Threshold phenomena in an "easy" class of random lattices and efficient reduction algorithms

Ali Akhavi

In 70's Donaldson proved a conjecture of Zassenhaus and Ford: "Lattices of the uniform model and with a fixed dimension are almost surely reduced (in the sense of Minkowski), as the dimension of the ambient space tends to infinity." Then he deduces some more efficient heuristic for finding a Minkowski reduced basis in a lattice. Our talk shows (in a more precise sense) an analogous phenomenon, for the Lovasz (LLL) reduction. In particular, our method exhibits some sharp threshold phenomenon concerning lattices of the uniform model.

Then we present and analyze two new lattice reduction algorithms. These algorithms, called the Schmidt reduction and the Gram reduction, are obtained by relaxing some of the constraints of the classical LLL algorithm. Analyzing the worst case behavior and the average case behavior in a tractable model, proves that the new algorithms still produce "good" reduced basis while requiring fewer steps on average. In addition, these results (about the relative behavior of the different reduction algorithms) are confirmed by empirical tests on random lattices coming from applications.

Random Assignment Problem and Asymptotic Essential Uniqueness in Optimization over Random Data

David J. Aldous

University of California, Berkeley

We discuss a general question and a specific question. To motivate the general question, consider the TSP on n random points in the square. The optimal tour is trivially a.s. unique. But there's a deeper uniqueness question. One can perturb the optimal tour by changing a proportion δ of edges used, to create a sub-optimal tour whose length is $(1+\varepsilon)$ times optimal. Is the converse true? That is, given a sub-optimal tour whose length is $(1+\varepsilon)$ times optimal, is it necessarily a perturbation of the optimal tour which changes a proportion $\leq \delta$ of edges, where $\delta(\varepsilon) \to 0$ as $\varepsilon \to 0$? This *asymptotic essential uniqueness* (AEU) question can be posed for most continuous-valued optimization-over-random-data problems. It is easy to see, for instance, that AEU holds for the minimal spanning tree problem but not for a continuous version of maximal clique in random graph. Resolving AEU for the TSP is a major challenge for the 21st century. More approachable today is the *random assignment* (RA) or *bipartite matching* problem which studies

$$A_n = \min_{\pi} \sum_{i=1}^{n} c_{i,\pi(i)}$$

for a $n \times n$ matrix (c_{ij}) with independent exponential(1) entries. It has long been conjectured [4] that $EA_n \to \pi^2/6$ and indeed that

$$(*)\ EA_n = \sum_{i=1}^{n} i^{-2}.$$

See [2] for recent rigorous work. Perhaps (*) can be proved by clever elementary methods (a recent argument of this type [3] seems incomplete). A more sophisticated approach [1] to the asymptotic conjecture identifies the limit $\lim_n EA_n$ as the value of a minimum matching problem on a certain random weighted infinite tree. I will describe ongoing work which exhibits a particular limit matching with renormalized mean $\pi^2/6$. This suggests an approach to proving both the $\pi^2/6$ limit conjecture and the AEU property for the RA problem, though we have not carried through the details.

References

[1] D.J. Aldous. Asymptotics in the random assignment problem. *Probab. Th. Rel. Fields*, 93:507–534, 1992.

[2] D. Coppersmith and G. Sorkin. Constructive bounds and exact expectations for the random assignment problem. *Random Structures and Algorithms*, 15:113–144, 1999.

[3] V. S. Dotsenko. Exact solution of the random bipartite matching model. *J. Phys. A*, 33:2015–2030, 2000.

[4] M. Mézard and G. Parisi. On the solution of the random link matching problem. *J. Physique*, 48:1451–1459, 1987.

The Kernel Method

Cyril Banderier

Algorithms Project, INRIA

The kernel method belongs to mathematical folklore since the 1970's; e.g., it has been used by combinatorialists (lattice paths problems [7], graph enumeration [4]) and probabilists [6]. There are also some more recent works which make a deep use of it [3, 1, 2].

