THE PATH LENGTH OF RANDOM SKIP LISTS

Peter Kirschenhofer and Helmut Prodinger

Department of Algebra and Discrete Mathematics
Technical University of Vienna, Austria

Abstract. The skip list is a recently introduced data structure that may be seen as an alternative to (digital) tries. In the present paper we analyze the path length of random skip lists asymptotically, i.e. we study the cumulated successful search costs. In particular we derive a precise asymptotic result on the variance, being of order n^2 (which is in contrast to tries under the symmetric Bernoulli model, where it is only of order n). We also intend to present some sort of technical toolkit for the skilful manipulation and asymptotic evaluation of generating functions that appear in this context.

1. Introduction and Main Results

The *skip list* has recently been introduced by Pugh [15] as an alternative data structure to search trees. We give a short description in the next paragraphs and refer the reader to [13] for more detailed information. In [1], [13] some interesting analytic aspects are obtained about the average case performance of random skip lists, especially on the search costs. Analyzing search trees, probably the most important parameter is the *path length*, i.e. *the sum of the costs to find all the elements in the data structure*, ("successful search"), compare [4] and [12]. This parameter has not yet been analyzed for the skip list, and we devote this paper to its study.

In a skip list n elements ($n \geq 1$) are stored in a set of sorted linear linked lists according to the following rule: all elements are stored in sorted order in a linked list named "level 1", and each element stored in the linked list "level i" is included with (independent) probability q ($0 < q < 1$) in the linked list "level $i+1$". A header refers to the first element in each linked list and also contains the total number of linked lists, i.e. the "height" of the data structure. The number of linked lists which an element belongs to is constant as long as the element exists in the skip list. Therefore it suffices to store each element only once, together with an array of horizontal pointers that refer to the respective consecutive elements of the linked lists the element belongs to.

If we denote by a_i the number of linear linked lists that contain the i-th element, it follows that a skip list of size n can also be described by the corresponding n-tuple $(a_1, \dots, a_n)$. In the sequel we adopt the following model of *random skip lists* (compare [1], [13]). Each $a_i \in \mathbb{N}$ is the *outcome* of a random variable G_i

Key words and phrases. Data structures, generating functions, probabilistic analysis.

Typeset by $\mathcal{A}_{\mathcal{M}}\mathcal{S}$-TEX

that is distributed according to the *geometric distribution* with parameter p, i.e. $\mathrm{Prob}\{G_i = k\} = pq^{k-1}$, where $q = 1 - p$. The random variables are *independent*.

The reader is warned that in the earlier papers the roles of p and q are interchanged. Nevertheless, we find it easier to stay with the present notation that is classical in the context of the geometric distribution.

The *search* for the i-th element starts at the header of the top level linked list. This list is followed until the index of the next element in the list is greater or equal to i (or the reference is null). In this instance the search continues one level below. The search terminates at level 1 with an equality test as the last comparison.

In the figure below we depict the search path for the 4-th element in a random skip list of size 4.

Following [13] we define the *search cost* for the i-th element as the number of pointer expections excluding the last one for the equality test (i.e. 13 in the example above). Each instance of a pointer inspection is marked by a dot in the figure. Observe that the last pointer expection comes up with the *third element* at level 1. The search cost may be split up into the sum of a "*vertical cost*" corresponding to pointer inspections of pointers refering to elements at a position $\geq i$ (that initiate the continuation of the search one level below, 11 in the example) and a "*horizontal cost*" corresponding to the remaining pointer inspections, i.e. the number of horizontal steps on the path from the header to the i-th element (2 in the example). We define the *horizontal path length* $X_n = X_n(\rho)$ of a skip list ρ containing n elements as the sum of the horizontal search costs of all elements in ρ. The *vertical path length* $Y_n = Y_n(\rho)$ is defined analogously. Finally, the total path length or *total cost* C_n is $X_n + Y_n$. X_n, Y_n and C_n are random variables under the above defined probability model for random skip lists.

It is an easy observation that the vertical search cost for each element equals the height of the skip list, so that $Y_n(\rho)$ is n times the height of ρ. The horizontal search cost is more intricate. Let us return to our alternative description of the skip lists as n-tuples $(a_1, \ldots, a_n)$ from above. The height clearly equals the *maximum* of $(a_1, \ldots, a_n)$. The horizontal search cost is zero for the first element. Let now $i \geq 2$. If we follow the path from the header to the $i-1$-th element in reverse order we find that for the i-th element the horizontal cost is the *number of right-to-left maxima* in $(a_1, \ldots, a_{i-1})$, i.e. the number of elements a_j $(1 \leq j \leq i-1)$ that are larger than or equal to $a_j, a_{j+1}, \ldots, a_{i-1}$. (Observe that the convention to call a_{i-1} itself a right-to-left maximum ensures that the horizontal step connecting the path with the header is counted, too.) We may conclude that the horizontal path length of $(a_1, \ldots, a_n)$ is the sum of the numbers of right-to-left maxima of the sequences $a_1, \ldots, a_{i-1}$ for $2 \leq i \leq n$.

The number of right-to-left maxima of $a_1, \ldots, a_n$ is a well understood parameter in the context of *random permutations*, see [10], [11] and [16]. In the context of the geometric distribution it was analyzed in [14].

With respect to *expectations*, the horizontal path length is simply the sum of all the horizontal search costs. This is also easy to check, comparing [14, Lemma 2] and our formula (2.7). However, with respect to *variances* the parameters are different, and the path length has proved to be more interesting. (For tries, Patricia tries and digital search trees, these analyses are quite challenging and are to be found in [7], [8], [9].) The results about the variance of the horizontal path length X_n

and of the total cost C_n are thus considered to be our main findings in the present paper. We also intend to present some sort of *technical toolkit* for the manipulation of generating functions that will frequently occur with the analysis of skip list parameters.

In the following we summarize our main results. Here and in the whole paper, we will use the handy abbreviations $Q = q^{-1}$ and $L = \log Q$.

For the sake of completeness we start with a formulation of the asymptotics of the *expectations*. Observe that these estimates can also be computed from the results of [13] on the search cost of specified elements, though the notion of path length is not considered there. (iii) is equivalent to [13, Theorem 4].

Proposition 1. (i) *The expected value* $\mathrm{E}(X_n)$ *of the horizontal path length in a random skip list containing* n *elements fulfills for* $n \to \infty$

$$\mathrm{E}(X_n) = (Q-1)n\Big(\log_Q n + \frac{\gamma - 1}{L} - \frac{1}{2} + \frac{1}{L}\delta_1(\log_Q n)\Big) + O(\log n)$$

where γ *is Euler's constant and* $\delta_1(x) = \sum_{k\neq 0} \Gamma(-1 - \frac{2k\pi i}{L})e^{2k\pi i x}$ *is a continuous periodic function of period 1 and mean zero.* (Γ *is the Gamma-function.*)

(ii) *The expected value of the vertical path length fulfills for* $n \to \infty$

$$\mathrm{E}(Y_n) = n\log_Q n + n\Big(\frac{\gamma}{L} + \frac{1}{2} - \frac{1}{L}\delta_2(\log_Q n)\Big) + O(\log n),$$

where $\delta_2(x) = \sum_{k\neq 0} \Gamma(-\frac{2k\pi i}{L})e^{2k\pi i x}$.

