PREFACE

This special issue is devoted to the *Average-Case Analysis of Algorithms.* Analysis of algorithms aims at a precise prediction of the expected performances of algorithms and data structures of general use in computer science. Quantification of performance goes from mean value estimates of costs to analysis of the probability distributions involved, under natural randomness models of the data.

The area of analysis of algorithms was born on July 27, 1963 when D. E. Knuth analyzed in his "Notes on Open Addressing" hashing tables with the linear probing. The reader of the Knuth note can find for the first time the connection between the analysis of full tables and what was to become later known as the Knuth-Ramanujan function $Q(n)$. Knuth's note of 1963 concludes with a conjecture about the asymptotics of the Knuth-Ramanujan function (which Knuth proved in 1965). In his note Knuth also asked for the analysis of the variance of the construction cost of linear probing hashing. This problem turns out to be significantly more difficult and only in 1997 Knuth and independently Poblete and Viola solved the problem. We have had the privilege to handle both papers for this special issue (Poblete and Viola merged forces with Flajolet and as a result the variance analysis was extended to a full distributional analysis of linear probing hashing.)

Since 1963 the field was undergoing tangible changes. We see now the emergence of combinatorial and asymptotic methods that permit us to classify data structures into broad categories that are susceptible of a unified treatment. Probabilistic methods that have been so successful in the study of random graphs and hard combinatorial optimization problems also play an important rôle in the field. These developments have two important consequences for the analysis of algorithms: it becomes possible to predict average-case behaviour under more *general* probabilistic models; at the same time it becomes possible to analyze much more structurally *complex* algorithms.

In July 1993, the first seminar entirely devoted to analysis of algorithms took place in Dagstuhl, Germany. On that occasion a special issue of *Theoretical Computer Science* (vol. 144, No. 1–2, 1995) was prepared. In 1995 a second seminar on *Average-Case Analysis of Algorithms* was held again in Dagstuhl and a special issue of *Random Structures & Algorithms* (vol. 10, No. 1–3, 1997) was edited by P. Flajolet and W. Szpankowski. Our current special issue is an outcome of the *Third Dagstuhl Seminar on Average-Case Analysis of Algorithms* that took place in 1997.

Since 1997, the Analysis of Algorithms community has a homepage (`http://www-rocq.inria.fr/algo/AofA/index.html`). There, one can find e. g. the fascinating story about linear probing hashing in more detail.

This special issue is dedicated to our colleague, teacher, and friend **Philippe Flajolet** on the occasion of his 50-th birthday (December 1, 1998). In the article "Philippe Flajolet's Research in Analysis of Algorithms and Combinatorics" we tried to give a snapshot of Flajolet's research and his lasting contributions to the field. This paper was written without giving any prior notice to Philippe Flajolet. It reflects our view and any misunderstandings and shortcomings should be put on our account. In fact, this special issue is a "surprise present" to Philippe Flajolet for his selfless and inspiring work. *Sto lat, Philippe! Ad multos annos!*

Contributions

The classical models of randomness in the analysis of algorithms include permutations, words, trees, strings, numbers and graphs.

Random permutations lie at the heart of all sorting and searching algorithms that operate by comparisons between items. Recurrences arising in Batcher's sorting algorithm are investigated by H.-K. Hwang; as a second showcase of his powerful machinery (elementary arguments and/or Mellin type analysis) he treats a divide-and-conquer recursion occurring in a paper of Reingold and Supowit about Euclidean matching. Mahmoud and Smythe consider the limiting distribution of the number of comparisons made by the multiple QuickSelect algorithm. They prove the existence of a limiting distribution by a Martingale argument; many fascinating questions in this area still wait to be discovered. Frieze, de la Vega and Santha derive the average running time for the merging algorithm of Hwang and Lin. An interesting permutation model is considered by Bruss, Drmota, and Louchard who analyzed a variant of the secretary problem: Two decision makers observe a permutation of n equally likely options (say, numbers $1, 2, \ldots, n$) and each has one draw to select an option. What is the optimal strategy and what is the corresponding winning probability? The problem is solved by a mixture of analytic, probabilistic and computer algebra tools.

