Theoretical Computer Science 72 (1990) 147-167
North-Holland

ANALYSIS OF DYNAMIC ALGORITHMS IN KNUTH'S MODEL *

J. FRANÇON
Département Informatique, Université Louis Pasteur, 7 Rue René Descartes, 67000 Strasbourg, France

B. RANDRIANARIMANANA and R. SCHOTT
C.R.I.N., Université Nancy 1, 54506 Vandœuvre-lès-Nancy, France

Abstract. This paper analyzes the average behaviour of algorithms that operate on dynamically varying data structures subject to insertions I, deletions D, positive (resp. negative) queries Q^+ (resp. Q^-) under the following assumptions: if the size of the data structure is k ($k \in \mathbb{N}$), then the number of possibilities for the operations D and Q^+ is a linear function of k, whereas the number of possibilities for the ith insertion or negative query is equal to i. This statistical model was introduced by Françon [6, 7] and Knuth [12] and differs from the model used in previous analyses [2-7]. Integrated costs for these dynamic structures are defined as averages of costs taken over the set of all their possible histories (i.e. evolutions considered up to order isomorphism) of length n. We show that the costs can be calculated for the data structures serving as implementations of linear lists, priority queues and dictionaries. The problem of finding the limiting distributions is also considered and the linear list case is treated in detail. The method uses continued fractions and orthogonal polynomials but in a paper in preparation, we show that the same results can be recovered with the help of a probabilistic model.

1. Introduction

Since the pioneer works in [10-12], the analysis of dynamic algorithms has made great progress: it was shown in [6, 7] and [3, 4] that several list and tree organizations can be analyzed in a dynamic context. Integrated costs for these dynamic structures were defined as averages of costs taken over the set of all possible evolutions of the structure, considered up to order isomorphism. Using a method of continued fractions and orthogonal polynomials Flajolet et al. obtained explicit expressions for the expected costs and in some cases for the variances but with a statistic called markovian which is briefly described in Section 3. The same results were proved in [14] with a probabilistic analysis. Taking account of the remarks made by Françon and Knuth, we introduce a more natural statistic: the number of possibilities for the ith insertion or negative query is equal to i but if after some operations the structure contains k records, the number of possibilities for a deletion or positive query is a linear function of k (Section 3). Since we have to work with two indices (i and k), the analysis of dynamic algorithms is more difficult in this model.

* A first version of this paper was presented at CAAP'88. Partially supported by the ESPRIT II Basic Research Actions Program of the European Community under contract No. 3075 (Project ALCOM) and by the P.R.C. Mathématiques et Informatique.

The plan of the paper is as follows: Section 2 provides the set of necessary definitions for dynamic data structures. Section 3 describes the two models. The integrated cost is defined in Section 4. In Section 5, we prove a continued fraction theorem and we use it to derive enumeration results relative to sequences of operations. These results are then applied in Section 6 to provide explicit evaluations for the integrated costs. Section 7 concerns the limiting distributions of the costs. Finally, Section 8 discusses some of the implications of our work and outlines some of its further aspects.

2. Data types, histories

We consider here data structures subject to the following natural operations: insertion (I), deletion (D) and query (Q); one may distinguish between successful queries (Q^+) and unsuccessful ones (Q^-). The data types to be studied here are:

- *Linear lists* (LL): support I and D only.
- *Priority queues* (PQ): support I and D only. D is performed only on the key of minimal value.
- *Dictionaries*: support I, D and Q without restriction.

Definition 2.1. A *schema* is a word $o_1 o_2 \dots o_n \in \{I, D, Q^+, Q^-\}^*$ such that for all j, $1 \leq j \leq n$,

$$|o_1 o_2 \dots o_j|_I \geq |o_1 o_2 \dots o_j|_D. \tag{1}$$

A schema is to be interpreted as a sequence of requests (the keys operated on not being represented). The condition (1) is to be interpreted as follows: after the operations $o_1 o_2 \dots o_j$ have been performed on the structure, the resulting size is $\alpha_j(\Omega) = |o_1 o_2 \dots o_j|_I - |o_1 o_2 \dots o_j|_D$, which should always be nonnegative.

Definition 2.2. (i) A *linear list history* is a sequence of the form

$$h = o_1(r_1) o_2(r_2) \dots o_n(r_n) \tag{2}$$

where $\Omega = o_1 o_2 \dots o_n$ is a schema over the alphabet $\{I, D\}$ and the r_j are integers satisfying

$$0 \leq r_j < \alpha_{j-1}(\Omega) \qquad \text{if } o_j = D,$$

$$0 \leq r_j \leq |o_1 o_2 \dots o_{j-1}|_I \quad \text{if } o_j = I.$$

(ii) A *priority queue history* is a sequence of the form (2), the schema Ω is also over the alphabet $\{I, D\}$, the r_j satisfy

$$r_j = 0 \qquad \text{if } o_j = D,$$

$$0 \leq r_j \leq |o_1 o_2 \dots o_{j-1}|_I \quad \text{if } o_j = I.$$

(iii) A *dictionary history* is a sequence of the form (2) but the schema Ω is over the alphabet $\{I, D, Q^+, Q^-\}$ and the r_j satisfy

$$0 \leq r_j < \alpha_{j-1}(\Omega) \qquad \text{if } o_j = D \text{ or } Q^+,$$

$$0 \leq r_j \leq |o_1 o_2 \ldots o_{j-1}|_I + |o_1 o_2 \ldots o_{j-1}|_{Q^-} \qquad \text{if } o_j = I \text{ or } Q^-.$$

r_j is the rank (or position) of the key operated upon at step j.

Definition 2.3. If $O \in \{I, D, Q^+, Q^-\}$, *the number of possibilities for* O is the number of keys k for which $O(k)$ is defined. If this number is zero, O is said to be *impossible.*

3. The two models

In [12], Knuth considers the following operations on a data structure containing d keys (or numbers):

D_r stands for random deletion, in the sense that if d keys are present each is chosen for deletion with probability $1/d$,

D_q stands for priority queue deletion, i.e. deletion of the smallest key,

I_0 stands for insertion of a random number by order, in the sense that *the new number is equally likely to fall into any of the $d+1$ intervals defined by the d numbers still present as keys after previous insertions and deletions*; this is to be independent of the history by which these d numbers were actually obtained,

I stands for insertion of a random real number from (... the uniform distribution) on the interval $[0, 1]$. *Each random number inserted is assumed to have the same distribution, and it is to be independent of all previously inserted numbers.* Thus, if we look at n such random numbers (...) the $n!$ possible orderings (of these numbers) are equally likely, and the particular distribution involved has no effect on the behaviour of the data organization (i.e. the class of data structures together with associated algorithms for operating on these structures).