(iii) *The expected value* $\mathrm{E}(C_n)$ *of the total cost fulfills for* $n \to \infty$

$$\mathrm{E}(C_n) = Qn\log_Q n + n\Big(\frac{(\gamma-1)Q+1}{L} - \frac{Q}{2} + 1 + \frac{1}{L}\delta_3(\log_Q n)\Big) + O(\log n)$$

where $\delta_3(x) = \delta_1(x) - \delta_2(x)$.

Remark. As a consequence of the exponential decrease of the Γ-function along vertical lines (compare (2.37)) we have that the amplitude of the functions $\delta_i(x)$ here and in the sequel is very small for all reasonable values of q, e.g. $0.1 \le q \le 0.9$.

The asymptotic equivalents for the second moments of the path lengths are given in the sequel, and the two corresponding theorems are the main results of this paper.

Theorem 1. *The variance* $\mathrm{Var}(X_n)$ *of the horizontal path length in a random skip list containing* n *elements fulfills for* $n \to \infty$

$$\mathrm{Var}(X_n) = (Q-1)^2 n^2 \left\{ \frac{Q+1}{2(Q-1)L} + \frac{1}{L^2} - \frac{\pi^2}{6L^2} + \frac{8\pi^2}{L^2} h\Big(\frac{4\pi^2}{L}\Big) - \alpha_1 + \delta_4(\log_Q n) \right\} + O(n^{1+\varepsilon}), \qquad \varepsilon > 0,$$

where $h(x) = \sum_{k\geq 1} \frac{e^{kx}}{\left(e^{kx}-1\right)^2}$, $\alpha_1 = L\sum_{k\geq 1} \frac{1}{k(L^2+4k^2\pi^2)\sinh(2k\pi^2/L)}$ *and* $\delta_4(x)$ *is again continuous, periodic of period 1 and mean zero.*

With regard to the vertical path length, i.e. n times the maximum element, we have from [5] or [17]

$$\operatorname{Var}(Y_n) = n^2\left\{\frac{\pi^2}{6L^2} + \frac{1}{12} - \alpha_2 + \delta_5(\log_Q n)\right\} + O(n^{1+\varepsilon}),$$

where $\alpha_2 = \frac{1}{L}\sum_{k\geq 1}\frac{1}{k\sinh(2k\pi^2/L)}$. Finally the following theorem holds.

Theorem 2. (i) *The covariance* $\operatorname{Cov}(X_n, Y_n)$ *of the horizontal and the vertical path lengths in a random skip list containing* n *elements fulfills for* $n\to\infty$

$$\begin{aligned}\operatorname{Cov}(X_n, Y_n) = (Q-1)n^2\Bigg\{&\frac{1}{L^2} - \frac{\pi^2}{6L^2} + \frac{8\pi^2}{L^2}h\Big(\frac{4\pi^2}{L}\Big) - \alpha_1 + \frac{1}{L}\sum_{k\geq 1}\frac{1}{Q^k-1}\\ &- \sum_{k\geq 1}\frac{k}{\left(Q^k-1\right)^2} + \delta_6(\log_Q n)\Bigg\} + O(n^{1+\varepsilon}),\end{aligned}$$

with h *and* α_1 *from Theorem 1.*

(ii) *The variance* $\operatorname{Var}(C_n)$ *of the total cost fulfills for* $n\to\infty$

$$\begin{aligned}\operatorname{Var}(C_n) = {}&\left(Q^2-1\right)n^2\left\{\frac{1}{2L} + \frac{1}{L^2} - \frac{\pi^2}{6L^2} + \frac{8\pi^2}{L^2}h\Big(\frac{4\pi^2}{L}\Big) - \alpha_1\right\}\\ &+ 2(Q-1)n^2\left\{\frac{1}{L}\sum_{k\geq 1}\frac{1}{Q^k-1} - \sum_{k\geq 1}\frac{k}{\left(Q^k-1\right)^2}\right\}\\ &+ n^2\left\{\frac{\pi^2}{6L^2} + \frac{1}{12} - \alpha_2 + \delta_7(\log_Q n)\right\} + O(n^{1+\varepsilon}).\end{aligned}$$

Theorems 1 and 2 show that the variances of the path lengths are of order n^2, whereas the variances of the corresponding parameter for regular symmetric tries (under the Bernoulli model or the Poisson model) are of order n. This means that random skip list parameters are not as closely concentrated around their expectations as the corresponding parameters for tries.

It is interesting to investigate the leading term of the variance of the total cost (Theorem 2 (ii)) for different values of q (resp. $Q = q^{-1}$) numerically. It turns out that the constant K in $\operatorname{Var}(C_n) \sim n^2(K + \delta_7(\log_Q n))$ takes its minimum value for $q = 0.31\ldots$, which differs a little from $q = \frac{1}{e} = 0.36\ldots$, where the main term of the expectation takes its minimum. Nevertheless we can conclude that for values of q close to $\frac{1}{e}$, the variance is close to its minimum, too. Below we give a small table of $q = Q^{-1}$ and the corresponding values of the constant K.

q	K
0.1	$10.57\ldots$
0.2	$3.16\ldots$
0.3	$2.15\ldots$
$\mathbf{0.31}\ldots$	$\mathbf{2.13}\ldots$
0.4	$2.44\ldots$
0.5	$3.66\ldots$
0.6	$6.41\ldots$
0.7	$12.96\ldots$
0.8	$33.03\ldots$
0.9	$149.35\ldots$

The proof of the theorems is given in the next two sections. Section 2 contains the considerations on the horizontal path length. It starts with a set-up of the appropriate generating functions, presents a transformation that allows to express the coefficients as complex contour integrals in a very efficient manner, and finally reveals the asymptotic equivalents of the expectation and the second moment as the sums of residues at certain poles. Section 3 contains the technical details concerning the second moment of the total cost. In Section 4 we present an alternative representation for the constants in the leading terms of Theorems 1 and 2 which is better suited for numerical evaluations if q is close to 0. For this purpose a series transformation result has to be proved that is of its own interest. In the final Section 5 we add some results on parameters that can be analyzed using the same toolkit as prepared for the analysis of the path length.

2. The horizontal path length

We start from the interpretation of a skip list ρ of size $n+1$ as a finite sequence $(a_1, \ldots, a_{n+1})$, $n \geq 0$, $a_i \in \{1, 2, 3, \ldots\}$. The horizontal path length $X_{n+1}(\rho)$ is the sum $b(\rho')$ of all numbers of right-to-left maxima in the subsequences $(a_1, \ldots, a_i)$, $1 \leq i \leq n$, of $\rho' = (a_1, \ldots, a_n)$. In order to obtain an expression for the corresponding probability generating function it is convenient to start from the following combinatorial decomposition:

Let m be the maximal element occuring in ρ'. Then we can write ρ' in a unique way as

$$\rho = \sigma m \tau, \quad \text{where} \quad \sigma \in \{1, \ldots, m-1\}^*, \;\; \tau \in \{1, \ldots, m\}^* \tag{2.1}$$

(fixing the leftmost occurence of the maximum m). From this decomposition it follows that

$$b(\rho') = b(\sigma m \tau) = b(\sigma) + b(\tau) + |\tau| + 1, \quad b(\varepsilon) = 0, \tag{2.2}$$

since the contribution of the leftmost maximum is 1 plus the number of the succeding elements. The reader might observe that relation (2.2), in the context of *permutations* ρ' (and thus binary search trees), describes the *right-sided* path length. However the probabilistic models are quite different. (Compare e.g. [10], [16].)

