*Discrete Mathematics and Theoretical Computer Science* DMTCS vol. **8**, 2006, 273–284

# *$d$–records in geometrically distributed random variables*

Helmut Prodinger[1] †

[1] *Stellenbosch University, Department of Mathematics, 7602 Stellenbosch, South Africa.* hproding@sun.ac.za

*received Aug 28, 2006*, *revised Sep 4, 2006*, *accepted Sep 4, 2006*.

We study $d$–records in sequences generated by independent geometric random variables and derive explicit and asymptotic formulæ for expectation and variance. Informally speaking, a $d$–record occurs, when one computes the $d$–largest values, and the variable maintaining it changes its value while the sequence is scanned from left to right. This is done for the "strict model," but a "weak model" is also briefly investigated. We also discuss the limit $q \to 1$ ($q$ the parameter of the geometric distribution), which leads to the model of random permutations.

**Keywords:** Records, geometric random variables, asymptotics

## 1 Introduction

Records (left–to–right maxima) of a sequence of elements $x_1 \dots x_n$ are a well studied subject (Nev01): The sequence is read from left to right, and whenever an element is encountered which is larger than the previously seen ones, we speak of a record. The total number of them is of interest. This is also of interest in Computer Science, see (Knu73).

As we learn from (DLB05), the notation of $d$–records is not uniform in the literature (e.g., (Nev87)). What works best for us in this context can be seen from the following simple program which computes the $d$–largest element. We count how often the variable $C_d$ changes its value. This is what we will call *number of $d$–records* in this paper.

In this note, we assume that the elements $n \in \mathbb{N}$ are drawn independently from a geometric distribution: $\mathbb{P}\{X = k\} = pq^{k-1}$, with $p + q = 1$. This is a situation for which we can still derive attractive results, and consider the limit $q \to 1$ as well. This leads us then to the model of random permutations.

We compute expectation and variance of the parameter "number of $d$–records;" the question about the limiting distribution is left open here, although we strongly believe that it is gaussian. Since we do not have access to the probabilities or a probability generating function, this seems to be more difficult than the instance $d = 1$. This one appears first in (Pro96); the limiting distribution was studied by Hwang and coauthors in (BHL98).

The following example shows how the parameters develop, for $d = 3$.

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

**Algorithm 1** Computation of the $d$–largest element

**Input:** $x_1, \dots, x_n$ sequence.
**Output:** $C_d$, which is the $d$–record of the sequence.

```
C_1 ← −∞, ..., C_d ← −∞
for k = 1 to n do
   X ← x_k
   for i = 1 to d do
      if X > C_i then
         C_i ↔ X
      end if
   end for
end for
```

| | 2 | 1 | 2 | 5 | 4 | 8 | 3 | 6 | 2 | 4 |
|---|---|---|---|---|---|---|---|---|---|---|
| $C_1 = -\infty$ | 2 | 2 | 2 | 5 | 5 | 8 | 8 | 8 | 8 | 8 |
| $C_2 = -\infty$ | $-\infty$ | 1 | 2 | 2 | 4 | 5 | 5 | 6 | 6 | 6 |
| $C_3 = -\infty$ | $-\infty$ | $-\infty$ | 1 | 2 | 2 | 4 | 4 | 5 | 5 | 5 |

We also study the instance of weak $d$–records, i.e., we replace $X > C_i$ in the above algorithm by $X \ge C_i$. This is not very practical and leads to messy computations as well, so we only compute the average value here.

We could say that this paper analyzes the algorithm decribed before.

A few abbreviations are useful: $Q = 1/q$, $L = \log Q$.

**Remark.** The last value of the variable $C_d$ for the computation of $d$–records is the $d$–largest value. These were investigated in (KP93).