In order to solve

$$K(z,u)F(z,u) = G(z,u) + U(z)$$

(K and G are known whereas F and U are unknown), the method consists in forcing the left hand-side to be zero by coupling z and u so that the coefficient of the (unknown) quantity $F(z,u)$ is zero. This constraint defines u as one of the branches of an algebraic function of z. If enough branches can be substituted analytically, then enough relations will be generated so that one can solve for the (unknown) quantities gathered in $U(z)$, (usually a linear combination of $F(z,1)$ and $\partial_u^j F(z,0)$). These unknown functions are then obtained as algebraic functions. From there, an expression for $F(z,u)$ also results in the form of a bivariate algebraic function.

We give applications to generating trees (a infinite model of rewriting rules used for random generation), walks over $\mathbb{N}$ (e.g. with a set of jumps $(+2,-5,+7)$), paths under a line of a rational slope, chess knight walks, generalized Dyck words, average number of Dyck factors of a Dyck word. The kernel method can also be exploited in order to get asymptotics.

References

[1] Cyril Banderier and Mireille Bousquet-Mélou and Alain Denise and Philippe Flajolet and Daniéle Gardy and Dominique Gouyou-Beauchamps. On Generating Functions of Generating Trees. *FPSAC'99*, 1999.

[2] M. Petkovšek and M. Bousquet-Mélou, Linear recurrences with constant coefficients: the multivariate *c*. *Discrete Mathematics*, to appear, 2000.

[3] A. Odlyzko, F. Chung, R. Graham and J. Morrison. Pebbling a chessboard. *American Mathematical Monthly*, 103, 1995.

[4] R. Cori and J. Richard. Énumération des graphes planaires á l'aide des séries formelles en variables non commutatives. *Discrete Mathematics*, 2:115–162, 1972.

[5] S. Dulucq, S. Gire and O. Guibert. A combinatorial proof of J. West's conjecture. *Discrete Mathematics*, 187:71–96, 1998.

[6] G. Fayolle and R. Iasnogrodski. Solutions of functional equations arising in the analysis in the analysis of two-server queueing models. Performance of computer systems, 289–303, North-Holland, 1979.

[7] D. E. Knuth. The Art of Computer Programming. Vol. 1: Fundamental Algorithms. Exercises 4 and 11, section 2.2.1, Addison-Wesley, 1968.

The height and width of simple trees

P. Chassaing, J. F. Marckert, M. Yor

Laboratoire e Mathématiques, Institut Elie Cartan
BP 239 54 506 Vandoeuvre les Nancy Cedex

The limit law of the couple height-width for simple trees can be seen as a consequence of deep results of Aldous, Drmota and Gittenberger, and Jeulin. We plan to give an elementary proof in the case of binary trees, using an idea of Odlyzko.

Let $Z_i(t)$ denote the number of nodes at distance i from the root of a rooted tree t. The *profile* of the tree t is the sequence $(Z_i(t))_{i\geq 0}$. The *width* $w(t)$ and *height* $h(t)$ of the tree t are defined by:

$$\begin{aligned} w(t) &= \max_i \{Z_i(t)\}, \\ h(t) &= \max\{i|Z_i(t) > 0\}. \end{aligned}$$

Let $T_B^{(n)}$ denote the set of *binary trees* with n leaves ($2n-1$ nodes), endowed with the uniform probability, and let $H_B^{(n)}$ (resp. $W_B^{(n)}$) be the restriction of h (resp. w) to $T_B^{(n)}$. One can also see $H_B^{(n)}$ and $W_B^{(n)}$ as the height and width of a Galton-Watson tree with offspring distribution 0 or 2 with probability 1/2, conditioned to have total progeny $2n-1$ (see [1, pp. 27-28]). Then, the limit law of the height [11, 13] and of the width [6, 10, 14] are given by:

$$\begin{aligned} \frac{H_B^{(n)}}{\sqrt{2n}} &\xrightarrow[n\to+\infty]{law} 2V, \\ \frac{W_B^{(n)}}{\sqrt{2n}} &\xrightarrow[n\to+\infty]{law} V, \end{aligned}$$

where:

$$\Pr(V \leq x) = \sum_{-\infty<k<+\infty} (1-4k^2x^2)\exp\left(-2k^2x^2\right).$$