We introduce the bivariate generating functions $P^{=m}(z, y)$, where the coefficient of $z^n y^j$ is the probability that a skip list ρ with $n+1$ elements has maximum element (of the prefix ρ') m and horizontal path length j (of ρ). $P^{\leq m}(z,y)$ and $P^{<m}(z,y)$ are defined analogously. Then relation (2.2) translates into the functional equation

$$P^{=m}(z,y)=pq^{m-1}zyP^{<m}(z,y)P^{\leq m}(zy,y),\quad P^{=0}(z,y)=1. \tag{2.3}$$

Although no simple closed formula for the solution of (2.3) is available, it allows to extract sufficient information on the moments of the path length.

If we set $F^{*}(z)=\frac{\partial P^{*}}{\partial y}(z,1)$, where $*$ stands for $=m$, $\leq m$ or $<m$, the generating function $F(z)=\sum_{n\geq 0}\mathrm{E}(X_{n+1})z^n$ of the *expectations* is obtained as

$$F(z)=\lim_{m\to\infty}F^{\leq m}(z).$$

The derivative of (2.3) w.r.t. y at $y=1$ involves the terms

$$P^{\leq m}(z,1)=\frac{1}{1-z(1-q^m)}=:\frac{1}{[\![m]\!]}\quad\text{and}\quad \frac{\partial}{\partial z}P^{\leq m}(z,1)=\frac{1-q^m}{[\![m]\!]^2}$$

and reads

$$\frac{F^{\leq m}(z)-F^{<m}(z)}{pq^{m-1}z}=\frac{1}{[\![m]\!][\![m-1]\!]}+\frac{z(1-q^m)}{[\![m]\!]^2[\![m-1]\!]}+\frac{F^{\leq m}(z)}{[\![m-1]\!]}+\frac{F^{<m}(z)}{[\![m]\!]}.$$

Performing some algebra it translates into

$$F^{\leq m}(z)[\![m]\!]^2=F^{<m}(z)[\![m-1]\!]^2+\frac{pq^{m-1}z}{[\![m]\!]},$$

which can be solved by *iteration*, observing that $F^{\leq 0}(z)=0$:

$$F^{\leq m}(z)=\frac{p}{q}\frac{z}{[\![m]\!]^2}\sum_{i=1}^{m}\frac{q^i}{[\![i]\!]}. \tag{2.4}$$

Now we let m tend to infinity and obtain

$$F(z)=\frac{p}{q}\frac{z}{(1-z)^2}\sum_{i\geq 1}\frac{q^i}{[\![i]\!]}. \tag{2.5}$$

The generating function $H(z)$ of the second factorial moments can be obtained in the same way, although the computations are much more involved. We use the notations $H^{*}(z)=\frac{\partial^2 P^{*}}{\partial y^2}(z,1)$. The second derivative of (2.3) w.r.t. y at $y=1$ involves additionally the quantities

$$\frac{\partial^2 P^{\leq m}}{\partial z^2}(z,1)=\frac{2\big(1-q^m\big)^2}{[\![m]\!]^3},$$

and

$$K^{\leq m}(z):=\frac{d}{dz}F^{\leq m}(z)=\frac{p}{q}\cdot\left(\frac{1}{[\![m]\!]^2}\sum_{j=1}^{m}\frac{q^j}{[\![j]\!]^2}+\frac{2z(1-q^m)}{[\![m]\!]^3}\sum_{j=1}^{m}\frac{q^j}{[\![j]\!]}\right).$$

In this way we find

$$\begin{aligned}\frac{H^{\le m}(z)-H^{<m}(z)}{pq^{m-1}z} &= \frac{2z\,(1-q^m)}{[\![m-1]\!][\![m]\!]^3}+\frac{2z}{[\![m-1]\!]}K^{\le m}(z)+\frac{1}{[\![m-1]\!]}H^{\le m}(z)\\ &\quad+\frac{1}{[\![m]\!]}H^{<m}(z)+\frac{2}{[\![m-1]\!]}F^{\le m}(z)+\frac{2}{[\![m]\!]}F^{<m}(z)\\ &\quad+\frac{2z(1-q^m)}{[\![m]\!]^2}F^{<m}(z)+2F^{\le m}(z)F^{<m}(z).\end{aligned}$$

Again we get a first order difference equation

$$\begin{aligned}H^{\le m}(z)[\![m]\!]^2 = H^{<m}(z)[\![m-1]\!]^2+2pq^{m-1}z\cdot\Bigg\{&\frac{z\,(1-q^m)}{[\![m]\!]^2}+z[\![m]\!]K^{\le m}(z)\\ &+[\![m]\!]F^{\le m}(z)+\frac{[\![m-1]\!]}{[\![m]\!]}F^{<m}(z)\\ &+[\![m]\!][\![m-1]\!]F^{\le m}(z)F^{<m}(z)\Bigg\}.\end{aligned}$$

This equation can be solved for $H^{\le m}(z)$ by iteration; performing the limit for $m\to\infty$ we finally achieve the generating function $H(z)$:

$$\begin{aligned}H(z) &= 2\frac{p}{q}\frac{z}{(1-z)^2}\sum_{i\ge1}\frac{q^i}{[\![i]\!]^2}\\ &\quad-2\frac{p}{q}\frac{z}{(1-z)^2}\sum_{i\ge1}\frac{q^i}{[\![i]\!]}\\ &\quad-2\left(\frac{p}{q}\right)^2\frac{z^2}{(1-z)^2}\sum_{1\le j\le i}\frac{q^{i+j}}{[\![i]\!][\![j]\!]}\\ &\quad+4\left(\frac{p}{q}\right)^2\frac{z^2}{(1-z)^2}\sum_{1\le j\le i}\frac{q^{i+j}}{[\![i]\!]^2[\![j]\!]}\\ &\quad+2\left(\frac{p}{q}\right)^2\frac{z^2}{(1-z)^2}\sum_{1\le j\le i}\frac{q^{i+j}}{[\![i]\!][\![j]\!]^2}\\ &\quad+2\left(\frac{p}{q}\right)^2\frac{z^2}{(1-z)^2}\sum_{1\le j<i}\frac{q^{i+j}}{[\![i]\!][\![i-1]\!][\![j]\!]}\\ &\quad+2\left(\frac{p}{q}\right)^3\frac{z^3}{(1-z)^2}\sum_{i\ge1}\sum_{1\le j\le i}\sum_{1\le h<i}\frac{q^{i+j+h}}{[\![i]\!][\![i-1]\!][\![j]\!][\![h]\!]}.\end{aligned}\tag{2.6}$$

Now we turn to the *evaluation of the expectations* $\mathrm{E}(X_n)$. Since the generating function $F(z)$ is not too complicated, we find, e.g. by partial fraction decomposition and extracting the coefficients, that

$$\mathrm{E}(X_n)=[z^{n-1}]F(z)=\frac{p}{q}\sum_{i\ge1}\left[-Q^i+n+Q^i\left(1-q^i\right)^n\right]\tag{2.7}$$

where $Q = q^{-1}$, or, by expanding the binomial and interchanging the summations,

$$\mathrm{E}(X_n) = \frac{p}{q}\sum_{k=2}^{n}\binom{n}{k}(-1)^k\frac{1}{Q^{k-1}-1}. \tag{2.8}$$