## 2 Expectation and variance

The random variable $X$ (number of $d$–records) can be written as $X = \chi_1 + \chi_2 + \cdots$, where $\chi_k$ is one if the variable $C_d$ (that represents the $d$–largest element) will eventually change to the value $k$, zero otherwise. The expectation $\mathbb{E}(\chi_k)$ is just the probability that this happens. To compute this, note that $d$ values must be $\ge k$, but not all of them $> k$. The other elements are smaller, and right from the element of interest, everything is allowed. In other words, we want to count the weight (probability) of all the words of length $n$ of the form $w_1 a_1 \dots w_d a_d y$, with letters $a_i \in \{k, k+1, \dots\}$ (not all of them larger than $k$) and words $w_i \in \{1, \dots, k-1\}^*$, $y \in \{1, 2, \dots\}^*$:

$$\begin{aligned}\mathbb{E}(X) &= [z^n] \sum_{k \ge 1} \Big( (q^{k-1} z)^d - (q^k z)^d \Big) \frac{1}{(1 - z(1 - q^{k-1}))^d} \frac{1}{1-z} \\ &= [z^n] \sum_{k \ge 0} (q^k z)^d (1 - q^d) \frac{1}{(1 - z(1 - q^k))^d} \frac{1}{1-z}.\end{aligned}$$

(We use the customary notation $[z^n] f(z)$ for the coefficient of $z^n$ in the power series $f(z)$.)

Now we use the substitution $z = w/(w-1)$. The tutorial (KP96) is a good reading for the type of analysis that is done here. The paper (KMP95) explains this substitution in more detail; of particular interest to us here is the formula

$$[z^n]f(z) = (-1)^n[w^n](1-w)^{n-1}f\big(w/(w-1)\big),$$

which can be proved for instance by formal residue calculus. Then

$$\begin{aligned}
\mathbb{E}(X) &= [w^n](-1)^{n+d}(1-q^d)(1-w)^n\sum_{k\ge 0}(q^kw)^d\frac{1}{(1-wq^k)^d}\\
&= (-1)^{n+d}(1-q^d)\sum_{j=0}^{n}[w^{n-j}](1-w)^n\cdot[w^j]\sum_{k\ge 0}(q^kw)^d\frac{1}{(1-wq^k)^d}\\
&= (-1)^d(1-q^d)\sum_{j=d}^{n}\binom{n}{j}(-1)^j[w^j]\sum_{k\ge 0}(q^kw)^d\frac{1}{(1-wq^k)^d}\\
&= (-1)^d(1-q^d)\sum_{j=d}^{n}\binom{n}{j}(-1)^j[w^j]\frac{1}{1-q^j}w^d\frac{1}{(1-w)^d}\\
&= (-1)^d(1-q^d)\sum_{j=d}^{n}\binom{n}{j}(-1)^j\frac{1}{1-q^j}\binom{j-1}{d-1}\\
&= (-1)^d(1-q^d)\frac{1}{2\pi i}\int_{\mathcal{C}}\frac{(-1)^n n!}{z(z-1)\dots(z-n)}\frac{(z-1)\dots(z-d+1)}{(d-1)!}\frac{1}{1-q^z}dz.
\end{aligned}$$

The last step (writing the alternating sum as a contour integral with a contour encircling the points $1,\dots,n$) is prominent in Rice's method, see (FS95).

The asymptotic evaluation now proceeds by looking at the residues outside of this curve, taking them with a negative sign. Here, the main contributions come from $z=0$:

$$\begin{aligned}
(1-q^d)[z^{-1}]\frac{1}{(1-\frac{z}{d})\dots(1-\frac{z}{n})}\frac{1}{1-q^z} &\sim (1-q^d)[z^{-1}]\Big(1+z(H_n-H_{d-1})\Big)\frac{1}{Lz}\Big(1+\frac{Lz}{2}\Big)\\
&\sim (1-q^d)\frac{1}{L}\Big(H_n-H_{d-1}+\frac{L}{2}\Big)\\
&\sim (1-q^d)\Big(\log_Q n+\frac{\gamma}{L}-\frac{H_{d-1}}{L}+\frac{1}{2}\Big).
\end{aligned}$$