Connections between the distribution of V on one hand, the Brownian motion and Jacobi's Theta function on the other hand, are discussed in [5, 7, 12]. For instance, let $(e(s))_{0\leq s\leq 1}$ denote a standard normalized Brownian excursion. Then the random variables

$$(H,W)=\Big(\int_0^1\frac{ds}{e(s)}\ ,\ \max_{0\leq s\leq 1}e(s)\Big),$$

satisfy

$$V\overset{law}{=}W\overset{law}{=}\frac{H}{2}. \tag{1}$$

The first identity is due to Chung [7], the second was first stressed in [4, p. 69]. The aim of this paper is to give a simple proof of the following theorem:

Theorem 1.

$$\left(\frac{H_B^{(n)}}{\sqrt{2n}},\frac{W_B^{(n)}}{\sqrt{2n}}\right)\xrightarrow[n\to+\infty]{law}(H,W).$$

Note that the obvious negative correlation between height and width of a tree with given size n, is reflected in the dependence between $\int_0^1\frac{ds}{e(s)}$ and $\max_{0\leq s\leq 1}e(s)$. Previous results [11, 13] about height and width of simple trees belongs to the foundations of computer science. Surprisingly, Theorem 1 does not seem to be stated anywhere, though it can be deduced easily from deep results of Aldous on one hand (about the continuum random tree [1, 2]) and on the other hand of Drmota & Gittenberger [9], using a clever idea due to Aldous [3, Th. 3] again. We feel that this consequence of [3, Th. 3] deserves to be pointed out, and that an 'elementary' and direct proof should be welcome.

The joint law of (H,W) has been investigated recently by Catherine Donati-Martin [8]. Let $\Phi(\alpha,\gamma,z)$ denote the confluent hypergeometric function, defined, for $|z|<+\infty,\gamma\neq 0,-1,-2,\cdots$, by:

$$\Phi(\alpha,\gamma,z)=\sum_{k=0}^{+\infty}\frac{(\alpha)_k\,z^k}{(\gamma)_k\,k!}$$

where $(\lambda)_k=(\lambda)(\lambda+1)\dots(\lambda+k-1)$. With the help of the agreement formula, Catherine Donati-Martin obtains the following results:

Theorem 2. *For $\lambda\geq 0,\alpha\geq 0$,*

$$E\Big(W\exp(-\frac{\lambda^2}{2W^2}-\frac{\alpha^2H}{2W})\Big)=\sqrt{\frac{\pi}{2}}\frac{\exp(2\lambda)}{\Phi^2(1+\alpha^2/(2\lambda),2,2\lambda)}.$$

As a consequence, for $Re(s)>1, Re(t)<0$ and $Re(s+t)>1$:

$$E(W^sH^t)=\sqrt{\frac{\pi}{2}}\frac{2^{\frac{5+t-s}{2}}}{\Gamma(-t)\Gamma(\frac{s+t-1}{2})}\int_0^{+\infty}\int_0^{+\infty}\frac{\lambda^{s+t-2}\alpha^{-(1+2t)}\exp(2\lambda)}{\Phi^2(1+\alpha^2/(2\lambda),2,2\lambda)}\,d\alpha\,d\lambda.$$

References

[1] D. Aldous, (1991) *The continuum random tree II: An overview,* Stochastic analysis, Proc. Symp., Durham. UK 1990, Lond. Math. Soc. Lect. Note Ser. 167, 23-70.

[2] D. Aldous, (1993) *The continuum random tree III,* Ann. of Probab. 21, No.1, 248-289.

[3] D. Aldous, (1998) *Brownian excursion conditionned on its local time*, Elect. Comm. in Probab., **3** , 79-90.

[4] P. Biane, M. Yor, (1987) *Valeurs principales associées aux temps locaux browniens*, Bull. Sci. Maths 111, 23-101.