In order to get the asymptotic equivalent for $\mathrm{E}(X_n)$ we might apply a Mellin transform approach (cf. [4]) to (2.7). Alternatively, the expression (2.8) can be rewritten as a complex contour integral (Rice's method), namely

$$\mathrm{E}(X_n) = \frac{p}{q}\cdot\frac{1}{2\pi i}\int_{\mathcal{C}} B(n+1,-z)\frac{1}{Q^{z-1}-1}dz, \tag{2.9}$$

where $B(x,y) = \dfrac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$ is the *Beta function* and the contour $\mathcal{C}$ encircles the poles $z = 2, 3, \ldots, n$. Now the asymptotics are obtained by extending the contour $\mathcal{C}$ "to the left" and collecting the additional residues. (Compare [3] for technical details of this procedure.)

In the integral (2.9) we have additional poles with real part larger than zero at $z = \chi_k$, $\chi_k = \frac{2k\pi i}{\log Q}$, $k \in \mathbb{Z}$. The computation of the residues is standard and can be performed e.g. by MAPLE. It yields for the double pole $z = 1$

$$\mathrm{Res}_{z=1} B(n+1,-z)\frac{1}{Q^{z-1}-1} = \frac{n}{L}\Big(H_{n-1} - 1 - \frac{L}{2}\Big),$$

with $L = \log Q$ and H_n the n-th Harmonic number, and thus the asymptotic contribution to $\mathrm{E}(X_n)$

$$\frac{p}{q}\,\frac{n}{L}\Big(\log n + \gamma - 1 - \frac{L}{2}\Big) + O(\log n).$$

Similarily the simple poles $z = 1 + \chi_k$, $k \neq 0$, contribute

$$\frac{p}{q}\,\frac{1}{L}\sum_{k\neq 0}\Gamma(-1-\chi_k)n^{1+\chi_k}\Big(1 + O(\frac{1}{n})\Big). \tag{2.11}$$

Collecting all contributions we have Proposition 1 (i).

In the sequel we concentrate on our main goal, namely the computation of the *variance* $\mathrm{Var}(X_n)$. We have

$$\mathrm{Var}(X_n) = [z^{n-1}]H(z) + \mathrm{E}(X_n) - \mathrm{E}(X_n)^2, \tag{2.11}$$

so that we need a well suited expression for the coefficients of $H(z)$.

For this purpose it is extremely useful to start with the following technical observations. In [2] Flajolet and Richmond made extensive use of the relation

$$A(z) = \sum_{n\geq 0} a_n z^n \longleftrightarrow \frac{1}{1-w}A\Big(\frac{w}{1-w}\Big) = \sum_{n\geq 0}\Big(\sum_{k=0}^{n}\binom{n}{k}a_k\Big)w^n. \tag{2.12}$$

Of course this relation can be inverted to read

$$A(z) = \sum_{n\geq 0}\Big(\sum_{k=0}^{n}\binom{n}{k}(-1)^k f(k)\Big)z^n \longleftrightarrow \frac{1}{1-w}A\Big(\frac{w}{w-1}\Big) = \sum_{n\geq 0} f(n)w^n. \tag{2.13}$$

The alternating sum can now be rewritten by Rice's method (compare above), and in this way we find

Lemma 1. *Let $A(z)$ be a formal power series and*

$$f_n = [w^n] \frac{1}{1-w} A\Big(\frac{w}{w-1}\Big), \qquad n \geq 0 \tag{2.14}$$

with $f_0 = f_1 = \cdots = f_{s-1} = 0$. Then

$$[z^n]A(z) = \frac{1}{2\pi i} \int\limits_{\mathcal{C}(s,\ldots,n)} B(n+1,-z) f(z) dz, \tag{2.15}$$

where $\mathcal{C}(s,\ldots,n)$ encircles the poles $z = s, s+1, \ldots, n$, B is the Beta function and $f(z)$ is analytic inside $\mathcal{C}$ with $f(n) = f_n$ for all $n \geq s$.

The following generalization will be useful.

Lemma 2. *With the assumptions of Lemma 1 we have*

$$[z^n] \frac{1}{\big(1-z\big)^r} A(z) = \frac{(-1)^r}{2\pi i} \int\limits_{\mathcal{C}(s+r,\ldots,n+r)} B(n+r+1,-z) f(z-r) dz. \tag{2.16}$$

Proof. With $z = \dfrac{w}{w-1}$ we have $(1-z)^{-r} = (-1)^r \dfrac{w^r}{z^r}$, so that

$$[z^n](1-z)^{-r}A(z) = (-1)^r [z^{n+r}] w^r A(z).$$

The result follows now from Lemma 1. □

In order to compute the coefficients f_n in (2.14) for the formal series $A(z)$ occuring in our problem we will frequently apply

Lemma 3. *Let C and D be formal power series. Then*

$$[w^n] \sum_{i \geq 0} C(wq^i) = \frac{1}{1-q^n} [w^n] C(w), \qquad n \geq 0, \tag{2.17}$$

$$[w^n] \sum_{i \geq 1} C(wq^i) = \frac{1}{Q^n - 1} [w^n] C(w), \qquad n \geq 1, \tag{2.18}$$

$$[w^n] \sum_{i \geq 1} iC(wq^i) = \frac{Q^n}{\big(Q^n - 1\big)^2} [w^n] C(w), \qquad n \geq 1, \tag{2.19}$$

$$[w^n] \sum_{1 \leq j \leq i} C(wq^i) D(wq^j) = \frac{1}{Q^n - 1} \sum_{m=1}^{n} \frac{1}{1-q^m} [w^m] C(w) \cdot [w^{n-m}] D(w), \tag{2.20}$$

and

$$[w^n] \sum_{1 \leq j < i} C(wq^i) D(wq^j) = \frac{1}{Q^n - 1} \sum_{m=1}^{n} \frac{1}{Q^m - 1} [w^m] C(w) \cdot [w^{n-m}] D(w), \tag{2.21}$$

Proof. Immediate. $\square$

Now we are ready to express $[z^{n-1}]H(z)$ as a complex contour integral. We confine ourselves here to show how the method works for two of the seven terms in (2.6).