Knott [11] has shown that I_0 is a concept different from I (see also [13, Section 6.2.2]); this result has stimulated further research, notably [10, 12], and the present work.

The above assumptions can be generalized for dictionary case. Thus, the assumption about I_0 is also valid for Q_0^- (negative query of a random number by order), and the assumption about D_r is valid for Q_r^+ (random positive query). The assumption about I can be modified as follows in order to take account of Q^-:

I or Q^- stands for insertion or negative query of a random real number from (... the uniform distribution) on the interval $[0, 1]$. Each random number inserted or negatively searched is assumed to have the same distribution, and it is to be independent of all previously inserted and negatively searched numbers. In this paper, considering the Is and Q^-s (resp. I_0s and Q_0^-s) kind of insertion and negative query only is called *Knuth's model* (resp. *markovian model*). The markovian model has been introduced and studied by combinatorial methods in [6] and [7]; Flajolet

[5] has shown how the theory of continued fractions and orthogonal polynomials remarkably fits this model; further developments appear in [1-3]; distributions of costs, average costs, limiting profiles (defined in [1]) have been calculated for some sequences of operations for various data types, including priority queues and linear lists.

The following questions were raised in [6]: how to compute the corresponding costs in Knuth's model and are the costs sensitive to the model? The first answers for linear lists and priority queues were given in [8, 15], after reducing the calculations in Knuth's model to calculations in the markovian model. In this paper, we develop an algebraic method which permits us to recover all the results of [15] and to treat the dictionary case. The first step is to express the problem in a combinatorial way. Following [10], let us consider the sequence of operations *IIIDI*, the initial data structure being empty; let $x<y<z$ be the three keys inserted during the sequence *III*; let us consider a linear list, that is x or y or z is deleted with equal probability; let w be the key inserted by the fourth *I*s of this sequence; then, all four cases $w<x<y<z$, $x<w<y<z$, $x<y<w<z$, $x<y<z<w$ do occur with equal probability, whatever the key deleted. More generally, let us consider a sequence of operations $O_1O_2\ldots O_j$ of dictionary, the initial data structure being empty; any data type may be considered, linear list, priority queue, dictionary; assume O_j is the ith I or Q^- of the sequence; let $x_1<x_2<\cdots<x_{i-1}$ be the keys inserted and negatively searched during the sequence $O_1O_2\ldots O_{j-1}$, and let w be the ith inserted or negatively searched key. Then, all the cases $w<x_1<x_2<\cdots<x_{i-1}$, $x_1<w<x_2<\cdots<x_{i-1},\ldots,x_1<x_2<\cdots<x_{i-1}<w$ are equally likely, whatever the deleted keys. Put into combinatorial words: after j operations, whose i are I and Q^-s, thus $j-i$ are D and Q^+s, the size of the data structure is $k\leqslant 2i-j$; the keys of the data structure can be considered as a subset of k distinct objects of a set of size i any of the $\binom{i}{k}$ possible subsets being equally likely. We say that *the number of possibilities of the ith I or* Q^- (in a sequence of operations) *is equal to i* (*for Knuth's model*) whatever the size of the data when this insertion or negative query occurs. On the contrary, in the markovian model, we say that *the number of possibilities of an* I_0 (*resp.* Q_0^-) *operation is* $k+1$ *iff* k *is the size of the data structure* when this insertion (resp. negative query) occurs, whatever the past of the sequence and of the data structure. The differences between the two models appear in Tables 1 and 2.

Table 1
Possibility functions in the markovian model

Data type	Npos(I, k)	Npos(D, k)	Npos(Q^+, k)	Npos(Q^-, k)
Dictionary	$k+1$	k	k	$k+1$
Priority queue	$k+1$	1		
Linear list	$k+1$	k		

Table 2
Possibility functions in Knuth's model

Data type	Npos(ith I or Q^-)	Npos(D, k)	Npos(Q^+, k)
Dictionary	i	k	k
Priority queue	i	1	
Linear list	i	k	

Here Npos(O, k) is the number of possibilities of an operation O performed on a data structure of size k.

4. Integrated costs

If **H** is a finite set of histories then we can define the average cost (i.e. integrated cost) by

$$\text{cost}(H) = \frac{\sum_{h \in H} \text{cost}(h)}{\text{card}(H)}.$$

In our applications $\mathbf{H} = \mathbf{H}_{0,0,n} = \mathbf{H}_n$ the set of all possible histories of length n, starting and finishing with an empty file. Computing cost (**H**) is possible for data representations having a "randomness" or "stationary" property which we define now.

Let S_k be the set of states of size k.

Definition 4.1. The *standard probability distribution on* S_k is the probability distribution induced on the set S_k by all possible histories consisting of k insertions (with schema $I_k = II \dots I$ (k times)).

We let $p(s)$ denote the standard probability of the state s.

Definition 4.2. A data representation is *stationary* if for all k, the three probability distributions induced over S_k by all possible histories of schema $I^{k+1}D$, I^kQ^- and I^kQ^+ coincide with the standard probability on S_k.

Definition 4.3. For a stationary data representation, we define the *individual cost* CO_k of the operation $O \in \{I, D, Q^+, Q^-\}$ on a state of size k by the formula

$$\text{CO}_k = \sum_{s \in S_k} p(s) \text{cost}(O, s).$$

Definition 4.4. The *level crossing number* $\text{NO}_{k,n}$ is the number of operations of type O performed on a file of size k in the course of all histories of $\mathbf{H}_{0,0,n}$.

For stationary data representations we have

$$\text{cost}(\mathbf{H}_n) = \frac{1}{\text{card } \mathbf{H}_n} \sum_{k \geq 0} \sum_{O \in \{I, D, Q^+, Q^-\}} \text{NO}_{k,n} \text{CO}_k.$$

Table 3

Data type	Data representation	CI_k	CD_k	CQ_k^+	CQ_k^-
Dictionary	Sorted list	$\frac{1}{2}(k+2)$	$\frac{1}{2}(k+1)$	$\frac{1}{2}(k+1)$	$\frac{1}{2}(k+2)$
	Unsorted list	0	$\frac{1}{2}(k+1)$	$\frac{1}{2}(k+1)$	k
Priority queue	Sorted list	$\frac{1}{2}(k+2)$	0		
	Unsorted list	0	$k-1$		
	Binary tournament	$H_{k+1}-\frac{1}{2}$	$2\left(H_k-2+\frac{1}{k}\right)$		
	Pagodas	$2\left(1-\frac{1}{k+1}\right)$	$2\left(H_k-2+\frac{1}{k}\right)$		
Linear list	Sorted list	$\frac{1}{2}(k+2)$	$\frac{1}{2}(k+1)$		
	Unsorted list	0	$\frac{1}{2}(k+1)$		

Table 3 gives us the individual costs for each data representation. Here $H_k = 1+1/2+\cdots+1/k$.