[5] P. Biane, J. Pitman, M. Yor, (1999) *Probability laws related to the Jacobi theta and Riemann zeta functions, and Brownian excursions*, (See http://www.stat.berkeley.edu/~pitman/)

[6] P. Chassaing, J.F. Marckert, (1999) *Parking functions, empirical processes and the width of rooted labeled trees*, preprint Elie Cartan, Université Nancy I.

[7] K.L. Chung, (1976) *Excursions in Brownian motion*, Ark. för Math., **14**, 155-177.
[8] C. Donati-Martin, (1999) *Some remarks about the identity in law for the Bessel bridge* $\int_0^1 \frac{ds}{r(s)} \stackrel{(law)}{=} 2 \sup_{s \leq 1} r(s)$, unpublished.
[9] M. Drmota, B. Gittenberger, (1997) *On the profile of random trees*, Random Structures Algorithms 10, no. 4, 421–451.
[10] M. Drmota, B. Gittenberger, (2000) *The width of Galton-Watson trees*, preprint.
[11] P. Flajolet, A. Odlyzko, (1982) *The average height of binary trees and other simple trees*, J. Comp. and Sys. Sci., Vol. 25, No.2.
[12] G. Louchard, (1984) *Kac's formula, Levy's local time and Brownian excursion*, J. Appl. Prob. **21** , 479-499.
[13] A. Renyi, G. Szekeres, (1967) *On the height of trees*, J. Aust. Math. Soc. 7, 497-507.
[14] L. Takács, (1993) *Limit distributions for queues and random rooted trees*, J. Appl. Math. Stoch. Ana., **6** , No.3, p.189 - 216.

A Martingale Method for the Profile of Binary Search Trees

Brigitte Chauvin

Department of Mathematics
Batiment Fermat
Universite de Versailles

Among algorithms to sort data, "quicksort" has been largely studied, especially because data are nodes of a so-called binary search tree. Sorting n data gives a labeled tree with n nodes. When n is large, the height of the tree, its width, the number of external and internal nodes at each level, etc ... are of interest. Considering these trees as a Markov process indexed by n, and the above quantities as random variables, probabilistic methods can be applied. Exhibiting a martingale, proving large deviations results on the distribution of internal nodes, getting an almost sure central limit theorem, allow us to get new asymptotic, especially on the width of binary search trees.

Limit distributions of the space requirements of random m-ary search trees

Hua-Huai, Félix, Chern

Department of Mathematics and Science Education
Taipei Municipal Teachers College

We use a uniform approach based on differential equations and method of moments to show that the limit distribution of the space requirements of an m-ary search tree built from iid continuous random variables is normal for $3 \leq m \leq 26$ and does not exist for $m \geq 27$. The tools we introduce are general and applicable to other problems for which the bivariate generating function in question satisfies a differential equation, linear or nonlinear. Typical examples include: rotations in $2k+1$-fringe-balanced BSTs, elastic buckets in multiway search trees, space of bucket recursive trees, cost of greedy matching heuristics on trees, etc.
(This is a joint work with H.K. Hwang)

On Greedy Reservation Algorithms

Y. Baryshnikov[1], E. Coffman[2], and P. Jelenković[2]

1. Mathematics Dept., University of Osnabrück, Osnabrück, Germany
2. Electrical Engineering Dept., Columbia University, New York, NY.

1. Introduction

In the model to be studied, a long interval $[0, X]$ is given, with X an integer; the interval is divided into X slots each of unit length. A *claim* is a short interval consisting of L consecutive slots. Thus, there are $R = X - L + 1$ possible positions where such claims can be placed. The claims arrive at rate 1 at each possible position (so the total arrival rate is R).

There are k possible channels (or lanes) where the arriving claims can be placed. Each claim belongs to a channel; overlapping within a channel is prohibited. Let $n(i)$ be the number of claims covering the i-th slot, $1 \leq i \leq X$. Then, clearly, for any configuration which can be placed on k channels, one has $n(i) \leq k$ for all i. A simple (but crucial) proposition shows that the converse is also true: if $n \leq k$ on $[0, X]$, then the configuration of claims can be placed on k channels. We study the *greedy* acceptance policy: if a claim arrives such that the total consistency condition $n(\cdot) \leq k$ is nowhere violated, the claim is accepted, otherwise rejected.