Let us start with the relatively simple first one, namely

$$A_1(z) = \frac{z}{(1-z)^2}\sum_{i\geq 1}\frac{q^i}{[\![i]\!]^2} =: \frac{1}{(1-z)^2}\bar{A}_1(z). \tag{2.22}$$

With the substitution $z = \dfrac{w}{w-1}$ we have

$$\frac{1}{[\![i]\!]} = \frac{1-w}{1-wq^i} \tag{2.23}$$

and therefore

$$\frac{1}{1-w}\bar{A}_1\Big(\frac{w}{w-1}\Big) = -w\sum_{i\geq 1}\frac{q^i}{\big(1-wq^i\big)^2}.$$

From Lemma 3 (2.18) we find

$$[w^n]\sum_{i\geq 1}\frac{wq^i}{\big(1-wq^i\big)^2} = \frac{1}{Q^n-1}[w^n]\frac{w}{\big(1-w\big)^2} = \frac{n}{Q^n-1}.$$

Therefore, by Lemma 2,

$$[z^{n-1}]A_1(z) = \frac{-1}{2\pi i}\int\limits_{\mathcal{C}(3,\ldots,n+1)} B(n+2,-z)\frac{z-2}{Q^{z-2}-1}dz. \tag{2.24}$$

The most complicated term in $H(z)$ is

$$A_7(z) = \frac{z^3}{(1-z)^2}\sum_{i\geq 1}\sum_{1\leq j\leq i}\sum_{1\leq h<i}\frac{q^{i+j+h}}{[\![i]\!][\![i-1]\!][\![j]\!][\![h]\!]} =: \frac{1}{(1-z)^2}\bar{A}_7(z). \tag{2.25}$$

Here we have

$$\frac{1}{1-w}\bar{A}_7\Big(\frac{w}{w-1}\Big) = -\sum_{i\geq 1}\frac{wq^i}{\big(1-wq^i\big)\big(1-wq^{i-1}\big)}S^{\leq i}(w)S^{\leq i-1}(w), \tag{2.26}$$

where

$$S^{\leq i}(w) = \sum_{1\leq j\leq i}\frac{wq^j}{1-wq^j}.$$

Now, a partial fraction decomposition gives

$$\frac{wq^i}{\big(1-wq^i\big)\big(1-wq^{i-1}\big)} = \frac{1}{Q-1}\Big(\frac{wq^{i-1}}{1-wq^{i-1}} - \frac{wq^i}{1-wq^i}\Big),$$

so that the whole expression (2.26) telescopes to

$$\frac{-1}{Q-1}\left(\frac{w}{1-q}S^{\le 1}(w)S^{\le 0}(w)+\sum_{i\ge 1}\frac{wq^i}{1-wq^i}\left(\frac{wq^i}{1-wq^i}+\frac{wq^{i+1}}{1-wq^{i+1}}\right)S^{\le i}(w)\right).$$

Since $S^{\le 0}(w)=0$ we get with Lemma 3 (2.20)

$$\begin{aligned}[w^n]\frac{1}{1-w}\bar{A}_7\Big(\frac{w}{w-1}\Big)&=-\frac{1}{Q-1}\frac{1}{Q^n-1}\sum_{m=2}^{n-1}\frac{1}{1-q^m}\times\\&\times[w^m]\left(\Big(\frac{w}{1-w}\Big)^2+\frac{w^2q}{(1-w)(1-wq)}\right)\cdot 1.\end{aligned}\tag{2.27}$$

A short calculation shows that

$$[w^m]\left(\Big(\frac{w}{1-w}\Big)^2+\frac{w^2q}{(1-w)(1-wq)}\right)=m-1+\frac{1}{Q-1}(1-Q^{1-m}),$$

so that finally (2.27) is equal to

$$-\frac{1}{Q-1}\frac{1}{Q^n-1}\left(\binom{n-1}{2}+\frac{n-2}{Q-1}+\sum_{m=2}^{n-1}\frac{m-2}{Q^m-1}\right).\tag{2.28}$$

It is not difficult to find an analytic continuation, if we observe that

$$\sum_{m=2}^{n-1}\frac{m-2}{Q^m-1}=\sum_{m\ge 2}\frac{m-2}{Q^m-1}-\sum_{m\ge 0}\frac{n+m-2}{Q^{n+m}-1}.$$

Altogether Lemma 2 yields

$$\begin{aligned}[z^{n-1}]A_7(z)=\frac{-1}{2\pi i}\frac{1}{Q-1}&\int\limits_{\mathcal{C}(4,\dots,n+1)}B(n+2,-z)\frac{1}{Q^{z-2}-1}\times\\&\times\left(\binom{z-3}{2}+\frac{z-4}{Q-1}+\sum_{m\ge 2}\frac{m-2}{Q^m-1}-\sum_{m\ge 0}\frac{z+m-4}{Q^{z+m-2}-1}\right)dz.\end{aligned}\tag{2.29}$$

Observe that $\mathcal{C}(4,\dots,n+1)$ may be replaced by $\mathcal{C}(3,\dots,n+1)$ since $z=3$ is a removable singularity of the integrand.

The remaining sums $A_i(z)$ in expression (2.6) for $H(z)$ can be treated in a similar way. Collecting all contributions we find

Lemma 4.

$$\begin{aligned}[z^{n-1}]H(z)=2(Q-1)^2\cdot\frac{1}{2\pi i}\Bigg\{&\int\limits_{\mathcal{C}(2,\dots,n)}B(n+1,-z)\frac{1}{Q^{z-1}-1}\times\\&\times\left(z-2-\frac{1}{Q-1}+\alpha-\sum_{m\ge 0}\frac{1}{Q^{z+m-1}-1}\right)dz\\&+\int\limits_{\mathcal{C}(3,\dots,n+1)}B(n+2,-z)\frac{1}{Q^{z-2}-1}\times\\&\times\left((z-3)(z-1)-\frac{z-2}{Q-1}+z\alpha-z\sum_{m\ge 0}\frac{1}{Q^{z+m-2}-1}\right)dz\Bigg\}\end{aligned}\tag{2.30}$$

with $\alpha = \sum_{m\geq 1} \frac{1}{Q^m - 1}$.

In order to get the asymptotic behaviour of $[z^{n-1}]H(z)$ for $n\to\infty$ we extend the contour of integration to a large rectangle and shift the left side to the line $\operatorname{Re} z = 1+\varepsilon$, $\varepsilon > 0$. Then $[z^{n-1}]H(z)$ equals the sum of residues at the additional poles included in the contour with an error term of order $O(n^{1+\varepsilon})$ (compare [3] for technical details of the estimate of the error term). In our instance the additional poles originate from the second integral, and they are located at $z = 2$ (triple) resp. $z = 2+\chi_k$, $k \in \mathbb{Z}\setminus\{0\}$ (simple). The computation of the residues can be performed by hand, or, more preferably, e.g. by MAPLE, and we find for the residue at $z = 2$

$$2(Q-1)^2\binom{n+1}{2}\left\{\frac{H_{n-1}^2}{L^2}+\frac{H_{n-1}^{(2)}}{L^2}-\frac{H_{n-1}}{L}\left(1+\frac{2}{L}\right)\right.$$
$$\left.+\frac{1}{3}+\frac{2}{L^2}-2(\alpha+\beta)+\frac{1}{2L}+\frac{1}{(Q-1)L}\right\}, \tag{2.31}$$

where $L = \log Q$, H_n resp. $H_n^{(2)}$ are the Harmonic numbers of first resp. second order, and $\beta = \sum_{m\geq 1}\frac{1}{(Q^m-1)^2}$.