5. Continued fraction theorem and enumerations in Knuth's model

It appears that Flajolet's continued fraction expansion for the generating function $H(z)$ of $H_n = \operatorname{card} \mathbf{H}_n$ is not true in our model since we have to work with the two parameters k and i. (For this reason, we used first a purely combinatorial approach [8, 15], but the dictionary case cannot be solved in this way).

Nevertheless if we denote by $H_{p+q+r}^{[h,p+q]}$ the number of histories of length $n=p+q+r$ with p insertions, q negative queries, $r=r_1+r_2$ where r_1 (resp. r_2) is the number of deletions (resp. positive queries) and if we consider the following generating function

$$H(t,x,z)=\sum_{p,q,r} H_{p+q+r}^{p+q} \frac{x^p t^q}{(p+q)!} z^r,$$

then we can prove the following theorem.

Theorem 5.1. *$H(t,x,z)$ has the following continued fraction expansion:*

$$H(t,x,z)=\cfrac{1}{1-t-i_0 z-\cfrac{s_2 xz}{1-t-i_1 z-\cfrac{s_2 xz}{\cdots}}}$$

where $i_k = \operatorname{Npos}(Q^+, k)$, $s_k = \operatorname{Npos}(D, k)$.

Sketch of the proof. Define the alphabet $X=\{I, Q^-, Q_0^+, Q_1^+, Q_2^+, \ldots, D_1, D_2, \ldots\}$ where O_j, $O \in \{D, Q^+\}$, denotes the operation O performed on a file of size j. Let $S^{[h]}$ denote the set of schemas represented by words over X having height $\leq h$, initial and final levels 0. The $S^{[h]}$ have the following regular descriptions: $S^{[0]} = (Q^- + Q_0^+)^*$; $S^{[1]} = (Q^- + Q_0^+ + I(Q^- + Q_1^+)^* D_1)^*$; $S^{[2]} = (Q^- + Q_0^+ + I(Q^- + Q_1^+ + I(Q^- + Q_2^+)^* D_2)^* D_1)^*$ and in general $S^{[h]}$ is obtained by substituting $Q^- + Q_h^+ + I(Q^- + Q_{h+1}^+)^* D_{k+1}$ for $Q^- + Q_h^+$ in the expression of $S^{[h]}$. If we let $H_{p+q+r}^{[h,p+q]}$ denote the number of histories of height $\leqslant h$, length $p+q+r$, with p insertions, q negative queries and r deletions and positive queries, and

$$H^{[h]}(t, x, z) = \sum_{p,q,r} H_{p+q+r}^{[h,p+q]} \frac{x^p t^q}{(p+q)!} z^r,$$

then using the morphism

$$I \to x, \qquad Q^- \to t, \qquad Q_k^+ \to q_k z, \qquad D_k \to d_k z,$$

we have

$$H^{[0]}(t, x, z) = \frac{1}{1 - t - q_0 z},$$

$$H^{[1]}(t, x, z) = \frac{1}{1 - t - q_0 z - \dfrac{d_1 xz}{1 - t - q_1 z}},$$

etc.; in general $H^{[h+1]}(t, x, z)$ is obtained by substituting

$$t + q_h z + \frac{d_{h+1} xz}{1 - t - q_{h+1} z}$$

for $t + q_h z$ in $H^{[h]}(t, x, z)$. The theorem follows by letting h go to infinity. □

Using the more economical notation,

$$H(t, x, z) = 1/1 - t - q_0 z - d_1 xz/\ldots/1 - t - q_h z - d_{h+1} xz/\ldots,$$

we apply Theorem 5.1 to our data structures and obtain

(a) *dictionary*

$$H^{\mathrm{DICT}}(t, x, z) = 1/1 - t - 0.z - 1.xz/1 - t - 1.z - 2xz/\ldots/1 - t - kz - (k+1)xz/\ldots,$$

(b) *linear list*

$$H^{\mathrm{LL}}(0, x, z) = 1/1 - 1.xz/1 - 2xz/\ldots/1 - kxz/\ldots,$$

(c) *priority queue*

$$H^{\mathrm{PQ}}(0, x, z) = 1/1 - xz/1 - xz/\ldots/1 - xz/\ldots$$

Remark 5.2. For *priority queues*,

$$H^{\mathrm{PQ}}(x, z) = \sum_{n\geq 0} H_{2n}^{\mathrm{PQ}} \frac{x^n}{n!} z^n,$$

verify

$$H^{\mathrm{PQ}}(x, z) = \frac{1}{1 - xzH^{\mathrm{PQ}}(x, z)}$$

and we have

$$H^{\mathrm{PQ}}(x, z) = \frac{1-\sqrt{1-4xz}}{2xz}.$$

This gives us

$$H_{2n}^{\mathrm{PQ}} = \frac{n!}{n+1}\binom{2n}{n}.$$

We have obtained the same result with the combinatorial approach [8, 15].

Remark 5.3. For *linear lists*,

$$H^{\mathrm{LL}}(x, z) = \sum_{n\geq 0} H_{2n}^{\mathrm{LL}} \frac{x^n}{n!} z^n.$$

Using the hypergeometric function

$$F(a, b, z) = 1 + ab\frac{z}{1!} + a(a+1)b(b+1)\frac{z^2}{2!} + \cdots,$$

we have

$$\frac{F(a, b+1; z)}{F(a, b; z)} = \frac{1}{1 - az\dfrac{F(a+1, b+1; z)}{F(a, b+1; z)}}$$

$$= 1/1 - az/1 - (b+1)z/1 - (a+1)z/1 - (a+1)z/1$$

$$-(b+2)z/\ldots$$

The substitution $a \to 1/2$, $b \to 0$, $z \to 2xz$ gives us

$$\frac{F(\frac{1}{2}, 1; 2xz)}{F(\frac{1}{2}, 0; 2xz)} = 1/1 - 1.xz/1 - 2xz/\ldots = H^{\mathrm{LL}}(x, z),$$

hence $H^{\mathrm{LL}}(x, z) = F(\frac{1}{2}, 1; 2xz)$ and $H_{2n}^{\mathrm{LL}} = n!\,n?$ where $n? = 1 \cdot 3 \cdot 5 \cdots (2n-1)$. The same result has been obtained with the combinatorial approach [8, 15].

5.1. Histories of bounded height

As direct consequence of Theorem 5.1, we have the following.