Random words and random allocation problems classically arise in hashing algorithms. Hashing with linear probing is analyzed by Knuth (variance of the number of probes). The key to success is the skillful manipulation of the (implicitly defined) tree function, exploiting in this way beautiful connections with the enumeration of graphs.

As already mentioned, Flajolet, Poblete and Viola dealt independently with the same subject. Additionally, they are able to identify the limit distributions of the construction cost of hash tables for two models, namely, full tables and sparse tables. In this way, another appearance of the otherwise not so well known Airy distribution has been brought to light.

The simple data compression algorithm by Merlini, Sprugnoli and Verri is analyzed by using Schützenberger's methodology to code events by words and thus translates appropriate languages into suitable generating functions.

Algorithms in number theory. Don Knuth referred to us "a very exciting paper" by Brigitte Vallée who for the first time analyzed rigorously the binary Euclidean `gcd` algorithm proving a twenty years old conjecture of R. Brent. The method of the proof is analytic and it is based on functional analysis and operators. This approach has already found many successful applications (e.g., digital tree analysis, algorithm of Gauss, continued fractions, etc.).

String models are important in understanding the complexity of efficient string searching algorithms (e.g., for data compression). Certain languages, generating functions and analytic asymptotics are used by Régnier and Szpankowski to provide a complete characterization (moments, limiting distribution, and large deviations) of the number of a given pattern occurrences in a random text.

Random tree models exist in many versions. One of the most prominent is the binary search tree which is at the heart of the quicksort algorithm. L. Devroye a couple of years ago presented a probabilistic derivation for the typical behavior of the height of such a tree. We still do not know whether the variance of the height is $O(1)$ or $O(\log\log n)$. We expect to answer this question once we find an analytic approach to deal with the height. M.

Drmota gives in his paper the first promising attempt to handle this problem. A retarded differential equation and its careful analysis are the key ingredients in this approach. There is still need for more work in this area!

The combinatorial model of random trees where all trees of a given size are taken with equal likelihood has found many applications in the past. In this issue Panholzer and Prodinger analyze priority trees through a combination of analytic tools and computer algebra. These trees were invented in the 70's by Jonassen and Dahl, and the present paper greatly extends the old analytic results, by using a natural combinatorial decomposition of the objects and by solving the associated system of differential equations.

On the other hand, the last decade has witnessed a blossoming of works that focus on the digital tree or "trie" model as this model underlies dynamic hashing schemes, distributed leader election algorithms, communication protocols, data compression, the organization of dictionaries, etc. Knessl studies the limiting distribution of the depth in tries. He obtains a new limiting distribution that interpolates the well known extreme (symmetric model) and normal (asymmetric model) distributions.

Probabilistic methods are used by D. Aldous to study Metropolis-type optimization algorithm, while Devroye, Mücke and Zhu investigate Delaunay triangulation of random points and show that a simple heuristic takes expected time $O(n^{1/3})$. Dynamic reorganization of a linear list is studied by probabilistic method by Shachnai and Hofri.

Finally, *analytic methods* are used by Coffman, Flato, Jelenković and Poonen, to explore packing random intervals on line and also by Hofri and Jacquet who study saddle points in random matrices (Ex. 1.3.2-12 of Knuth's *The Art of Computer Programming*). This open (M-47) problem was solved independently by Hofri and Jacquet, and the initiative of Don Knuth resulted in the present paper that contains the combined efforts.

In terms of methodology, this issue therefore illustrates the two major paradigms of analysis of algorithms. Objects of low structural complexity (for instance, strings, trees, permutations) can often be described exactly by analytic models, where generating functions, integral transforms, and complex asymptotics are central tools. Objects of higher structural complexity (e.g., graphs) are best approached by probabilistic inequalities, since an analytic model is often out of reach. There is however no firm dividing line, as exemplified by many papers in this volume: on the one hand, the analytic method is often capable of providing information on probability distributions of interest; on the other hand, the probabilistic method is often applicable to problems of low computational complexity.

We are very grateful to C. K. Wong, the Editor–in–Chief of *Algorithmica* for having provided the opportunity to prepare this special issue. The papers in this issue examine problems where randomness is central to algorithmic performance. We hope that the readers will find this volume interesting as well as useful in their scientific endeavours.

Helmut Prodinger
Wojciech Szpankowski

Guest Editors