The numbers $H_m := 1+\frac{1}{2}+\cdots+\frac{1}{m}$ that appear here are called *harmonic numbers*; the numbers $H_m^{(2)} := 1+\frac{1}{2^2}+\cdots+\frac{1}{m^2}$ will appear later. For $d=1$ we find the old value (cf. (Pro96))

$$p\Big(\log_Q n+\frac{\gamma}{L}+\frac{1}{2}\Big).$$

And now, for the variance, let us compute $\mathbb{E}(\chi_k\chi_l)$ for $k<l$, which is the probability that the variable $C_d$ changes to $k$, and (later) to $l$. So, there must be $1\le i\le d$ values between $k$ and $l-1$, and not all of them $>k$, and the remaining $d-k$ are $\ge l$. Together with $i$ other values, which are also $\ge l$, we have $d$

values $\ge l$; not all of them can be $> l$. So

$$\begin{aligned}
\sum_{1\le k<l}\mathbb{E}(\chi_k\chi_l) &= [z^n]\sum_{1\le k<l}\sum_{i=1}^{d}\binom{d}{i}\Big(((q^{k-1}-q^{l-1})z)^i-((q^{k}-q^{l-1})z)^i\Big)\frac{1}{(1-z(1-q^{k-1}))^d}\\
&\qquad\times\big((q^{l-1}z)^d-(q^lz)^d\big)\frac{1}{(1-z(1-q^{l-1}))^i}\frac{1}{1-z}\\
&=(1-q^d)[z^n]\sum_{0\le k<l}\sum_{i=1}^{d}\binom{d}{i}\Big(((q^{k}-q^{l})z)^i-((q^{k+1}-q^{l})z)^i\Big)\frac{1}{(1-z(1-q^{k}))^d}\\
&\qquad\times(q^lz)^d\frac{1}{(1-z(1-q^{l}))^i}\frac{1}{1-z}\\
&=(1-q^d)[z^n]\sum_{0\le k<l}\sum_{i=1}^{d}\binom{d}{i}((q^k)z)^i\Big((1-q^{l-k})^i-(q-q^{l-k})^i\Big)\frac{1}{(1-z(1-q^{k}))^d}\\
&\qquad\times(q^lz)^d\frac{1}{(1-z(1-q^{l}))^i}\frac{1}{1-z}\\
&=(1-q^d)[w^n](1-w)^n\sum_{0\le k,\,l\ge1}\sum_{i=1}^{d}\binom{d}{i}(-1)^{n+d+i}\Big((1-q^l)^i-(q-q^l)^i\Big)\\
&\qquad\times\frac{((q^k)w)^i}{(1-wq^k)^d}\frac{(q^{k+l}w)^d}{(1-wq^{k+l})^i}\\
&=(1-q^d)[w^n](1-w)^n\sum_{k\ge0,\,l\ge1}\sum_{i=1}^{d}(-1)^{n+d+i}\binom{d}{i}\sum_{\lambda=0}^{i}\binom{i}{\lambda}(-1)^\lambda q^{\lambda l}(1-q^{i-\lambda})\\
&\qquad\times\frac{((q^k)w)^i}{(1-wq^k)^d}\frac{(q^{k+l}w)^d}{(1-wq^{k+l})^i}\\
&=(1-q^d)\sum_{i=1}^{d}\binom{d}{i}(-1)^{d+i}\sum_{\lambda=0}^{i}\binom{i}{\lambda}(-1)^\lambda(1-q^{i-\lambda})\sum_{j=i+d}^{n}\binom{n}{j}(-1)^j\frac{1}{1-q^j}\\
&\qquad\times[w^j]\sum_{l\ge1}q^{\lambda l}\frac{w^i}{(1-w)^d}\frac{(q^lw)^d}{(1-wq^l)^i}\\
&=(1-q^d)\sum_{i=1}^{d}\binom{d}{i}(-1)^{d+i}\sum_{\lambda=0}^{i}\binom{i}{\lambda}(-1)^\lambda(1-q^{i-\lambda})\sum_{j=i+d}^{n}\binom{n}{j}(-1)^j\frac{1}{1-q^j}\\
&\qquad\times\sum_{m=d+\lambda}^{j+\lambda-i}\sum_{l\ge1}[w^{j+\lambda-m}]\frac{w^i}{(1-w)^d}\cdot[w^m]\frac{(q^lw)^{d+\lambda}}{(1-wq^l)^i}
\end{aligned}$$