The asymptotic equivalent of (2.31) is

$$2(Q-1)^2\left\{\frac{n^2\log_Q^2 n}{2}+n^2\log_Q n\left(\frac{\gamma}{L}-\frac{1}{2}-\frac{1}{L}\right)\right.$$
$$+n^2\left(\frac{1}{6}+\frac{1}{L^2}-\alpha-\beta-\frac{\gamma}{2L}-\frac{\gamma}{L^2}+\frac{\gamma^2}{2L^2}+\frac{\pi^2}{12L^2}+\frac{1}{4L}+\frac{1}{2(Q-1)L}\right) \tag{2.32}$$
$$\left.+O\left(n\log^2 n\right)\right\}$$

The first order poles at $z = 2+\chi_k$, $k\in\mathbb{Z}\setminus\{0\}$, give a contribution of the form

$$n^2\delta_8(\log_Q n)+O(n), \tag{2.33}$$

where $\delta_8(x)$ is a continuous periodic function of period 1 and mean zero. The Fourier coefficients of $\delta_8(x)$ can be obtained explicitly from the residues at the above mentioned poles. They can be used to show that the amplitude of $\delta_8(x)$ is very small (because of the exponential decrease of the Γ-function along vertical lines). Since from the practical point of view $\delta_8(x)$ has almost no influence on the variance, we omit these calculations here.

Combining (2.32), (2.33) and Proposition 1 we get the asymptotic equivalent of $\operatorname{Var}(X_n)$ according to (2.11). The reader should take notice of the fact that the term of order n^2 in $\mathrm{E}(X_n)^2$ contains the square $\delta_1^2(\log_Q n)$ of the fluctuating term $\delta_1(\log_Q n)$, where $\delta_1(x)$ has mean zero. Therefore we have

$$\operatorname{Var}(X_n) = (Q-1)^2n^2\left\{\frac{1}{L^2}+\frac{1}{12}-\frac{1}{2L}+\frac{\pi^2}{6L^2}-2(\alpha+\beta)+\frac{1}{(Q-1)L}-\frac{1}{L^2}\left[\delta_1^2\right]_0\right\}$$
$$+n^2\delta_4(\log_Q n)+O\left(n^{1+\varepsilon}\right),\qquad \varepsilon>0, \tag{2.34}$$

where $\left[\delta_1^2\right]_0$ is the mean of $\delta_1^2(x)$, and $\delta_4(x)$ is a small periodic function of period 1 and mean *zero*.

The quantity $\alpha+\beta=\sum_{k\geq 1}\frac{Q^k}{\left(Q^k-1\right)^2}$ can be rewritten as

$$\alpha+\beta=h(L)=\frac{\pi^2}{6L^2}-\frac{1}{2L}+\frac{1}{24}-\frac{4\pi^2}{L^2}h\Big(\frac{4\pi^2}{L}\Big), \tag{2.35}$$

where

$$h(x)=\sum_{k\geq 1}\frac{e^{kx}}{\left(e^{kx}-1\right)^2}. \tag{2.36}$$

Equation (2.35) follows from the functional equation for the Dedekind η-function and is proved e.g. in [6]. The alternative representation for $\alpha+\beta$ given by (2.35) is extremely useful for numerical purposes, since the term $\frac{4\pi^2}{L^2}h(\frac{4\pi^2}{L})$ is very small for "reasonable" values of q resp. $Q=q^{-1}$.

The term $\left[\delta_1^2\right]_0$ can be computed explicitly from the Fourier expansion in Proposition 1. Since (compare e.g. [18])

$$|\Gamma(iy)|=\frac{\pi}{y\sinh\pi y}, \tag{2.37}$$

we find

$$\frac{1}{L^2}\left[\delta_1^2\right]_0=\alpha_1 \tag{2.38}$$

with α_1 defined in the theorem. Therefore the proof of Theorem 1 is complete.

3. The total cost

As already mentioned in the Introduction, the total cost C_n is the sum of the horizontal path length X_n and the vertical path length Y_n, where Y_n is n times the height of the skiplist. The variance of the latter parameter was already studied in [5] resp. [17] and we have the asymptotic equivalents for $\mathrm{E}(Y_n)$ resp. $\mathrm{Var}(Y_n)$ as presented in the Introduction. It is an easy consequence that $\mathrm{E}(C_n)=\mathrm{E}(X_n)+\mathrm{E}(Y_n)$ fulfills the asymptotic relation from Proposition 1 (ii).

In order to compute $\mathrm{Var}(C_n)=\mathrm{Var}(X_n+Y_n)$ we use

$$\mathrm{Var}(C_n)=\mathrm{Var}(X_n)+\mathrm{Var}(Y_n)+2\mathrm{Cov}(X_n,Y_n), \tag{3.1}$$

where the *covariance* $\mathrm{Cov}(X_n,Y_n)=\mathrm{E}(X_nY_n)-\mathrm{E}(X_n)\mathrm{E}(Y_n)$ will be studied in the sequel.

We start with the term $\mathrm{E}(X_nY_n)$. The generating function of $\frac{1}{n+1}\mathrm{E}(X_{n+1}Y_{n+1})$ is

$$\sum_{m\geq 1}F^{=m}(z)\Big(m\mathrm{Prob}(G\leq m)+\sum_{k>m}k\mathrm{Prob}(G=k)\Big) \tag{3.2}$$

where G stands for the random variable producing the last elment of the skip list $(a_1,\ldots)$. (Observe that the random variables in question were assumed to be i.i.d.) From (4.2) we get immediately the expression

$$\sum_{m\ge1} F^{=m}(z)\Big(m+\frac{q^m}{1-q}\Big)=\sum_{m\ge0}\Big(F(z)-(1-q^m)F^{\le m}(z)\Big). \tag{3.3}$$

Now from (2.4) and (2.5)

$$\begin{aligned}&\sum_{m\ge0}\Big(F(z)-(1-q^m)F^{\le m}(z)\Big)\\&=(Q-1)z\Bigg(\sum_{m\ge0}\Big(\frac{1}{(1-z)^2}-\frac{1-q^m}{[\![m]\!]^2}\Big)\sum_{i=1}^{m}\frac{q^i}{[\![i]\!]}+\sum_{m\ge0}\frac{1}{(1-z)^2}\sum_{i>m}\frac{q^i}{[\![i]\!]}\Bigg),\end{aligned}$$

which can be rewritten as

$$=(Q-1)z\Bigg(\frac{1}{1-z}\sum_{1\le i\le j}\frac{q^{i+j}}{[\![i]\!][\![j]\!]^2}+\frac{z}{(1-z)^2}\sum_{1\le i\le j}\frac{q^{i+j}}{[\![i]\!][\![j]\!]}+\frac{1}{(1-z)^2}\sum_{i\ge1}\frac{iq^i}{[\![i]\!]}\Bigg). \tag{3.4}$$

The evaluation of the coefficients may now be performed by the same technique as in Section 2. Observe that we use Lemma 3 (2.19) for the third sum. We find

$$\begin{aligned}\mathrm{E}(X_nY_n)=(Q-1)n\cdot\Bigg\{&\frac{1}{2\pi i}\int_{\mathcal{C}(2,\ldots,n)}B(n+1,-z)\frac{1}{Q^{z-1}-1}\times\\&\times\Big(3-z-\alpha+\sum_{m\ge0}\frac{1}{Q^{m+z}-1}+\frac{2}{Q^{z-1}-1}\Big)dz\\&+\frac{1}{2\pi i}\int_{\mathcal{C}(2,\ldots,n+1)}B(n+2,-z)\frac{1}{Q^{z-1}-1}\times\\&\times\Big(-\binom{z-1}{2}-\sum_{m\ge1}\frac{m}{Q^m-1}+\sum_{m\ge0}\frac{m+z-1}{Q^{m+z-1}-1}\Big)dz\Bigg\}.\end{aligned} \tag{3.5}$$