Proposition 5.4. *If*

$$H^{[h]}(t, x, z) = \sum_{p,q,r} H^{[h,p+q]}_{p+q+r} \frac{x^p t^q}{(p+q)!} z^r,$$

then $H^{[h]}(t, x, z)$ *has a rational generating function*

$$H^{[h]}(t, x, z) = \frac{P_h(t, x, z)}{Q_h(t, x, z)}$$

where P_h *and* Q_h *are polynomials that satisfy the relations*:

$$P_{-1} = 0, \qquad P_0 = 1,$$

$$P_h(t, x, z) = (1 - t - q_h z) P_{h-1}(t, x, z) - d_h xz P_{h-2}(t, x, z),$$

$$Q_{-1} = 0, \qquad Q_0(t, x, z) = 1 - t - q_0 z,$$

$$Q_h(t, x, z) = (1 - t - q_h z) Q_{h-1}(t, x, z) - d_h xz Q_{h-2}(t, x, z),$$

$$\deg(P_h) = \deg(Q_{h-1}) = h \quad \textit{with} \quad \deg(t^i x^j z^l) = i + j + l.$$

From now we put $t = x = yz$ (i.e.) we use an auxillary variable y to mark insertions I and negative queries Q^- and principal variable z for histories. So, we can write the above propositions only by means of y and z

$$H(y, z) = \sum_{n \geq 0} \sum_i H^i_n \frac{y^i}{i!} z^n \quad \text{and} \quad H^{[h]}(y, z) = \frac{P_h(y, z)}{Q_h(y, z)}.$$

Proposition 5.5. *Let*

$$H_{k,p}(y, z) = \sum_{n \geq 0} \sum_i H^i_{k,p,n} \frac{y^i}{i!} z^n$$

where $H^i_{k,p,n}$ *is the number of histories going from level k to level p in n steps with i insertions and negative queries, we have*

$$H_{k,p}(y, z) = \frac{1}{d_1 d_2 \dots d_p y^k z^{k+p}}$$
$$\times [Q_{k-1}(y, z) Q_{p-1}(y, z) H(y, z) - P_{\lambda-1}(y, z) Q_{\mu-1}(y, z)]$$

where $\lambda = \max(k, p)$ *and* $\mu = \min(k, p)$. *In particular*

$$H_{0,p}(y, z) = \frac{1}{d_1 d_2 \dots d_p z^p} [Q_{p-1}(y, z) H(y, z) - P_{p-1}(y, z)],$$

$$H_{k,0}(y, z) = \frac{1}{y^k z^k} [Q_{k-1}(y, z) H(y, z) - P_{k-1}(y, z)].$$

An alternative way of looking at the relations between $H(y,z)$ and $Q_k(y,z)$ is by means of orthogonality relations. Starting from the numbers

$$a_n(y)=\sum_i H_n^i \frac{y^i}{i!}, \qquad n\geq 0,$$

we introduce the linear form $\langle P(y,z)\rangle$ over polynomials

$$P(y,z)=\sum_{j=0}^{k} p_j(y)z^j$$

defined by

$$\langle P(y,z)\rangle=\sum_{j=0}^{k} p_j(y)\sum_i H_{..}^i \frac{y^i}{i!}.$$

This induces a scalar product $\langle P\,|\,Q\rangle=\langle P\cdot Q\rangle$ and we have the following.

Proposition 5.6. *Let $Q_k(y,z)=z^{k+1}Q_k(y,1/z)$ be the reciprocal polynomial of $Q_k(y,z)$ relative to z. Then we have*

$$\langle z^i\,|\,Q_{k-1}(y,z)\rangle=\begin{cases}0 & \textit{if } 0\leq i<k,\\ d_1d_2\cdots d_ky^k & \textit{if } i=k.\end{cases}$$

Proposition 5.6 gives us the following using the scalar product $\langle\,|\,\rangle$.

Proposition 5.7.

$$\sum_i H_{k,p,n}^i \frac{y^i}{i!}=\frac{1}{d_1d_2\cdots d_py^k}\langle z^n\,|\,Q_{k-1}(y,z)Q_{p-1}(y,z)\rangle.$$

The proofs of these propositions follow the proofs in the markovian model [3, 5].

5.2. Data structures and orthogonal polynomials

5.2.1. Dictionaries and Charlier y-polynomials

For this data type, we have

$$Q_{-1}(y,z)=1, \qquad Q_0(y,z)=z-y,$$
$$Q_k(y,z)=(z-y-k)Q_{k-1}(y,z)-kyQ_{k-2}(y,z), \quad k\geq 1.$$

This recurrence relation translates over the generating function

$$Q(y,z,t)=\sum_{k\geq 0} Q_{k-1}(y,z)\frac{t^k}{k!}$$

into the differential equation

$$(1+t)\frac{\partial}{\partial t}Q(y,z,t)=(z-(1+t)y)Q(y,z,t)$$

whose solution is

$$Q(y, z, t) = (1+t)^z \, e^{-yt}.$$

We can remark that $Q(1, z, t)$ is the exponential generating function of Charlier polynomials.

Theorem 5.8. *The Charlier y-polynomials associated with dictionaries in Knuth's model admit*

$$\sum_{k \geq 0} Q_{k-1}(y, z) \frac{t^k}{k!} = (1+t)^z \, e^{-yt}$$

for exponential generating function. As for dictionary histories,

$$h(y, z) = \sum_{n \geq 0} \sum_i H_n^i \frac{y^i}{i!} \frac{z^n}{n!} = e^{y(e^z - 1)}$$

thus

$$\sum_{n \geq 0} H_n \frac{z^n}{n!} = \frac{1}{2 - e^z}$$

and

$$H_n = \sum_i H_n^i = \sum_i i! \, S_{n,i}$$

or

$$H_n = \sum_{i \geq 0} \frac{i^n}{2^{i+1}}$$

where $S_{n,i}$ are the Stirling numbers (of the second kind). We have also

$$h(y, u, v, z) = \sum_{n,k,p} \sum_i H_{k,p,n}^i \frac{y^{i+k}}{i!} \frac{u^k}{k!} v^p \frac{z^n}{n!}$$
$$= \exp y[(1+u)(1+v) \, e^z - u - v - 1]$$

and

$$\sum_{n,k,p} H_{k,p,n} \frac{u^k}{k!} v^p \frac{z^n}{n!} = \frac{1}{2 - (1+u)(1+v) \, e^z + u + v}.$$

Proof. Computing $\langle Q(y, z, t) \rangle$ in two different ways yields

$$\langle Q(y, z, t) \rangle = \left\langle \sum_{k \geq 0} Q_{k-1}(y, z) \frac{t^k}{k!} \right\rangle = \sum_{k \geq 0} \langle Q_{k-1}(y, z) \mid Q_{-1}(y, z) \rangle \frac{t^k}{k!} = 1$$

on one hand, and

$$\langle Q(y, z, t) \rangle = \langle (1+t)^z \rangle \, e^{-yt} = \sum_{n \geq 0} \langle z^n \rangle \, e^{-yt} \frac{[\log(1+t)]^n}{n!}$$

$$\langle Q(y, z, t) \rangle = e^{-yt} \sum_{n \geq 0} \sum_i H_n^i \frac{y^i}{i!} \frac{[\log(1+t)]^n}{n!}$$

on the other hand. Setting $u=\log(1+t)$ leads to

$$h(y,u)=\sum_{n\geq 0}\sum_i H_n^i\frac{y^i}{i!}\frac{u^n}{n!}=\mathrm{e}^{y(\mathrm{e}^u-1)},$$

which is the "double" generating function of Stirling numbers of second kind denoted by $S_{n,i}$,

$$h(y,u)=\sum_{n\geq 0}\sum_i S_{n,i}y^i\frac{u^n}{n!}=\mathrm{e}^{y(\mathrm{e}^u-1)}.$$