$$= (1-q^d)\sum_{i=1}^{d}\binom{d}{i}(-1)^{d+i}\sum_{\lambda=0}^{i}\binom{i}{\lambda}(-1)^{\lambda}(1-q^{i-\lambda})\sum_{j=i+d}^{n}\binom{n}{j}(-1)^j\frac{1}{1-q^j}$$
$$\times\sum_{m=d+\lambda}^{j+\lambda-i}\frac{1}{Q^m-1}\binom{j+\lambda-m-i+d-1}{d-1}\binom{m-d-\lambda+i-1}{i-1}$$
$$= (1-q^d)\sum_{i=1}^{d}\binom{d}{i}(-1)^{d+i}\sum_{\lambda=0}^{i}\binom{i}{\lambda}(-1)^{\lambda}(1-q^{i-\lambda})\sum_{j=i+d}^{n}\binom{n}{j}(-1)^j\frac{1}{1-q^j}$$
$$\times\sum_{m=0}^{j-d-i}\frac{1}{Q^{m+d+\lambda}-1}\binom{j-m-i-1}{d-1}\binom{m+i-1}{i-1}.$$

This is explicit, but we need the asymptotic behaviour of it, again with Rice's method.

Define

$$\psi(j)=\sum_{m=0}^{j-d-i}\frac{1}{Q^{m+d+\lambda}-1}\binom{j-m-i-1}{d-1}\binom{m+i-1}{i-1}$$
$$=\sum_{m\geq 0}\frac{1}{Q^{m+d+\lambda}-1}\binom{j-m-i-1}{d-1}\binom{m+i-1}{i-1}$$
$$-\sum_{m+d-j+i-1\geq 0}\frac{1}{Q^{m+d+\lambda}-1}\binom{j-m-i-1}{d-1}\binom{m+i-1}{i-1}$$
$$=\sum_{m\geq 0}\frac{1}{Q^{m+d+\lambda}-1}\binom{j-m-i-1}{d-1}\binom{m+i-1}{i-1}$$
$$-\sum_{m\geq 0}\frac{1}{Q^{m+j-i+\lambda+1}-1}\binom{-m+d-2}{d-1}\binom{m-d+j}{i-1}.$$

So we can rewrite $\psi(z)$ as follows:

$$\psi(z)=\sum_{m\geq 0}\frac{1}{Q^{m+d+\lambda}-1}\binom{z-m-i-1}{d-1}\binom{m+i-1}{i-1}$$
$$-\sum_{m\geq 0}\frac{1}{Q^{m+z-i+\lambda+1}-1}\binom{-m+d-2}{d-1}\binom{m-d+z}{i-1}.$$

But this form also makes sense for complex values of $z$, whence you use it as the extended definition of $\psi(z)$. Continuing, we can write

$$\sum_{1\leq k<l}\mathbb{E}(\chi_k\chi_l)=(1-q^d)\sum_{i=1}^{d}\binom{d}{i}(-1)^{d+i}\sum_{\lambda=0}^{i}\binom{i}{\lambda}(-1)^{\lambda}(1-q^{i-\lambda})$$
$$\times\frac{1}{2\pi i}\int_{\mathcal{C}}\frac{(-1)^n n!}{z(z-1)\dots(z-n)}\frac{1}{1-q^z}\psi(z)dz.$$

When shifting the contour of integration (and thus computing residues) it is essential that the integrand is small for large imaginary parts. See (FS95) for background information. Since the $\psi(z)$ function is defined in terms of Gamma functions, this smallness is inherited from the well known smallness of $|\Gamma(c+it)|$, when $|t|$ gets large.