The computation of the residues at $z=1$ (third order pole) resp. $z=1+\chi_k$, $k\in\mathbb{Z}\setminus\{0\}$ (simple poles) gives the main asymptotic terms, from which we obtain

$$\begin{aligned}\mathrm{Cov}(X_n,Y_n)=(Q-1)n^2\Bigg\{&\frac{1}{L^2}+\frac{1}{12}-\frac{1}{L}+\frac{\pi^2}{6L^2}+\frac{\alpha}{L}-\sum_{k\ge1}\frac{k}{\left(Q^k-1\right)^2}-2(\alpha+\beta)\\&+\frac{1}{L^2}\big[\delta_1\delta_2\big]_0+\delta_6(\log_Q n)\Bigg\}+O(n^{1+\varepsilon}),\end{aligned} \tag{3.6}$$

with $\delta_6(x)$ of mean zero. Now

$$\begin{aligned}
\left[\delta_1\delta_2\right]_0 &= \sum_{k\neq 0}\Gamma(\chi_k)\Gamma(-1-\chi_k)\\
&= \sum_{k\neq 0}(-1+\chi_k)\Gamma(-1+\chi_k)\Gamma(-1-\chi_k)\\
&= -\left[\delta_1^2\right]_0 + 2\mathrm{Re}\sum_{k>0}\chi_k\left|\Gamma(-1-\chi_k)\right|^2\\
&= -\left[\delta_1^2\right]_0,
\end{aligned} \tag{3.7}$$

so that $\frac{1}{L^2}\left[\delta_1\delta_2\right]_0$ is the negative value from (2.38). Inserting the last result in (3.6) we find Theorem 2 (i). The variance $\mathrm{Var}(C_n)$ of the total cost is now computed by (3.1), and this completes the proof of Theorem 2 (ii).

4. Alternative Representations for the Constants

For values of q that are very close to 0, i.e. Q very large, the representations of the constants in the theorems are not well-suited for numerical evaluation. Therefore we present the following alternative expressions. Concerning $\alpha+\beta$ we stay with the original representation

$$\alpha+\beta = \sum_{k\geq 1}\frac{Q^k}{\left(Q^k-1\right)^2}.$$

With regard to $\left[\delta_1^2\right]_0$ the following transformation can be performed.

From the Fourier expansion of $\delta_1(x)$ in Proposition 1 we have

$$\left[\delta_1^2\right]_0 = \sum_{k\neq 0}\Gamma(-1+\chi_k)\Gamma(-1-\chi_k), \qquad \chi_k = \frac{2k\pi i}{L}. \tag{4.1}$$

Let us consider the function

$$g(z) = L\frac{\Gamma(-1+z)\Gamma(-1-z)}{e^{Lz}-1}. \tag{4.2}$$

Then $\left[\delta_1^2\right]_0$ is the sum of residues of $g(z)$ at the poles $z=\chi_k$, $k\in\mathbb{Z}\setminus\{0\}$. Furthermore

$$\mathrm{Res}_{z=0}\,g(z) = -\frac{L^2}{12}-\frac{\pi^2}{6}-1. \tag{4.3}$$

Combining these observations we can easily conclude that $(0<\varepsilon<1)$

$$\frac{1}{2\pi i}\int_{\varepsilon-i\infty}^{\varepsilon+i\infty} g(z)dz - \frac{1}{2\pi i}\int_{-\varepsilon-i\infty}^{-\varepsilon+i\infty} g(z)dz = \left[\delta_1^2\right]_0 - \frac{L^2}{12}-\frac{\pi^2}{6}-1 \tag{4.4}$$

since the Γ-function decreases exponentially towards $\pm i\infty$. Because of

$$\frac{1}{e^{-Lz}-1} = -1-\frac{1}{e^{Lz}-1}$$

we have

$$\frac{1}{2\pi i}\int_{-\varepsilon-i\infty}^{-\varepsilon+i\infty} g(z)dz = \frac{1}{2\pi i}\int_{\varepsilon-i\infty}^{\varepsilon+i\infty} g(z)dz + \frac{L}{2\pi i}\int_{\varepsilon-i\infty}^{\varepsilon+i\infty} \Gamma(-1+z)\Gamma(-1-z)dz. \tag{4.4}$$

Altogether

$$\left[\delta_1^2\right]_0 = \frac{L^2}{12} + \frac{\pi^2}{6} + 1 + \frac{2}{2\pi i}\int_{\varepsilon-i\infty}^{\varepsilon+i\infty} g(z)dz + \frac{L}{2\pi i}\int_{\varepsilon-i\infty}^{\varepsilon+i\infty} \Gamma(-1+z)\Gamma(-1-z)dz. \tag{4.5}$$

It is easily verified that both remaining integrals coincide with the negative sum of the corresponding residues at the poles with real part greater than 0. Collecting the residues at $z=1$ (double pole) and $z=k\in\mathbb{N}$, $k\geq 2$ (simple poles) we finally obtain

$$\begin{aligned}\left[\delta_1^2\right]_0 = {} & \frac{L^2}{12} + \frac{\pi^2}{6} + 1 - \frac{QL^2}{(Q-1)^2} - \frac{3}{2}\frac{QL}{Q-1} \\ & + 2L - 2L\log 2 - 2L\sum_{k\geq 2}\frac{(-1)^k}{(k+1)k(k-1)(Q^k-1)}\end{aligned} \tag{4.6}$$

Inserting (4.6) in expression (2.34) gives for the variance of the horizontal path length the following alternative formula.

Corollary 1.

$$\begin{aligned}\mathrm{Var}(X_n) = (Q-1)^2 n^2 \Bigg\{ & \frac{1}{L}\left(2\log 2 - 1 + \frac{5}{2}\frac{1}{Q-1} + 2\sum_{k\geq 2}\frac{(-1)^k}{(k+1)k(k-1)(Q^k-1)}\right) \\ & + \frac{Q}{(Q-1)^2} - 2\sum_{k\geq 1}\frac{Q^k}{(Q^k-1)^2} + \delta_4(\log_Q n)\Bigg\} + O\big(n^{1+\varepsilon}\big), \qquad \varepsilon > 0.\end{aligned}$$

In a similar manner we obtain

Corollary 2.

(i) $$\mathrm{Var}(Y_n) = n^2\left(\frac{\log 2}{L} + \frac{2}{L}\sum_{k\geq 1}\frac{(-1)^{k-1}}{k(Q^k-1)} + \delta_5(\log_Q n)\right) + O(n^{1+\varepsilon}),$$

(ii) $$\begin{aligned}\mathrm{Cov}(X_n, Y_n) = (Q-1)n^2\Bigg\{ & \frac{1}{L}\Bigg(2\log 2 - \frac{3}{2} + \frac{3}{2}\frac{1}{Q-1} + \sum_{k\geq 1}\frac{1}{Q^k-1} \\ & + 2\sum_{k\geq 2}\frac{(-1)^k}{(k+1)k(k-1)(Q^k-1)}\Bigg) \\ & + \frac{Q}{(Q-1)^2} - \sum_{k\geq 1}\frac{(k+1)Q^k}{(Q^k-1)^2} + \delta_6(\log_Q n)\Bigg\} + O(n^{1+\varepsilon}).\end{aligned}$$

Again, $\mathrm{Var}(C_n)$ is easily computed from the terms above using (3.1).

