

Analysis of a new skip list variant

Guy Louchard¹ and Helmut Prodinger² †

¹ *Université Libre de Bruxelles, Département d'Informatique, CP 212, Boulevard du Triomphe, B-1050 Bruxelles, Belgium.* louchard@ulb.ac.be

² *Stellenbosch University, Department of Mathematics, 7602 Stellenbosch, South Africa.* hprodinger@sun.ac.za

For a skip list variant, introduced by Cho and Sahni, we analyse what is the analogue of horizontal plus vertical search cost in the original skip list model. While the average in Pugh's original version behaves like $Q \log_Q n$, with $Q = \frac{1}{q}$ a parameter, it is here given by $(Q + 1) \log_Q n$.

Keywords: Skip list, Rice's method, moments, functional equation, asymptotic expansion.

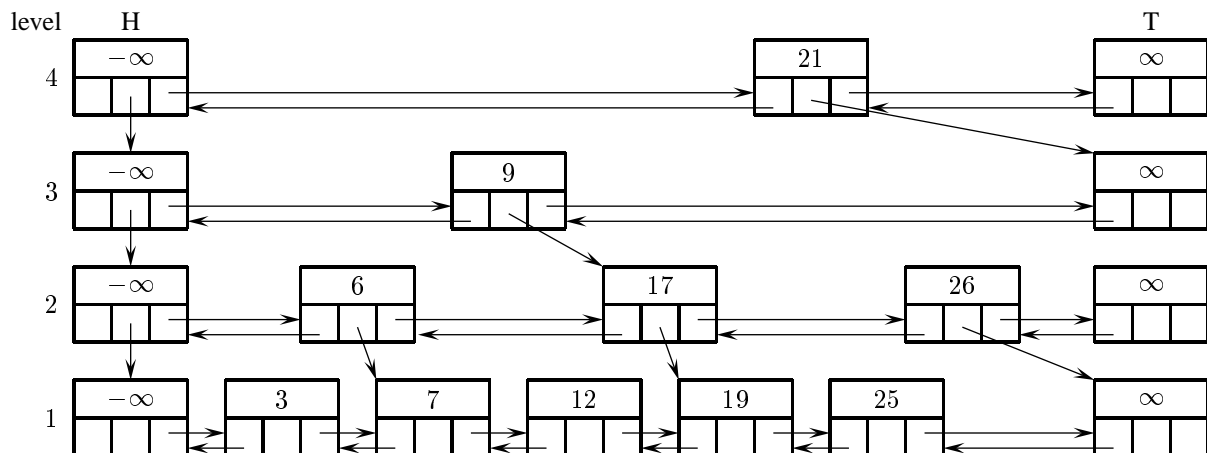
1 Introduction

Skip lists were introduced by Pugh (10) and analysed in (7; 4; 9) and also some other papers.

We assume that the reader has a certain familiarity with *skip lists*, if (s)he wants to understand the origin of the problem. To understand the analysis that we perform, no such prerequisites are necessary.

The variation that was suggested in (1) is best understood looking at the following example, taken from (1):

The data (3, 6, 7, 9, 12, 17, 19, 21, 25, 26) have a certain level associated to them, which follows the geometric law $\mathbb{P}\{\text{level is } = k\} = pq^{k-1}$. And they are linked as indicated, which one can easily understand from the diagram.



We want to study the length of a path to reach a certain element. For instance, to reach 25, we follow the path 9–17–19–25, and we record 4 steps. A step from the header to the data, as well as steps downwards between header elements are not counted.

The values of the data are completely immaterial; only the levels are of relevance. So, in our example the sequence is 1213121412. We start from the highest level that still allows us to reach the desired element, (here 3), and stay there as long as possible. Otherwise, we go down one level.

The process is easier understood, if we think about the sequence *reversed*, (here 2141213121).

The path of interest starts then at the element to be found (here the second element of sequence), which has level 1, and we scan the sequence, counting elements on the same level, until we find an element with the next level (one level higher), etc. We thus “visit” the elements marked in boldface: 2141**2**1**3**121, and we have 4 marked elements, which checks with the length of our search path.