Using Laplace transform relatively to y we obtain

$$\sum_{n\geq 0}\sum_i i!\,S_{n,i}y^i\frac{u^n}{n!}=\sum_{n\geq 0}\sum_i H_n^i y^i\frac{z^n}{n!}=\frac{1}{1-y(\mathrm{e}^z-1)}.$$

The first part of Theorem 5.8 is obtained by setting $y=1$. Let

$$A(y,z,u,v,w)=\langle Q(y,z,u)Q(y,z,v)Q(y,z,wh(y,w))\rangle$$
$$=\langle Q(y,z,u)Q(y,z,v)Q(y,z,w\,\mathrm{e}^{y(\mathrm{e}^w-1)})\rangle.$$

On one hand,

$$A(y,z,u,v,w)=\sum_{n,k,p}\langle Q_{k-1}(y,z)Q_{p-1}(y,z)z^n\rangle$$
$$\times\frac{1}{n!}\frac{u^k}{k!}\frac{v^p}{p!}[\log(1+w\exp[y(\mathrm{e}^w-1])]^n\exp[-yw\,\mathrm{e}^{y(\mathrm{e}^w-1)}].$$

Using Proposition 5.6 we have

$$A(y,z,u,v,w)$$
$$=\exp[-yw\,\mathrm{e}^{y(\mathrm{e}^w-1)}]\sum_{n,k,p}\left(\sum_i H_{k,p,n}^i\frac{y^{i+k}}{i!}\right)\frac{u^k}{k!}v^p\frac{[\log(1+w\exp[y(\mathrm{e}^w-1)])]^n}{n!}.$$

On the other hand

$$A(y,z,u,v,w)$$
$$=\exp[-y(u+v+w\,\mathrm{e}^{y(\mathrm{e}^w-1)})]\langle(1+u)^z(1+v)^z(1+w\,\mathrm{e}^{y(\mathrm{e}^w-1)})^z\rangle$$
$$=\exp[-y(u+v+w\,\mathrm{e}^{y(\mathrm{e}^w-1)})]\sum_{n\geq 0}[\log(1+u)(1+v)$$
$$\times(1+w\,\mathrm{e}^{y(\mathrm{e}^w-1)})]^n\frac{1}{n!}\langle z^n\rangle$$
$$=\exp[-y(u+v+w\,\mathrm{e}^{y(\mathrm{e}^w-1)})]h(y,\log(1+u)(1+v)(1+w\,\mathrm{e}^{y(\mathrm{e}^w-1)}))$$
$$=\exp[-y(u+v+w\,\mathrm{e}^{y(\mathrm{e}^w-1)})]\exp\{y[(1+u)(1+v)(1+w\,\mathrm{e}^{y(\mathrm{e}^w-1)})-1]\}.$$

Identifying the two expressions of $A(y,z,u,v,w)$ and setting

$$t=\log(1+w\,\mathrm{e}^{y(\mathrm{e}^w-1)})$$

we obtain

$$h(y, u, v, t) = \sum_{n,k,p} \sum_i H^i_{k,p,n} \frac{y^{i+k}}{i!} \frac{u^k}{k!} v^p \frac{t^n}{n!}$$
$$= \exp[y[(1+u)(1+v)e^t - u - v - 1]].$$

Using Laplace transform relatively to y and taking $y=1$ we have, with

$$H_{k,p,n} = \sum_i H^i_{k,p,n},$$

$$\sum_{n,k,p} H_{k,p,n} \frac{u^k}{k!} v^p \frac{t^n}{n!} = \frac{1}{2-(1+u)(1+v)e^t+u+v}. \qquad \square$$

This treatment applies mutadis mutandis to linear lists and we merely state the results.

5.2.2. Linear lists and Hermite y-polynomials

For this type, we have $Q_{-1}=1$, $Q_0(y,z)=z$, $Q_k(y,z)=zQ_{k-1}(z)-kyQ_{k-2}(z)$.

Theorem 5.9. *The Hermite y-polynomials associated with linear lists in Knuth's model admit for exponential generating function*

$$Q(y,z,t) = \sum_{k\geq 0} Q_{k-1}(y,z) \frac{t^k}{k!} = \exp\left(zt - y\frac{t^2}{2}\right)$$

and

$$Q_{k-1}(1,z) = \sum_i \frac{(-1)^i k!}{2^i i!(k-2i)!} z^{k-2i}.$$

As for histories

$$h(y,t) = \sum_{n\geq 0} H_n \frac{y^{\frac{1}{2}n}}{(\frac{1}{2}n)!} \frac{t^n}{n!} = \exp\left(y\frac{t^2}{2}\right).$$

Hence

$$H_{2n} = \frac{(2n)!}{2^n} \quad \text{and} \quad \sum_{n\geq 0} H_{2n} \frac{z^{2n}}{(2n)!} = \frac{2}{2-z^2}.$$

$$h(y,u,v,t) = \sum_{n,k,p} H_{k,p,n} \frac{y^{\frac{1}{2}(n+p+k)}}{[\frac{1}{2}(n+p-k)]!} \frac{u^k}{k!} v^p \frac{t^n}{n!}$$
$$= \exp\{y[uv+(u+v)t+\tfrac{1}{2}t^2]\}$$

and

$$\sum_{n,k,p} H_{k,p,n} \frac{u^k}{k!} v^p \frac{z^n}{n!} = \frac{2}{2-2uv-2(u+v)z-z^2}.$$

Remark 5.10

$$H^i_n = \delta_{i,\frac{1}{2}n} H^{\frac{1}{2}n}_n \quad \text{so} \quad H_n = H^{\frac{1}{2}n}_n,$$
$$H^i_{k,l,n} = \delta_{i,n+1-k} H^{\frac{1}{2}(n+1-k)}_{k,l,n} \quad \text{hence} \quad H_{k,l,n} = H^{\frac{1}{2}(n+l-k)}_{k,l,n}$$

where δ is the Kronecker symbol.