There is a triple pole at $z=0$, and the computation of the residue (with negative sign) is extremely tedious. We spare the reader the details.

$$\begin{aligned}\Xi &= -(1-q^d)\sum_{i=1}^{d}\binom{d}{i}(-1)^{d+i}\sum_{\lambda=0}^{i}\binom{i}{\lambda}(-1)^{\lambda}(1-q^{i-\lambda})\\ &\qquad\times[z^0]\Big(1+zH_n+z^2\frac{H_n^2+H_n^{(2)}}{2}\Big)\frac{1}{Lz}\Big(1+z\frac{L}{2}+z^2\frac{L^2}{12}\Big)\psi(z)\\ &=(1-q^d)^2\frac{H_n^2+H_n^{(2)}}{2}-(1-q^d)^2\frac{H_{d-1}-\gamma}{L^2}H_n\\ &\qquad-(1-q^d)\sum_{i=1}^{d}\binom{d}{i}(-1)^{d+i}\sum_{\lambda=0}^{i}\binom{i}{\lambda}(-1)^{\lambda}(1-q^{i-\lambda})\frac{1}{L}[z^1]\psi(z),\end{aligned}$$

and

$$\begin{aligned}[z^1]\psi(z) &= \sum_{m\ge 0}\frac{1}{Q^{m+d+\lambda}-1}\binom{m+i-1}{i-1}[z^1]\binom{z-m-i-1}{d-1}\\ &\quad-\sum_{m\ge d-1}\frac{1}{Q^{m-i+\lambda+1}-1}\binom{-m+d-2}{d-1}[z^1]\binom{m-d+z}{i-1}\\ &\quad+L\sum_{m\ge d-1}\frac{Q^{m-i+\lambda+1}}{(Q^{m-i+\lambda+1}-1)^2}\binom{-m+d-2}{d-1}\binom{m-d}{i-1}\\ &=\sum_{m\ge 0}\frac{1}{Q^{m+d+\lambda}-1}\binom{m+i-1}{i-1}\binom{-m-i-1}{d-1}(H_{m+d-1}-H_m)\\ &\quad-\sum_{m\ge d+i-1}\frac{1}{Q^{m-i+\lambda+1}-1}\binom{-m+d-2}{d-1}\binom{m-d}{i-1}(H_{m-d}-H_{m-d-i+1})\\ &\quad-\frac{1}{Q^{d-i+\lambda}-1}(-1)^{d-1+i}H_i\\ &\quad-\sum_{d\le m\le d+i-2}\frac{1}{Q^{m-i+\lambda+1}-1}\binom{-m+d-2}{d-1}\frac{(m-d)!(d-m+i-2)!(-1)^{d-m+i}}{(i-1)!}\\ &\quad+L\sum_{m\ge d-1}\frac{Q^{m-i+\lambda+1}}{(Q^{m-i+\lambda+1}-1)^2}\binom{-m+d-2}{d-1}\binom{m-d}{i-1}.\end{aligned}$$

In this sum, the term for $i=d$, $\lambda=0$, $m=d-1$ must be replaced by

$$\frac{H_{d-1}}{L}+\frac{1}{2}.$$

For the final computation of the variance, we must take $\Xi$ twice (because of the symmetry $k < l$ resp. $k > l$), add the diagonal terms $k = l$, which amounts to the expectation, and subtract the square of the expectation. Apart from the ungainly constant, that we no longer mention explicitly, we find that the contribution of $\log^2 n$ cancels out, and the term $\log_Q n$ has as a factor

$$-2(1-q^d)^2\frac{H_{d-1}-\gamma}{L}+(1-q^d)-(1-q^d)^2 2\Big(\frac{\gamma}{L}-\frac{H_{d-1}}{L}+\frac{1}{2}\Big)=(1-q^d)-(1-q^d)^2=q^d(1-q^d).$$

For $d = 1$, this reduces to $pq$, as it should (compare (Pro96)).

In all these problems, there are also poles at $2\pi ik/L$ for $k \in \mathbb{Z}$, $k \neq 0$. They contribute small periodic functions ("fluctuations"). We refrain from computing them explicitly, as they lead to *very* unpleasant terms, in the style of the previous constant.

Summarizing, we sketched the proof of the following theorem.