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

Since elements to the left of where we start are irrelevant, we assume that we start with the first element. We will study the parameter $K(a_1 a_2 \dots a_n)$, which we might call the number of *weak consecutive maxima*, as we count repetitions of the (current) maximum, and only allow the maximum to change to the next value ($=1 +$ the previous value).

We did not care about implementations of this data structure. It is quite likely that if one does this, then some parameters that are similar to our K , but slightly different, play a rôle. We leave such indoubtedly interesting variations to the interested reader.

For our probabilistic analysis, we assume that the levels a_i are independently generated by the geometric law with parameter q (with $p = 1 - q$).

We have two parameters, the level I that has been already reached, and the counter K , that counts how often the current maximum I has been either repeated or replaced by $I + 1$.

There is a small technical subtlety: Sometimes it is useful to assume that we start at level r , and with $K = 1$, *before* we start to read the word. We will call this version $K^{(r)}(n)$. For the skip list analysis, we assume that the first symbol read defines the starting value; this version will be called $\mathcal{K}(n)$. Of course, they are intimately linked, and in a slight abuse of notation, we can say that

$$\mathcal{K}(n) = \sum_{r \geq 1} p q^{r-1} K^{(r)}(n-1).$$

Being precise, this informal equality reads

$$\mathbb{P}\{\mathcal{K}(n) = k\} = \sum_{r \geq 1} p q^{r-1} \mathbb{P}\{K^{(r)}(n-1) = k\}.$$

For the sake of clarity, we give the list of values for the word 13112435351:

$K^{(1)}$	$I^{(1)}$	$K^{(2)}$	$I^{(2)}$	$K^{(3)}$	$I^{(3)}$	$K^{(4)}$	$I^{(4)}$	$K^{(5)}$	$I^{(5)}$
7	3	5	5	5	5	4	5	3	5

We use (standard) notation from q -analysis: $(x)_n = \prod_{i=0}^{n-1} (1 - xq^i)$ and $(x)_\infty = \prod_{i \geq 0} (1 - xq^i)$. Note that $(x)_n = (x)_\infty / (xq^n)_\infty$, and the latter form makes sense also for n a complex number.

Furthermore, we use $Q = 1/q$, and $L = \ln Q$.

2 Generating functions

We consider the random variables $K^{(r)}(n)$ and $I^{(r)}(n)$. Let

$$\pi(n; k, i) = \mathbb{P}\{K^{(r)}(n) = k, I^{(r)}(n) = i\}.$$

(We don't write the parameter r into this notation, in order not to overload it.)

As a warm-up, we start at level $r = 1$. We have the backwards equation

$$\pi(n; k, i) = \pi(n-1; k-1, i-1) p q^{i-1} + \pi(n-1; k-1, i) p q^{i-1} + \pi(n-1; k, i) (1 - p q^{i-1} - p q^i)$$

and $\pi(0; 1, 1) = 1$.

We want to translate this into a trivariate generating function

$$F(z, u, v) = \sum_{n, k, i \geq 0} \pi(n; k, i) z^n u^k v^i.$$

Multiplying the backwards equation by $z^n u^k v^i$ we eventually get after a few routine simplifications (note that $F(0, u, v) = uv$):

$$F(z, u, v) = \frac{1}{1-z} \left[uv + pz \left(uv + \frac{1}{q} (u - q - 1) F(z, u, qv) \right) \right].$$

Iterating, this gives

$$F(z, u, v) = \sum_{j \geq 0} \frac{(pz)^j}{(1-z)^{j+1}} uv q^j \prod_{l=0}^{j-1} \left(uv q^l + \frac{u - q - 1}{q} \right).$$

Setting $u = 1$ means ignoring the K -parameter and only counting the level. The corresponding generating functions can be found in our recent paper (8).