5. Variations

Here we first discuss the analogous notion of the path length where only strictly greater elements count as right-to-left maxima. This is quite natural from a combinatorial point of view. However, since there is no relevance with respect to a computer science problem, we confine ourselves to a sketch and, to ease the presentation, use the analogous notations without explicitly stating them.

We write ρ' in a unique way as

$$\rho' = \sigma m \tau, \quad \text{where} \quad \sigma \in \{1, \dots, m\}^*, \ \tau \in \{1, \dots, m-1\}^*;$$

then

$$b(\rho) = b(\sigma m \tau) = b(\sigma) + b(\tau) + |\tau| + 1, \quad b(\varepsilon) = 0.$$

Hence

$$P^{=m}(z,y) = pq^{m-1} zy P^{\leq m}(z,y) P^{<m}(zy,y), \quad P^{=0}(z,y) = 1,$$

and

$$\frac{F^{\leq m}(z) - F^{<m}(z)}{pq^{m-1}z} = \frac{1}{[\![m]\!][\![m-1]\!]} + \frac{z(1-q^{m-1})}{[\![m]\!][\![m-1]\!]^2} + \frac{F^{\leq m}(z)}{[\![m-1]\!]} + \frac{F^{<m}(z)}{[\![m]\!]}$$

or

$$F^{\leq m}(z)[\![m]\!]^2 = F^{<m}(z)[\![m-1]\!]^2 + \frac{pq^{m-1}z}{[\![m-1]\!]}.$$

Again we solve it by iteration and let m tend to infinity;

$$F(z) = p\frac{z}{(1-z)^2}\sum_{i\geq 0}\frac{q^i}{[\![i]\!]}.$$

So the expection $\mathrm{E}(X_n)$ is closely related to the old one; since there is the extra summand for $i = 0$, it is $p(n-1)$ plus q *times the expectation from before.* (The reader should remember that $\mathrm{E}(X_n)$ stands for the right-to-left maxima in a sequence of length $n-1$.)

Next, we discuss the case where "maximum" is replaced by "minimum". Let's start with the first variant (equality *counts*).

$$P^{=m}(z,y) = pq^{m-1} zy P^{>m}(z,y) P^{\geq m}(zy,y), \quad P^{=0}(z,y) = 1,$$

This time $F(z) = F^{\geq 1}(z)$ is the generating function of interest. With

$$[\![m]\!] := \frac{1}{1 - zq^m}$$

we get

$$F^{\geq m}(z)[\![m-1]\!]^2 = F^{>m}(z)[\![m]\!]^2 + \frac{pq^{m-1}z}{[\![m-1]\!]} = \sum_{j\geq m}\frac{pq^{j-1}z}{[\![j-1]\!]},$$

and therefore

$$F(z)=p\frac{z}{(1-z)^2}\sum_{j\ge0}\frac{q^j}{[\![j]\!]}.$$

To get the coefficient of z^{n-1} is now very simple. Apart from exponentially small terms, the expectation is

$$E_n\sim p\binom{n}{2}+pn\alpha-p\alpha_4,$$

with the constants

$$\alpha=\sum_{k\ge1}\frac{1}{Q^k-1}\qquad\text{and}\qquad\alpha_4=\sum_{k\ge1}\frac{k}{Q^k-1}=\alpha+\beta.$$

The other variant (equality does *not* count) leads to

$$P^{=m}(z,y)=pq^{m-1}zyP^{>m}(zy,y)P^{\ge m}(z,y)P^{>m}(zy,y),\qquad P^{=0}(z,y)=1,$$

and

$$F(z)=\frac{p}{q}\frac{z}{(1-z)^2}\sum_{j\ge1}\frac{q^j}{[\![j]\!]},$$

so that

$$E_n\sim\frac{p}{q}n\alpha-\frac{p}{q}\alpha_4.$$

Acknowledgment. Many computations in this paper have been performed resp. checked by the computer algebra system Maple.

References

1. L.Devroye, *A limit theory for random skip lists*, Annals of Applied Probability **2** (1992), 597–609.
2. P.Flajolet and L.B.Richmond, *Generalized Digital Trees and Their Difference-Differential Equations*, Random Structures and Algorithms **3** (1992), 305–320.
3. P.Flajolet and R.Sedgewick, *Digital Search Trees Revisited*, SIAM J. Comput. **15** (1986), 748–767.
4. P.Flajolet and J.Vitter, *Average-Case Analysis of Algorithms and Data Structures*, Handbook of Theoretical Computer Science Vol. A "Algorithms and Complexity", North-Holland, 1990, pp. 431–524.
5. P.Kirschenhofer and H.Prodinger, *A Result in Order Statistics Related to Probabilistic Counting*, Computing **51** (1993), 15-27.
6. P.Kirschenhofer, H.Prodinger und J.Schoißengeier, *Zur Auswertung gewisser numerischer Reihen mit Hilfe modularer Funktionen*, Zahlentheoretische Analysis II, E. Hlawka, Ed., Lecture Notes in Mathematics **1262** (1987), 108–110.
7. P.Kirschenhofer, H.Prodinger and W.Szpankowski, *On the Variance of the External Pathlength in a Symmetric Digital Trie*, Discrete Applied Mathematics **25** (1989), 129–143.
8. P.Kirschenhofer, H.Prodinger and W.Szpankowski, *On the balance property of Patricia tries: External Pathlength View*, Theoretical Computer Science **68** (1989), 1–17.
9. P.Kirschenhofer, H.Prodinger and W.Szpankowski, *Digital Search Trees again Revisited: The Internal Path Length Perspective*, submitted (1990).

10. P.Kirschenhofer, H.Prodinger and R.F.Tichy, *A Contribution to the Analysis of in situ Permutation*, Glasnik Mathematicki **22(42)** (1987), 269–278.
11. D.E.Knuth, *The Art of Computer Programming Vol.1*, Addison-Wesley, Reading, MA, 1968.
12. D.E.Knuth, *The Art of Computer Programming Vol.3*, Addison-Wesley, Reading, MA, 1973.
13. T.Papadakis, I.Munro and P.Poblete, *Average search and update costs in skip lists*, BIT **32** (1992), 316–332.
14. H.Prodinger, *Combinatorics of geometrically distributed random variables: Left-to-right maxima*, submitted (1992).
15. W.Pugh, *Skip lists: a probabilistic alternative to balanced trees*, Communications of the ACM **33** (1990), 668–676.
16. R.Sedgewick, *Mathematical analysis of combinatorial algorithms*, Probability Theory and Computer Science (G.Louchard and G.Latouche, eds.), Academic Press, 1983, pp. 125–205.
17. W.Szpankowski and V.Rego, *Yet another Application of a Binomial Recurrence: Order Statistics*, Computing **43** (1990), 401–410.
18. E.T. Whittaker and G.N. Watson, *A Course in Modern Analysis*, Cambridge University Press, 1927.

TU Vienna
Wiedner Hauptstrasse 8–10
A-1040 Vienna
Austria
E-mail address: kirsch@email.tuwien.ac.at, proding@email.tuwien.ac.at