**Theorem 1** *The parameter "number of changes of variable $C_d$" (= number of $d$–records), for random strings of length $n$, produced by independent geometric random variables, has the following asymptotic equivalents for $n \to \infty$ for expectation resp. variance:*

$$\textit{Expectation} \sim (1-q^d)\Big(\log_Q n+\frac{\gamma}{L}-\frac{H_{d-1}}{L}+\frac{1}{2}\Big)+\delta_E(\log_Q n),$$
$$\textit{Variance} \sim q^d(1-q^d)\log_Q n+\textit{constant}+\delta_V(\log_Q n).$$

*The (small) periodic functions $\delta_E(x)$, $\delta_V(x)$ could be determined in principle in terms of their Fourier coefficients. The term labelled "constant" could be collected from the computations sketched above.*

# 3 The weak model: expectation

Now we think about a new element bubbling down as before. If the one that is to be compared with the variable for the $d$–record is equal, we *also* count that as a change (compare the Introduction). Note that if $-\infty$ is replaced by $-\infty$ (at the beginning of the sequence), we do not count this.

So $m \geq d$ elements must have been $\geq k$, but at most $d-1$ of them $> k$:

$$\begin{aligned}
\mathbb{E}(X) &= [z^n]\sum_{k\geq 1}\sum_{m\geq d}\sum_{\lambda=0}^{d-1}\binom{m}{\lambda}(zq^k)^\lambda(zpq^{k-1})^{m-\lambda}\frac{1}{(1-z(1-q^{k-1}))^m}\frac{1}{1-z}\\
&=\sum_{j=d}^{n}\binom{n}{j}(-1)^j[w^j]\sum_{m\geq d}(-1)^m\sum_{\lambda=0}^{d-1}\binom{m}{\lambda}\sum_{k\geq 0}(wq^{k+1})^\lambda(wpq^k)^{m-\lambda}\frac{1}{(1-wq^k)^m}\\
&=\sum_{j=d}^{n}\binom{n}{j}(-1)^j[w^j]\sum_{m\geq d}(-1)^m\sum_{\lambda=0}^{d-1}\binom{m}{\lambda}\frac{1}{1-q^j}(wq)^\lambda(wp)^{m-\lambda}\frac{1}{(1-w)^m}\\
&=\sum_{j=d}^{n}\binom{n}{j}(-1)^j\frac{1}{1-q^j}\underbrace{[w^j]\sum_{m\geq d}\sum_{\lambda=0}^{d-1}\binom{m}{\lambda}\Big(\frac{q}{p}\Big)^\lambda\Big(\frac{pw}{w-1}\Big)^m}_{\psi(j)}.
\end{aligned}$$

In order to evaluate $\psi(j)$, let us do a little calculation:

$$\begin{aligned}\Xi &= \sum_{0\le\lambda<d}\sum_{m\ge d}\binom{m}{\lambda}a^\lambda b^m\\ &= \sum_{0\le\lambda<d}a^\lambda\biggl(\frac{b^\lambda}{(1-b)^{\lambda+1}}-\sum_{m<d}\binom{m}{\lambda}b^m\biggr)\\ &= \frac{1}{1-b}\frac{1-\bigl(\frac{ab}{1-b}\bigr)^d}{1-\frac{ab}{1-b}}-\underbrace{\sum_{0\le\lambda\le m<d}\binom{m}{\lambda}a^\lambda b^m}_{T_d}.\end{aligned}$$

Then $T_0=0$ and

$$T_{d+1}=T_d+\sum_{0\le\lambda\le d}\binom{d}{\lambda}a^\lambda b^d=T_d+(1+a)^d b^d,$$

so

$$T_d=\sum_{0\le h<d}\bigl((1+a)b\bigr)^h=\frac{1-((1+a)b)^d}{1-(1+a)b}.$$

Therefore

$$\Xi=\frac{1}{1-b}\frac{1-\bigl(\frac{ab}{1-b}\bigr)^d}{1-\frac{ab}{1-b}}-\frac{1-((1+a)b)^d}{1-(1+a)b}=\frac{((1+a)b)^d-\bigl(\frac{ab}{1-b}\bigr)^d}{1-(1+a)b}.$$