5.2.3. Priority queues and Tchebycheff y-polynomials

For this data type $Q_{-1} = 1$, $Q_0(y, z) = z$, $Q_k(y, z) = zQ_{k-1}(y, z) - yQ_{k-2}(y, z)$, $k \geq 1$. Thus

$$Q(y, z, t) = \sum_{k \geq 0} Q_{k-1}(y, z) t^k = \frac{1}{1 - zt + yt^2},$$
$$Q_{k-1}(1, z) = \sum_i (-1)^i \binom{k-i}{i} z^{k-2i}.$$

Theorem 5.11. *The Tchebycheff y-polynomials associated with priority queues in Knuth's model have the following generating function*:

$$Q(y, z, t) = \sum_{k \geq 0} Q_{k-1}(y, z) t^k = \frac{1}{1 - zt + yt^2}$$

and

$$Q_{k-1}(1, z) = \sum_i (-1)^i \binom{k-i}{i} z^{k-2i}.$$

As for histories

$$h(y, z) = \sum_{n \geq 0} \frac{H_n y^{\frac{1}{2}n}}{(\frac{1}{2}n)!} z^n = \frac{1 - \sqrt{1 - 4yz^2}}{2yz},$$

hence

$$H_{2n} = \frac{n!}{(n+1)} \binom{2n}{n}$$

and

$$\sum_{n \geq 0} H_{2n} \frac{z^{2n}}{(2n)!} = \frac{1}{z^2} (e^{z^2} - 1).$$

We have also

$$\sum_{n,k \geq 0} H_{0,k,n} \frac{y^{\frac{1}{2}(n+k)}}{[\frac{1}{2}(n+k)]!} v^k z^n = \sum_{n,k} H_{k,0,n} \frac{y^{\frac{1}{2}(n+k)}}{[\frac{1}{2}(n-k)]!} v^k z^n$$
$$= \frac{1 - \sqrt{1 - 4yz^2}}{2yz^2 - yvz(1 - \sqrt{1 - 4yz^2})}.$$

6. The integrated cost theorem in Knuth's model

The preceding section provides expressions for the number $H_{k,p,n}$ of histories of length n, starting at level k and finishing at level p.

Let $\mathrm{NO}^i_{k,n}$ be the level crossing number of operation $O \in \{I, D, Q^+, Q^-\}$ at level k for all histories with i insertions and negative queries, initial and final level 0. Setting

$$\mathrm{NO}_k(y, z) = \sum_{n,i} \mathrm{NO}^i_{k,n} \frac{y^i}{i!} z^n$$

we have the following.

Proposition 6.1. *In Knuth's model*

$$\mathrm{NI}_k(y, z) = yzH_{0,k}(y, z)H_{k+1,0}(y, z),$$

$$\mathrm{NQ}^-_k(y, z) = yzH_{0,k}(y, z)H_{k,0}(y, z),$$

$$\mathrm{NQ}^+_k(y, z) = q_k zH_{0,k}(y, z)H_{k,0}(y, z),$$

$$\mathrm{ND}_k(y, z) = d_k zH_{0,k}(y, z)H_{k-1,0}(y, z),$$

where

$$H_{a,b}(y, z) = \sum_{n,i} H^i_{a,b,n} \frac{y^i}{i!} z^n.$$

We are thus in possession of all the quantities needed in order to apply the integrated cost formula. If

$$\mathrm{KO}^i_n = \sum_{k \geq 0} \mathrm{CO}_k \mathrm{NO}^i_{k,n}$$

denotes the integrated cost of the operation $O \in \{I, D, Q^+, Q^-\}$ in the course of all histories of length n with i insertions and negative queries, initial and final level 0, then,

$$K_n = \sum_i (\mathrm{KI}^i_n + \mathrm{KD}^i_n + \mathrm{KQ}^{-i}_n + \mathrm{KQ}^{+i}_n)$$

represents the integrated cost for the histories in $\mathcal{H}_{n,0,0}$.

6.1. Integrated cost for priority queues

For this data type we have $\mathrm{NI}_{k,2n} = \mathrm{ND}_{k+1,2n}$. Setting

$$K(y, z) = \sum_{n \geq 0} \sum_i K^i_n \frac{Y^i}{i!} z^n$$

we have

$$K(y, z) = \sum_{n \geq 0} K_{2n} \frac{y^n}{n!} z^{2n}$$

where

$$K_{2n} = \sum_k (\mathrm{CI}_k + \mathrm{CD}_{k+1}) \mathrm{NI}_{k,2n}.$$

Theorem 6.2. *The generating function of the unitary costs*

$$C_{\mathrm{ID}}(x) = \sum_k (\mathrm{CI}_k + \mathrm{CD}_{k+1}) x^k$$

and integrated costs

$$K(y,z) = \sum_{n \geq 0} K_{2n} \frac{y^n}{n!} z^{2n}$$

for priority queues in Knuth's model are related by

$$K(y,z) = yz^2 B(y,z)^3 C_{\mathrm{ID}}(yz^2 B(y,z)^2)$$

with

$$B(y,z) = \frac{1-\sqrt{1-4yz^2}}{2yz^2}.$$

This theorem proceeds directly from Proposition 6.1. and Theorem 5.11 with $d_k = 1$.

6.2. Integrated cost for linear lists

For this data type we have also $\mathrm{NI}_{k,2n} = \mathrm{ND}_{k+1,2n}$. Setting

$$H_{k,p}(y,z) = \sum_n \sum_i H^i_{k,p,n} \frac{y^i}{i!} \frac{z^n}{n!}$$

and

$$\mathrm{NO}_k(y,z) = \sum_n \sum_i \mathrm{NO}^i_{k,n} \frac{y^i}{i!} \frac{z^n}{n!}$$

Proposition 6.1 becomes as follows.

Proposition 6.3.

$$\mathrm{NI}_k(y,z) = yH_{0,k}(y,z) * H_{k+1,0}(y,z),$$

$$\mathrm{ND}_k(y,z) = s_k H_{0,k}(y,z) * H_{k-1,0}(y,z),$$

$$\mathrm{NQ}^+_k(y,z) = q_k H_{0,k}(y,z) * H_{k,0}(y,z),$$

$$\mathrm{NQ}^-_k(y,z) = yH_{0,k}(y,z) * H_{k,0}(y,z),$$

where

$$(A * B)(y,z) = \int_0^z A(y, z-t) B(y,t)\, \mathrm{d}t$$

is the (Laplace) convolution.

This result proceeds from the following lemma.

Lemma 6.4. *If*

$$A(y,z)=\sum_{n\geq 0} a_n(y)\frac{z^n}{n!} \quad \text{and} \quad B(y,z)=\sum_{n\geq 0} b_n(y)\frac{z^n}{n!}$$

then

$$(A*B)(y,z)=\sum_{n\geq 0}\sum_{i=0}^{n} a_i(y)b_{n-i}(y)\frac{z^{n+1}}{(n+1)!}.$$

The treatment for linear lists in Knuth's model is the same as for priority queues in the markovian model (see [3, 5] for details) and we state merely the results.