However, we rather ignore the I -parameter here, which means that we set $v = 1$, to get

$$G(z, u) = F(z, u, 1) = \sum_{j \geq 0} \frac{(pz)^j}{(1-z)^{j+1}} uq^j \prod_{l=0}^{j-1} \left(uq^l + \frac{u-q-1}{q} \right).$$

The modifications for the starting level r are only minor: $F(0, u, v) = uv^r$, otherwise the same functional equation. Iteration produces the explicit form

$$F(z, u, v) = F^{(r)}(z, u, v) = \sum_{j \geq 0} \frac{(pz)^j}{(1-z)^{j+1}} u(vq^j)^r \prod_{l=0}^{j-1} \left(uvq^l + \frac{u-q-1}{q} \right).$$

The generating function that is relevant to the skip list, is

$$\begin{aligned} \mathcal{F}(z, u, v) &= z \sum_{r \geq 1} pq^{r-1} F^{(r)}(z, u, v) \\ &= \frac{u}{q} \sum_{j \geq 1} \frac{(pz)^j}{(1-z)^j} \frac{q^j v}{1-q^j v} \prod_{l=0}^{j-2} \left(uvq^l + \frac{u-q-1}{q} \right), \end{aligned} \tag{2.1}$$

and since the I -parameter is not relevant,

$$\mathcal{G}(z, u) = \mathcal{F}(z, u, 1) = \frac{u}{q} \sum_{j \geq 1} \frac{(pz)^j}{(1-z)^j} \frac{q^j}{1-q^j} \prod_{l=0}^{j-2} \left(uq^l + \frac{u-q-1}{q} \right).$$

Note that

$$[z^{n+1}] \mathcal{G}(z, u) = \frac{u}{q} \sum_{j=0}^n (pq)^{j+1} \binom{n}{j} \frac{1}{1-q^{j+1}} \prod_{l=0}^{j-1} \left(uq^l + \frac{u-q-1}{q} \right).$$

There is another way to get this generating function, which is more *combinatorial*. There is a unique decomposition of words:

$$\left(r(\mathbb{N} \setminus \{r, (r+1)\})^* \right)^+ \left((r+1)(\mathbb{N} \setminus \{(r+1), (r+2)\})^* \right)^+ \dots \left(s(\mathbb{N} \setminus \{s, s+1\})^* \right)^+.$$

This expresses the fact that the level starts at r and ends at s , which must be summed over all possible choices. If one translates this symbolic form accordingly, one gets

$$\bar{\mathcal{G}}(z, u) = \sum_{1 \leq r \leq s} \prod_{i=r}^s \frac{zupq^{i-1}}{1-z+zipq^{i-1}(1+q-u)}.$$

We will give a direct proof in the sequel that the two forms of $\mathcal{G}(z, u)$ coincide, which is surprisingly difficult. In order to avoid confusion, we temporarily call the second form $\bar{\mathcal{G}}(z, u)$; we will drop the bar once equality has been established.

We substitute $z = w/(w-1)$. This substitution is common in the analysis of skip-list structures (*Euler transform*). Here is the formula linking coefficients:

$$\begin{aligned} [z^n] f(z) &= \frac{1}{2\pi i} \oint f(z) \frac{dz}{z^{n+1}} = -\frac{1}{2\pi i} \oint f(z(w)) \frac{dw(w-1)^{n-1}}{w^{n+1}} \\ &= (-1)^n [w^n] f(z(w)) (1-w)^{n-1} = \sum_{k=0}^{n-1} \binom{n-1}{k} (-1)^{n-k} [w^{n-k}] f(z(w)). \end{aligned}$$

Thus

$$\bar{\mathcal{G}}(z, u) = \sum_{1 \leq r \leq s} (-1)^{s+1-r} \prod_{i=r}^s \frac{wupq^{i-1}}{1-wipq^{i-1}(1+q-u)}.$$