Let us use this with $a=\frac{q}{p}$ and $b=\frac{pw}{w-1}$. Then $1+a=\frac{1}{p}$, $(1+a)b=\frac{w}{w-1}$, $1-b=\frac{w-1-pw}{w-1}=\frac{1-qw}{1-w}$, and $\frac{ab}{1-b}=\frac{qw}{qw-1}$:

$$\begin{aligned}\psi(j) &= [w^j]\frac{(\frac{w}{w-1})^d-\bigl(\frac{qw}{qw-1}\bigr)^d}{1-\frac{w}{w-1}}\\ &= (-1)^d[w^j](1-w)\Bigl(\Bigl(\frac{w}{1-w}\Bigr)^d-\Bigl(\frac{qw}{1-qw}\Bigr)^d\Bigr)\\ &= (-1)^d\biggl(\binom{j-1}{d-1}-\binom{j-2}{d-1}-q^j\binom{j-1}{d-1}+q^{j-1}\binom{j-2}{d-1}\biggr)\\ &= (-1)^d\biggl(\binom{j-1}{d-1}(1-q^j)-\binom{j-2}{d-1}(1-q^{j-1})\biggr).\end{aligned}$$

So

$$\mathbb{E}(X)=\sum_{j=d}^{n}\binom{n}{j}(-1)^j\frac{1}{1-q^j}(-1)^d\biggl(\binom{j-1}{d-1}(1-q^j)-\binom{j-2}{d-1}(1-q^{j-1})\biggr)$$

$$= (-1)^d \sum_{j=d}^{n} \binom{n}{j}(-1)^j \binom{j-1}{d-1} - \sum_{j=d}^{n} \binom{n}{j}(-1)^j \frac{1}{1-q^j}\binom{j-2}{d-1}(1-q^{j-1})$$
$$= 1-(-1)^d \sum_{j=d}^{n} \binom{n}{j}(-1)^j \frac{1-q^{j-1}}{1-q^j}\binom{j-2}{d-1}$$
$$= 1-(-1)^d \frac{1}{2\pi i}\int_{\mathcal{C}} \frac{(-1)^n n!}{z(z-1)\dots(z-n)}\frac{1-q^{z-1}}{1-q^z}\binom{z-2}{d-1}$$
$$\sim 1-(-1)^d [z^{-1}]\frac{n!\Gamma(-z)}{\Gamma(n+1-z)}\frac{1-q^{z-1}}{1-q^z}\binom{z-2}{d-1}$$
$$\sim 1+d\frac{p}{q}\log_Q n - \frac{1}{L}\Big((H_{d-1}-\gamma)d\frac{p}{q}+\frac{p}{q}-d\frac{p}{q}\Big)-\frac{d}{2}\Big(1+\frac{1}{q}\Big).$$

That matches for $d=1$ with the old result (Pro96).

**Theorem 2** *The parameter "number of changes of variable $C_d$" (= number of $d$–records, weak model) for random strings of length $n$, produced by independent geometric random variables, has the following asymptotic equivalent for $n \to \infty$ for its expectation:*

$$\textit{Expectation} \sim d\frac{p}{q}\log_Q n - \frac{1}{L}\Big((H_{d-1}-\gamma)d\frac{p}{q}+\frac{p}{q}-d\frac{p}{q}\Big)-\frac{d}{2}\Big(1+\frac{1}{q}\Big)+1+\delta_{EW}(\log_Q n).$$

# 4 The permutation model

This model is obtained by taking the limit $q \to 1$. Doing this for

$$\mathbb{E}(X) = (-1)^d(1-q^d)\sum_{j=d}^{n}\binom{n}{j}(-1)^j\frac{1}{1-q^j}\binom{j-1}{d-1}$$

(strict model), we obtain

$$\mathbb{E}(X) = (-1)^d d\sum_{j=d}^{n}\binom{n}{j}(-1)^j\frac{1}{j}\binom{j-1}{d-1}.$$

Doing this for

$$\mathbb{E}(X) = 1-(-1)^d\sum_{j=d}^{n}\binom{n}{j}(-1)^j\frac{1-q^{j-1}}{1-q^j}\binom{j-2}{d-1}$$

(weak model), we obtain

$$\mathbb{E}(X) = 1-(-1)^d\sum_{j=d}^{n}\binom{n}{j}(-1)^j\frac{j-1}{j}\binom{j-2}{d-1}.$$