Theorem 6.5. *The generating function*

$$K(y,z)=\sum_{n\geq 0} K_{2n}\frac{y^n}{n!}\frac{z^n}{n!}$$

for linear lists is related to

$$C_{\mathrm{ID}}(x)=\sum_{k\geq 0}(\mathrm{CI}_k+\mathrm{CD}_{k+1})x^{k+1}$$

by

$$K(y,z)=\frac{1}{\sqrt{1-2yz}}\,C_{\mathrm{ID}}\left(\frac{yz}{1-yz}\right).$$

6.3. Integrated cost for dictionaries

Theorem 6.6. *Setting*

$$K_n^i=\mathrm{KI}_n^i+\mathrm{KD}_n^i+\mathrm{KQ}_n^{+i}+\mathrm{KQ}_n^{-i}$$

the generating function

$$K(z)=\sum_{n\geq 0}\sum_{i}\frac{K_n^i}{i!}\frac{z^n}{n!}$$

is related to the exponential generating functions of unitary costs by the relation

$$K(y,z)=y\,\mathrm{e}^{y(\mathrm{e}^z-1)}\int_0^{(\mathrm{e}^{z/2}-1)^2}[(\mathrm{e}^z-u-1)C_{\mathrm{ID}}(yu)+2C_{Q^-}(yu)+2uC_{Q^+}(yu)]$$

$$\times\frac{\mathrm{e}^{-yu}\,\mathrm{d}u}{\sqrt{(\mathrm{e}^z+1-u)^2-4\mathrm{e}^z}}$$

with

$$C_{\mathrm{ID}}(x)=\sum_k(\mathrm{CI}_k+\mathrm{CD}_{k+1})\frac{x^k}{k!}$$

$$C_{Q^-}(x)=\sum_k \mathrm{CQ}_k^-\frac{x^k}{k!} \quad \text{and} \quad C_{Q^+}(x)=\sum_k \mathrm{CQ}_{k+1}^+\frac{x^k}{k!}.$$

Table 4

Data type	Data representation	Integrated cost (as $n \to \infty$)
Dictionary	Sorted list	$\left(\frac{5}{8 \log 2} - \frac{3}{4(\log 2)^2}\right) n^2 + O(n)$
	Unsorted list	$\left(\frac{7}{4 \log 2} - \frac{19}{16(\log 2)^2}\right) n^2 + O(n)$
Linear list	Sorted list	$n(n+5)/3$
	Unsorted list	$n(n+5)/6$
Priority queue	Sorted list	$\frac{\sqrt{\pi}}{4} n\sqrt{n} + O(n)$
	Unsorted list	$\frac{\sqrt{\pi}}{2} n\sqrt{n} + O(n)$
	Binary tournament	$\frac{3}{2} n \log n + O(n)$
	Pagodas	$n \log n + O(n)$

Table 4 summarizes the results for the integrated (time) costs. Some technical aspects are omitted here.

7. Limiting distributions

In this section, we study the asymptotic behaviour of linear lists in Knuth's model. It is a continuation of [1, 2]. In Section 7.1 we recall briefly some notions in order to apply them to linear lists (Section 7.2).

7.1. Limiting profiles

In order to calculate the integrated costs on the histories of length $2n$ (i.e. the average cost of an operation $O \in \{I, D\}$ for any history of length $2n$), we introduce the *average profile* notion *of an operation O at level k* on $H_{0,0,2n}$, which is the probability to have O at level k in the course of history, and define it by the quantity

$$\mathrm{TO}_{k,n} = \mathrm{NO}_{k,2n}/2nH_{2n}$$

where $\mathrm{NO}_{k,2n}$ is the level crossing number.

Then we deduce the average cost of an operation O on all the histories of length $2n$ by the formula

$$C_{2n} = \sum_{O \in \{I, D\}} \sum_{k \geq 0} \mathrm{CO}_k \mathrm{TO}_{k,2n}$$

where CO_k is the unitary cost. We can remark that for linear lists

$$\mathrm{NI}_{k,2n} = \mathrm{ND}_{k+1,2n}.$$

Consequently, it suffices to calculate $\mathrm{TI}_{k,2n}$ for this data type.

In [8, 15] we proved respectively the following relations

$$H_{2n}^{\mathrm{KLL}} = n!\,H_{2n}^{\mathrm{MPQ}} \quad \text{and} \quad \mathrm{NO}_{k,2n}^{\mathrm{KLL}} = n!\,\mathrm{NO}_{k,2n}^{\mathrm{MPQ}}$$

where KLL (resp. MPQ) means linear list in Knuth's model (resp. priority queue in the markovian model). So, we can say that $\mathrm{TO}_{k,2n}^{\mathrm{KLL}} = \mathrm{TO}_{k,2n}^{\mathrm{MPQ}}$ for $O \in \{I, D\}$.

7.2. Limiting profiles for linear lists in Knuth's model

Limiting profiles for MPQ are studied in [1] and [2]. We state merely the results.

Proposition 7.1. *The average profile of the operation $O \in \{I, D\}$ at level k is given by*

$$\mathrm{TI}_{k,2n}^{\mathrm{KLL}} = \mathrm{TD}_{k+1,2n}^{\mathrm{KLL}} \approx \frac{1}{2n\sqrt{1-2(k/n)}}.$$

Proposition 7.2. *The average cost of an operation is given by the formula*

$$C_{2n}^{\mathrm{KLL}} \approx \int_0^{1/4} (\mathrm{CI}_{2n\phi} + \mathrm{CD}_{2n\phi+1}) \frac{1}{\sqrt{1-4\phi}}\,\mathrm{d}\phi, \quad \phi = \frac{k}{2n}.$$

Application: for the sorted list representation, we obtain: $\mathrm{CI}_{2n\phi} = \mathrm{CD}_{2n\phi+1} = n\phi + 1$ and an elementary integration gives $C_{2n}^{\mathrm{KLL}} \approx n/6$, and the leading term of the asymptotic expansion of the integrated cost is

$$K_{2n} = 2nC_{2n}^{\mathrm{KLL}} \approx \frac{n^2}{3}.$$

We can remark that:

(i) the *limiting profile* $\mathrm{TI}_{k,x}^{\mathrm{KLL}}$ (Fig. 1) represents the contribution of the individual costs to the integrated cost C_{2n}^{KLL} (see Proposition 7.1.).