Then, by the principle just explained,

$$\begin{aligned}
[z^{n+1}]\mathfrak{G}(z, u) &= (-1)^{n+1}[w^{n+1}](1-w)^n \sum_{1 \leq r \leq s} (-1)^{s+1-r} \prod_{i=r}^s \frac{wupq^{i-1}}{1-wpq^{i-1}(1+q-u)} \\
&= (-1)^{n+1} \sum_{k=0}^n \binom{n}{k} (-1)^k [w^{n+1-k}] \sum_{1 \leq r, h} (-1)^h \prod_{i=r}^{r+h-1} \frac{wupq^{i-1}}{1-wpq^{i-1}(1+q-u)} \\
&= \sum_{k=0}^n \binom{n}{k} (-1)^{k-1} [w^{k+1}] \sum_{1 \leq r, h} (-1)^h \prod_{i=1}^h \frac{wupq^{i+r-2}}{1-wpq^{i+r-2}(1+q-u)} \\
&= \sum_{k=0}^n \binom{n}{k} (-1)^{k-1} \frac{q^{k+1}}{1-q^{k+1}} [w^{k+1}] \sum_{h \geq 1} (-1)^h \prod_{i=1}^h \frac{wupq^{i-2}}{1-wpq^{i-2}(1+q-u)} \\
&= \sum_{k=0}^n \binom{n}{k} (-1)^{k-1} \frac{(pq)^{k+1}}{1-q^{k+1}} [w^{k+1}] \sum_{h \geq 1} (-1)^h \prod_{i=1}^h \frac{wupq^{i-2}}{1-wq^{i-2}(1+q-u)}.
\end{aligned}$$

We still have to prove that

$$\begin{aligned}
\frac{u}{q} \prod_{l=0}^{k-1} \left(uq^l + \frac{u-q-1}{q} \right) &= (-1)^{k-1} [w^{k+1}] \sum_{h \geq 1} (-1)^h \prod_{i=1}^h \frac{wupq^{i-2}}{1-wq^{i-2}(1+q-u)} \\
&= [w^{k+1}] \sum_{h \geq 1} \prod_{i=1}^h \frac{wupq^{i-2}}{1+wq^{i-2}(1+q-u)},
\end{aligned}$$

or, in equivalent form:

$$u \prod_{l=1}^k \left(uq^l + u - q - 1 \right) = [w^{k+1}] \sum_{h \geq 0} \prod_{i=0}^h \frac{wupq^i}{1+wq^i(1+q-u)}.$$

Now set, with $v = 1 + q - u$,

$$H(w) := \sum_{h \geq 0} \prod_{i=0}^h \frac{wupq^i}{1+wq^i v},$$

then

$$H(w) = \frac{wu}{1+wv} + \frac{wu}{1+wv} H(wq),$$

or

$$(1+wv)H(w) = wu + wuH(wq).$$

With $a_k = [w^{k+1}]H(w)$, we find⁽ⁱ⁾

$$a_k + va_{k-1} = u[k=0] + uq^k a_{k-1},$$

or

$$a_k = u[k=0] + (uq^k - v)a_{k-1},$$

from which we find by iteration

$$a_k = u \prod_{l=1}^k (uq^l - v),$$

as desired.

⁽ⁱ⁾ We use Iverson's notation: $[P] = 1$ if P is true, zero otherwise.

3 Moments

We start from

$$[z^{n+1}]G(z, u) = \frac{u}{q} \sum_{j=0}^n (pq)^{j+1} \binom{n}{j} \frac{1}{1 - q^{j+1}} \prod_{l=0}^{j-1} \left(uq^l + \frac{u - q - 1}{q} \right).$$

Differentiate this, then set $u = 1$ to get the average:

$$\begin{aligned} & 1 + Q(Q + 1) \sum_{j=1}^n (pq)^{j+1} \binom{n}{j} \frac{1}{1 - q^{j+1}} \prod_{l=1}^{j-1} (q^l - 1) \\ &= 1 + p(Q + 1) \sum_{j=1}^n (pq)^j \binom{n}{j} \frac{1}{1 - q^{j+1}} (-1)^{j-1} (q)_{j-1} \\ &= 1 + p(Q + 1) \sum_{j=1}^n (pq)^j \binom{n}{j} \frac{1}{(1 - q^{j+1})(1 - q^j)} (-1)^{j-1} (q)_j \\ &= 1 - p(Q + 1) \frac{1}{2\pi i} \int \frac{n!(-1)^n}{z(z-1)\dots(z-n)} \frac{(pq)^z}{(1 - q^{z+1})(1 - q^z)} \frac{(q)_\infty}{(q^{z+1})_\infty} dz. \end{aligned}$$