The two expressions coincide, which can be directly seen by the following computation:

$$\begin{aligned}
&(-1)^d d\sum_{j=d}^{n}\binom{n}{j}(-1)^j\frac{1}{j}\binom{j-1}{d-1}-1+(-1)^d\sum_{j=d}^{n}\binom{n}{j}(-1)^j\frac{j-1}{j}\binom{j-2}{d-1}\\
&=(-1)^d\sum_{j=d}^{n}\binom{n}{j}(-1)^j\binom{j-1}{d-1}-1\\
&=(-1)^d n\binom{n-1}{d-1}\sum_{j=d}^{n}\binom{n-d}{j-d}(-1)^j\frac{1}{j}-1\\
&=n\binom{n-1}{d-1}\sum_{j=0}^{n-j}\binom{n-d}{j}(-1)^j\frac{1}{j+d}-1\\
&=n\binom{n-1}{d-1}\frac{1}{d\binom{n}{d}}-1=0.
\end{aligned}$$

The permutation model itself is very easy, and one derives immediately the probability generating function

$$\prod_{k=d}^{n}\Big(\frac{k-d}{k}+\frac{xd}{k}\Big),$$

from which the expectation comes out as $d(H_n-H_{d-1})$ and the variance as $d(H_n-H_{d-1})-d^2(H_n^{(2)}-H_{d-1}^{(2)})$. From this we derive the formula

$$(-1)^d\sum_{j=d}^{n}\binom{n}{j}(-1)^j\frac{1}{j}\binom{j-1}{d-1}=H_n-H_{d-1}.$$

This is not too hard to do directly and we leave it as an exercise. However, the analogous limit concerning the variance is

$$\begin{aligned}
d\sum_{i=1}^{d}\binom{d}{i}(-1)^{d+i}\sum_{\lambda=0}^{i}\binom{i}{\lambda}(-1)^{\lambda}(i-\lambda)\sum_{j=i+d}^{n}\binom{n}{j}(-1)^j\frac{1}{j}\times&\\
\times\sum_{m=0}^{j-d-i}\frac{1}{m+d+\lambda}\binom{j-m-i-1}{d-1}\binom{m+i-1}{i-1}&\\
=\frac{d^2}{2}\Big((H_n-H_{d-1})^2-(H_n^{(2)}-H_{d-1}^{(2)})\Big).&
\end{aligned}$$

A direct proof of this is probably messy, but well within the reach of Carsten Schneider's impressive software Sigma (Sch01).

## References


[BHL98] Z.-D. Bai, H.-K. Hwang, and W.-Q. Liang. Normal approximations of the number of records in geometrically distributed random variables. *Random Structures and Algorithms*, 13:319–334, 1998.

[DLB05] A. Dembińska and F. López-Blázquez. $k$th records from discrete distributions. *Statistics and probability letters*, 71:203–214, 2005.

[FS95] P. Flajolet and R. Sedgewick. Mellin transforms and asymptotics: Finite differences and Rice's integrals. *Theoretical Computer Science*, 144:101–124, 1995.

[KMP95] P. Kirschenhofer, C. Martínez, and H. Prodinger. Analysis of an optimized search algorithm for skip lists. *Theoretical Computer Science*, 144:199–220, 1995.

[Knu73] D. E. Knuth. *The Art of Computer Programming*, volume 1: Fundamental Algorithms. Addison-Wesley, 1973. Third edition, 1997.

[KP93] P. Kirschenhofer and H. Prodinger. A result in order statistics related to probabilistic counting. *Computing*, 51:15–27, 1993.

[KP96] P. Kirschenhofer and H. Prodinger. The number of winners a in discrete geometrically distributed sample. *Annals in Applied Probability*, 6:687–694, 1996.

[Nev87] V. Nevzorov. Distributions of $k$th records in the discrete case. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)*, 158:133–137, 172, 1987.

[Nev01] V. Nevzorov. *Records: mathematical theory*, volume 194 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 2001. Translated from the Russian manuscript by D. M. Chibisov.

[Pro96] H. Prodinger. Combinatorics of geometrically distributed random variables: Left-to-right maxima. *Discrete Mathematics*, 153:253–270, 1996.

[Sch01] C. Schneider. *Symbolic Summation in Difference Fields*. PhD thesis, RISC, J Kepler University, Linz, 2001.