} - \frac{q(1-pz)}{(pqz)^2} + \frac{q(pz)_\infty}{(pqz)^2}. \qquad (6.3)
\end{aligned}$$

Then using (6.1) together with (6.2) and (6.3),

$$\begin{aligned}
X &= pz\xi - pz \\
&= \left[\frac{1}{(pqz)_\infty}\Big(1-(pz)_\infty\Big) - \frac{q}{(pqz)_\infty}\Big(\frac{1}{pz} - \frac{(pz)_\infty}{pz} - \frac{q(1-pz)}{pq^2 z} + \frac{q(pz)_\infty}{pq^2 z}\Big)\right] - pz \\
&= \left[\frac{1}{(pqz)_\infty} - 1 + pz - \frac{q}{(pqz)_\infty}\frac{1}{pz} + \frac{q(1-pz)}{pz} + \frac{1-pz}{pz(pqz)_\infty} - \frac{(1-pz)}{pz}\right] - pz \\
&= -q + \frac{1}{z(pqz)_\infty} - \frac{1}{z}.
\end{aligned}$$

□

Proof of Proposition 2:

Compute

$$A := \sum_{n \geq 0} pzq^{2h+1}(-pz)_h = pqzB,$$

with

$$B := \sum_{n \geq 0} q^{2h}(-pz)_h.$$

Apply Heine's formula with $t = q^2$, $a = q$, $b = -pz$, $c = 0$:

$$\begin{aligned}
B &= \frac{(-pz)_\infty (q^3)_\infty}{(q^2)_\infty} \sum_{m \geq 0} \frac{(q^2)_m}{(q)_m (q^3)_m} (-pz)^m \\
&= (-pz)_\infty \sum_{m \geq 0} \frac{1 - q^{m+1}}{(q)_{m+2}} (-pz)^m \\
&= \frac{(-pz)_\infty}{p^2 z^2} \sum_{m \geq 2} \frac{1}{(q)_m} (-pz)^m - \frac{q(-pz)_\infty}{p^2 q^2 z^2} \sum_{m \geq 2} \frac{1}{(q)_m} (-pqz)^m \\
&= \frac{(-pz)_\infty}{p^2 z^2} \left[\frac{1}{(-pz)_\infty} - 1 + z \right] - \frac{(-pz)_\infty}{p^2 q z^2} \left[\frac{1}{(-pqz)_\infty} - 1 + qz \right] \\
&= \frac{(-pz)_\infty}{pqz^2} - \frac{1}{pqz^2} - \frac{1}{pqz}.
\end{aligned}$$

And finally

$$B = pqzA = \frac{(-pz)_\infty}{z} - \frac{1}{z} - 1.$$

$\square$