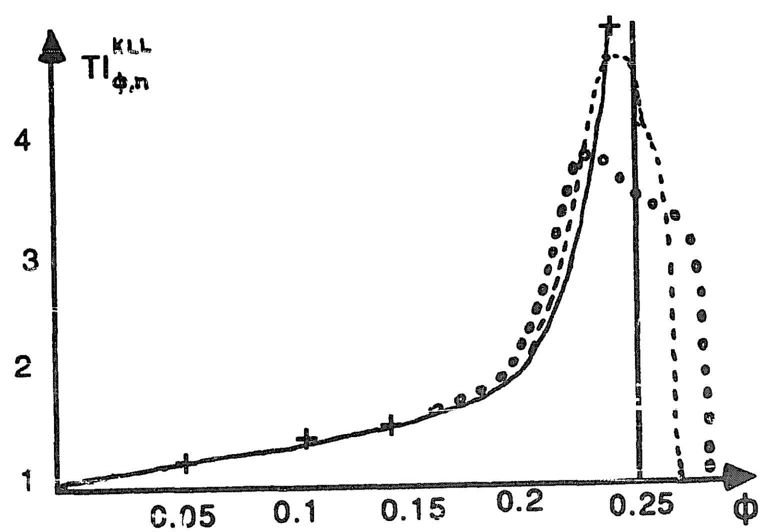

Fig. 1. *Geometrical interpretations*: ——, limit distribution $(1-4\Phi)^{-1/2}$; ∘ ∘ ∘, curve for $n = 800$; - - -, curve for $n = 1600$.

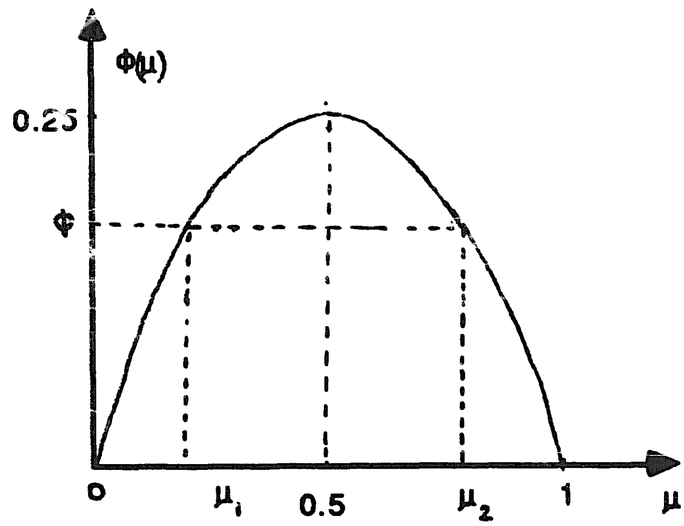

Fig. 2.

(ii) it is easier to calculate the integrated costs with Proposition 7.2 (rather than using Theorem 6.5).

Limiting profile for histories of $H_{0,0,2n}^{\mathrm{KLL}}$

According to [1, 2] (relative to MPQ) and the above remarks, Fig. 2 is the limiting profile of a random history of KLL which is a parabola representing the evolution of (the size of) the file

$$\Phi(\mu) = \mu - \mu^2.$$

The same curve has been obtained by Louchard [14] for priority queues in the markovian model by a probabilistic method. This curve reveals interesting properties concerning the behaviour of linear lists in Knuth's model:

(a) the top of the parabola tells us that, for large values of n, the size of the file never exceeds $\frac{1}{4}n$.

(b) in the first half-time there are more insertions than deletions, and the inverse phenomenon in the second half-time.

(c) the size of the file passes only twice to the same value.

(d) $|\mathrm{d}\mu/\mathrm{d}\Phi)| = (1-4\Phi)^{-1/2}$ represents the individual costs CO_k contribution to the integrated cost C_{2n}^{KLL}. Inversely, by integration of this expression we can recover the results of Propositions 7.1 and 7.2.

Limiting profiles for the other data structures can be obtained in the same way.

8. Conclusion

We have developed an algebraic method in order to analyze dynamic algorithms in Knuth's model. Our results prove that for dictionaries and linear lists the integrated costs are of the same kind in the two models but this is not true for priority queues represented by lists (whose cost is $O(n^2)$ in the markovian model and $O(n^{3/2})$ in Knuth's model). That means *the costs may be sensitive to the model.* Several works are in preparation, one of which concerns the probabilistic approach which permits us to find the explicit form of the limiting distribution of the costs measures and of course the variances. The case where the universe of keys is finite is the object of another study.

Acknowledgment

The authors are grateful to Ph. Flajolet, P. Lescanne, J.L. Rémy and J.M. Steyaert for several discussions on this topic.

References

[1] L. Chéno, Profils limites d'histoires sur les dictionnaires et les files de priorité. Application aux files binomiales. Thèse de 3è cycle, Université d'Orsay, 1981.

[2] L. Chéno, Ph. Flajolet, J. Françon, C. Puech and J. Vuillemin, Finite files, limiting profiles and variance analysis, in: *Proc. 18th Allerton Conf. on Com. Control and Computing* (1980).

[3] Ph. Flajolet, J. Françon and J. Vuillemin, Sequence of operations analysis for dynamic data structures, *J. Algorithms* **1** (1980) 111-141.

[4] Ph. Flajolet, C. Puech and J. Vuillemin, The analysis of simple lists structures, *Inform. Sci.* **38** (1986) 121-146.

[5] Ph. Flajolet, Analyse d'algorithmes de manipulation d'arbres et de fichiers, *B.U.R.O. Cahier* (1981) 34-35.

[6] J. Françon, Combinatoire des structures de données, Thèse de doc. d'Etat, Université de Strasbourg, 1979.

[7] J. Françon, Histoires de fichiers, *RAIRO Inform. Théor.* **12** (1978) 49-62.

[8] J. Françon, B. Randrianarimanana and R. Schott, Analysis of dynamic data structures in D.E. Knuth's model, Rapport C.R.I.N., 1986 (submitted).

[9] J. Françon, B. Randrianarimanana and R. Schott, Analysis of dynamic algorithms in D.E. Knuth's model, in: *Proc CAAP'88*, Lecture Notes in Computer Science **299** (Springer, Berlin, 1988) 72-88.

[10] A. Jonassen and D.E. Knuth, A trivial algorithm whose analysis isn't, *J. Comput System Sci.* **16** (1978) 301-332.

[11] G.D. Knott, Deletion in binary storage trees, Report Stan-CS, May 1975, 75-491.

[12] D.E. Knuth, Deletions that preserve randomness, *Trans. Software Eng.* (1977) 351-359.

[13] D.E. Knuth, *The Art of Computer Programming: Sorting and Searching*, Vol. 3 (Addison-Wesley, Reading, MA, 1975).

[14] G. Louchard, Random walks, Gaussian processes and list structures, *Theoret. Comput. Sci.* **53** (1987) 99-124.

[15] B. Randrianarimanana, Analyse des structures de données dynamiques dans le modèle de D.E. Knuth, Thèse de 3ème cycle, Université Nancy 1, 1986.