This integral representation comes from *Rice's method*: We cite the survey paper (3) for that. The contour includes the poles $1, 2, \dots, n$ and no others. Changing the contour, one is lead to compute the outside poles as a compensation. Here, we have to consider the poles at $z = 0$ and at $z = \chi_k = 2\pi ik/L$. The machinery is explained in more detail in the earlier skip list papers (7; 4).

Thus we compute

$$\begin{aligned} & 1 + p(Q + 1)[z^{-1}] \frac{n!(-1)^n}{z(z-1)\dots(z-n)} \frac{(pq)^z}{(1 - q^{z+1})(1 - q^z)} \frac{(q)_\infty}{(q^{z+1})_\infty} \\ &= 1 + \frac{Q + 1}{L} H_n + \frac{Q + 1}{L} \ln(p) - (Q + 1)\alpha - \frac{(1 + q)^2}{2pq}. \end{aligned}$$

We use the (standard) abbreviation

$$\alpha = \sum_{i \geq 1} \frac{q^i}{1 - q^i}.$$

So the average $\mathbb{E}\mathcal{K}(n + 1)$ is asymptotic to

$$(Q + 1) \log_Q n + \frac{(Q + 1)\gamma}{L} + \frac{Q + 1}{L} \ln(p) - (Q + 1)\alpha - \frac{(1 + q)^2}{2pq} + \delta(\log_Q n) + 1.$$

To compute the second (factorial) moment, we have to differentiate

$$\frac{u}{q} \sum_{j=0}^n (pq)^{j+1} \binom{n}{j} \frac{1}{1 - q^{j+1}} \prod_{l=0}^{j-1} \left(uq^l + \frac{u - q - 1}{q} \right)$$

twice with respect to u , and then set $u = 1$. This leads to

$$2(\mathbb{E}\mathcal{K}(n + 1) - 1) + 2p(Q + 1) \sum_{j=1}^n \binom{n}{j} (-1)^{j-1} \frac{(pq)^j}{(1 - q^{j+1})(1 - q^j)} (q)_j \cdot T(j),$$

with

$$T(j) = \sum_{l=1}^{j-1} \frac{q^l + \frac{1}{q}}{q^l - 1} = -Q(j - 1) - (Q + 1) \sum_{l=1}^{j-1} \frac{q^l}{1 - q^l}.$$

Now it is easy to replace the discrete parameter j by a complex variable z :

$$T(z) = -Q(z - 1) - (Q + 1) \left(\alpha - \sum_{l \geq 1} \frac{q^{l+z}}{1 - q^{l+z}} - \frac{q^z}{1 - q^z} \right).$$

Around $z \sim 0$,

$$T(z) \sim \frac{Q+1}{Lz} + \frac{p}{2q} - z \left(L\beta(Q+1) + Q - \frac{L(Q+1)}{12} \right),$$

with

$$\beta = \sum_{l \geq 1} \frac{q^l}{(1-q^l)^2}.$$

Thus the residue computation at $z = 0$ leads to

$$2p(Q+1)[z^{-1}] \frac{n!(-1)^n}{z(z-1)\dots(z-n)} \frac{(pq)^z}{(1-q^{z+1})(1-q^z)} \frac{(q)_\infty}{(q^{z+1})_\infty} T(z).$$

Notice the following expansion:

$$\frac{(q)_\infty}{(q^{z+1})_\infty} \sim 1 - zL\alpha + z^2 \frac{L^2}{2} (\alpha^2 + \beta).$$

Now Maple computes the variance (fluctuations must be added, coming from the poles at χ_k). To summarize:

Theorem 1 *Expectation and variance of the $\mathcal{K}(n)$ -parameter are asymptotic to*

$$\mathbb{E}\mathcal{K}(n) \sim (Q+1) \log_Q n + \frac{(Q+1)\gamma}{L} + \frac{Q+1}{L} \ln(p) - (Q+1)\alpha - \frac{(1+q)^2}{2pq} + \delta(\log_Q n) + 1,$$

$$\begin{aligned} \mathbb{V}\mathcal{K}(n) &\sim Q(Q+1) \log_Q n + Q(Q+1) \frac{\gamma}{L} + (Q+1)^2 \frac{\pi^2}{6L^2} + Q(Q+1) \frac{\ln(p)}{L} - Q(Q+1)\alpha - (Q+1)^2 \beta \\ &\quad - \frac{2Q(Q+1)}{L} - \frac{(Q+1)^2(5Q^2 - 16Q - 1)}{12(Q-1)^2} + \delta_{\text{var}}(\log_Q n). \end{aligned}$$

The constants α and β are given by

$$\alpha = \sum_{i \geq 1} \frac{q^i}{1-q^i} \quad \text{and} \quad \beta = \sum_{i \geq 1} \frac{q^i}{(1-q^i)^2};$$

$\delta(x)$ and $\delta_{\text{var}}(x)$ are small periodic functions. Their Fourier coefficients could be given in principle.

Note that we actually did the computations for $\mathcal{K}(n+1)$ instead of $\mathcal{K}(n)$, but that does not make a difference for the main terms displayed here.

4 The cumulative \mathcal{K} -parameter

So far, in terms of the proposed new skip list, we computed the (average) cost to get to the last element. (Recall that for the analysis we think about the reversed sequence, thus we start with the first element.) But one also wants to know the cost to get to any of the other elements. So, we compute here the *average cumulative cost*

$$\mathbb{E}\mathcal{K}(1) + \dots + \mathbb{E}\mathcal{K}(n).$$

This is easy on the level of generating functions; it just means a multiplication by $1/(1-z)$:

$$\frac{1}{1-z} \mathcal{G}(z, u) = \frac{u}{q} \sum_{j=0}^{\infty} \frac{(pqz)^{j+1}}{(1-z)^{j+2}} \frac{1}{1-q^{j+1}} \prod_{l=0}^{j-1} \left(uq^l + \frac{u-q-1}{q} \right).$$

Now we differentiate this with respect to u , and set $u = 1$:

$$\frac{z}{(1-z)^2} + (Q+1) \sum_{j=1}^{\infty} \frac{(pz)^{j+1}}{(1-z)^{j+2}} \frac{q^j}{1-q^{j+1}} (-1)^{j-1} (q)_{j-1}.$$

This time it is more convenient to read off the coefficient of z^n :

$$\begin{aligned} & n + (Q + 1) \sum_{j=1}^{\infty} p^{j+1} [z^{n-j-1}] \frac{1}{(1-z)^{j+2}} \frac{q^j}{1-q^{j+1}} (-1)^{j-1} (q)_{j-1} \\ &= n + (Q + 1) \sum_{j=1}^{n-1} p^{j+1} \binom{n}{j+1} \frac{q^j}{1-q^{j+1}} (-1)^{j-1} (q)_{j-1} \\ &= n + Q(Q + 1) \sum_{j=2}^n p^j \binom{n}{j} \frac{q^j}{(1-q^j)(1-q^{j-1})} (-1)^j (q)_{j-1}. \end{aligned}$$

Thus, using Rice's integral once again, which this time uses a contour that enclosed the poles $2, \dots, n$ we must look at

$$\begin{aligned} & n - Q(Q + 1) [(z-1)^{-1}] \frac{n!(-1)^n}{z(z-1)\dots(z-n)} p^z \frac{q^z}{(1-q^z)(1-q^{z-1})} (q)_{z-1} \\ &= n - Q(Q + 1) [z^{-1}] \frac{n!(-1)^n}{(z+1)z(z-1)\dots(z-n+1)} p^{z+1} \frac{q^{z+1}}{(1-q^{z+1})(1-q^z)} (q)_z \\ &= n + nQ(Q + 1) [z^{-1}] \frac{(n-1)!(-1)^{n-1}}{(z+1)z(z-1)\dots(z-n+1)} p^{z+1} \frac{q^{z+1}}{(1-q^{z+1})(1-q^z)} (q)_z. \end{aligned}$$

This is eventually evaluated with the help of Maple.

Theorem 2 *The expected value of the cumulative $\mathcal{K}(n)$ -parameter is given by*

$$n \left[(Q + 1) \frac{H_{n-1}}{L} + (Q + 1) \frac{\ln(p)}{L} - (Q + 1)\alpha - \frac{Q + 1}{L} - \frac{Q^2 + 3}{2(Q - 1)} + \delta_{cum}(\log_Q n) \right].$$

Of course, one could replace H_{n-1} by $\ln n + \gamma$.

5 Additional analysis

As we explained in the introduction, in our parameter, steps downwards in the header structure as well as the single step from to the header to the actual data are not included. The latter one is just one step, so it does not require analysis. The first one, however, is the difference between the *maximum* of the word and the *level I* finally reached.

Now it is a folklore result that the maximum (sometimes called height) is $\sim \log_Q n$ (see (11; 6)). The level that is reached at the end is a parameter typically encountered in *approximate counting*, see (2; 8; 5).

We find it from the generating function (2.1)

$$\mathcal{F}(z, u, v) = \frac{u}{q} \sum_{j \geq 1} \frac{(pz)^j}{(1-z)^j} \frac{q^j v}{1-q^j v} \prod_{l=0}^{j-2} \left(uvq^l + \frac{u-q-1}{q} \right)$$

by setting $u = 1$:

$$\mathcal{F}(z, 1, v) = \frac{1}{q} \sum_{j \geq 2} \frac{(pz)^j}{(1-z)^j} \frac{q^j v}{1-q^j v} (-1)^{j-1} \prod_{l=0}^{j-2} (1 - vq^l) + \frac{1}{q} \frac{pz}{1-z} \frac{qv}{1-qv}.$$

Differentiating w.r.t. v , followed by $v = 1$ gives the generating function of the average of the final level reached:

$$\left. \frac{\partial}{\partial v} \mathcal{F}(z, 1, v) \right|_{v=1} = \frac{1}{q} \sum_{j \geq 2} \frac{(pz)^j}{(1-z)^j} \frac{q^j}{1-q^j} (-1)^j (q)_{j-2} + \frac{1}{p} \frac{z}{1-z}.$$

Now, reading off the coefficient of z^{n+1} , we get

$$\frac{1}{q} \sum_{j=2}^{n+1} p^j \binom{n}{j-1} \frac{q^j}{1-q^j} (-1)^j (q)_{j-2} + \frac{1}{p} = \sum_{j=1}^n p^{j+1} \binom{n}{j} \frac{q^j}{1-q^{j+1}} (-1)^{j-1} (q)_{j-1} + \frac{1}{p}.$$

The residue computation as before produces the asymptotic expansion of the average level, reached after $n + 1$ (or also n) random steps:

$$\log_Q n + \frac{\ln(p)}{L} - \alpha + \frac{\gamma}{L} + \frac{1}{2} + \delta_{\text{level}}(\log_Q n).$$

Since the average of the maximum is

$$\log_Q n + \frac{\gamma}{L} + \frac{1}{2} + \delta_{\text{maximum}}(\log_Q n),$$

the difference of the two parameters is (apart from fluctuations) asymptotic to $\alpha - \frac{\ln(p)}{L}$.

6 The \mathcal{K} -parameter for permutations

In this section, we consider, instead of words, permutations (which can also be interpreted as special words). The definition of the $\mathcal{K}(n)$ -parameter is the same. The analysis is easy, and there are probably no applications, so we keep this section rather short.

Consider the probability $\pi(n, l)$ that the $\mathcal{K}(n)$ -parameter is $\geq l$. Then the permutation must look like

$$(r)w_1(r+1)w_2(r+2)\dots(r+l-1)w_l.$$

Note that $w_1 \dots w_l$ is any permutation of $n - l$ letters; there are $(n - l)!$ of them. The number of ways to split them into l groups is $\binom{n-l}{n-l}$. The number of choices for the first letter $r = 1, \dots, n + 1 - l$ is $(n + 1 - l)$. Thus

$$\pi(n, l) = \frac{1}{n!} (n - l)! \binom{n - 1}{n - l} (n + 1 - l) = \frac{1}{(l - 1)!} - \frac{1}{n} \frac{1}{(l - 2)!}.$$

Therefore we find the expectation:

$$\mathbb{E}\mathcal{K}(n) \sim \sum_{l=1}^n \pi(n, l) \sim \left(1 - \frac{1}{n}\right)e.$$

A similar computation gives the variance (again apart from exponentially small terms):

$$\mathbb{V}\mathcal{K}(n) \sim e(3 - e) + \frac{e(2e - 3)}{n} - \frac{e^2}{n^2}.$$

7 Conclusion

The \mathcal{K} -parameter analysed in this paper is the analogue of the combined horizontal and vertical cost in the original skip list. This parameter is *on average* asymptotic to $\sim Q \log_Q n$, as shown for instance in (9; 7). Our present analysis gives a leading term $(Q + 1) \log_Q n$, so the logarithmic behaviour is preserved (as predicted in (1)), but with a larger constant.

In the section on the cumulative \mathcal{K} -parameter we discussed the *average* of the total search cost, the parameter *path length*, which, for a given word $a_1 \dots a_n$ is defined to be

$$\mathcal{P}(a_1 \dots a_n) := \mathcal{K}(a_1 \dots a_n) + \mathcal{K}(a_2 \dots a_n) + \dots + \mathcal{K}(a_n).$$

For the average this does not make a difference, but higher moments and distribution *are* different. This parameter is, apparently, much harder to analyse. We hope to report on it in a subsequent publication.

References

- [1] S. Cho and S. Sahni. Weight-biased leftist trees and modified skip lists. *ACM J. Exp. Algorithmics*, 3:approx. 27 pp. (electronic), 1998.
- [2] P. Flajolet. Approximate counting: a detailed analysis. *BIT*, 25(1):113–134, 1985.
- [3] P. Flajolet and R. Sedgewick. Mellin transforms and asymptotics: Finite differences and Rice's integrals. *Theoretical Computer Science*, 144:101–124, 1995.

- [4] P. Kirschenhofer, C. Martínez, and H. Prodinger. Analysis of an optimized search algorithm for skip lists. *Theoretical Computer Science*, 144:199–220, 1995.
- [5] P. Kirschenhofer and H. Prodinger. Approximate counting: An alternative approach. *RAIRO Theoretical Informatics and Applications*, 25:43–48, 1991.
- [6] P. Kirschenhofer and H. Prodinger. A result in order statistics related to probabilistic counting. *Computing*, 51:15–27, 1993.
- [7] P. Kirschenhofer and H. Prodinger. The path length of random skip lists. *Acta Informatica*, 31:775–792, 1994.
- [8] G. Louchard and H. Prodinger. Generalized approximate counting revisited. <http://www.ulb.ac.be/di/mcs/louchard/louchard.papers/gac.ps>, 2006.
- [9] T. Papadakis, I. Munro, and P. Poblete. Average search and update costs in skip lists. *BIT*, 32:316–332, 1992.
- [10] W. Pugh. Skip lists: a probabilistic alternative to balanced trees. *Communications of the ACM*, 33:668–676, 1990.
- [11] W. Szpankowski and V. Rego. Yet another application of a binomial recurrence. Order statistics. *Computing*, 43(4):401–410, 1990.