We investigate two types of allocation mechanisms: static, where a claim once accepted remains in the configuration forever, and dynamic, where any accepted claim can depart. We will assume that these departures happen independently, in Markov fashion, and with rate p for each currently active claim. As we will show, the static and dynamic systems exhibit remarkable differences in their limiting behavior. In particular, the average wasted space per unit length, $w(X)/X$ in the dynamic model converges, as $X \to \infty$ to a certain value $v_{k,L}(p)$ whose asymptotics we study in the special cases $L = 2$ or $k = 1$. In particular, we find that

$v(p) \to 0$ as $p \to 0$. By contrast, in the static model the average wasted space $w(X)/X$ has a positive limit as $X \to \infty$, whatever $k \geq 1$ or $L \geq 2$ are. Here is a sample of our results.

2. Dynamic model

Proposition 1. *For any k, L, the average wasted space per unit time converges to a function $v_{k,L(p)}$ analytic for $p \in (0,1)$. The limit $v(0+) = 0$ for all k, L.*

For $k = 1$ (that is, for the single channel case) we find explicitly $v(p) = p^{1/L}/L + O(p^{2/L})$ for small p. In fact, a similar analysis can be carried out for a mixture of claims of different lengths $L_1, L_2, \ldots, L_m$ arriving with different rates $\lambda_i, \sum_i \lambda_i = 1$.

For $L = 2$ (all claims have length 2) we can analyze explicitly the situation with arbitrary k. Interestingly, the parity of k plays an important role.

Proposition 2. *If $k+1 = 2l$ then $v(p) = \Theta(p^{1/2l})$ while for $k+1 = 2l+1$, $v(p) = \Theta(p^{1/(l+1)})$.*

We use the transfer matrix method in this section.

3. Static model

We consider the final configuration of claims on k channels where no further claim can be placed without violating the condition $n(\cdot) \leq k$. The wasted space function $i \mapsto s(i) = k - n(i)$ is of primary interest to us here.

An easy proposition shows that for any i the random value $s(i)$ converges, as $X \to \infty$ to a limiting r.v. $s_\infty(i)$. Moreover, $s_\infty(i) \to s_\infty$ as $i \to \infty$, where s_∞ can also be defined in terms of the claim placement process on the infinite line. Clearly, the average wasted space for the static process is just the expected value of s_∞ We concentrate on the asymptotics of s_∞ for large k. Our main result here is

Proposition 3. *For fixed L, the r.v. s_∞ scales as $\sqrt{k}$, that is $s_\infty/\sqrt{k}$ converges to a nondegenerate r.v. as $k \to \infty$.*

The height and profile of binary and m-ary search trees

Michael Drmota

Department of Discrete Mathematics
Technical Universiyu of Vienna

Binary search trees are widely used to store (totally orderer) data, and many parameters have been discussed in the literature. Usually it is assumed that every permutation of $\{1, 2, \ldots, n\}$ is equally likely and any parameter of binary search trees may be considered as a random variable. (An alternative way of looking at is a Markov chain of trees which describes the evolution of a binary search tree.)

There are many interesting parameters, e.g. let $U_k(n)$ denote the number of (external) nodes at level k. The first part of this lecture is devoted to the profile, i.e. to the sequence $(U_k(n))_{k \geq 0}$.

It will be shown that $U_k(n)$ is concentrated around the level $k \sim 2\log n$ (compare with [1]), more precisely, as $n \to \infty$ we almost surely have

$$\frac{U_k(n)}{n/\sqrt{4\pi \log n}} = e^{-\frac{(k-2\log n)^2}{4\log n}} + O\left(\frac{1}{\sqrt{\log n}}\right),$$

uniformly for all $k \geq 0$. Especially if follows that the width satisfies, as $n \to \infty$ almost surely

$$\frac{\max\limits_{k\geq 0} U(n)}{n/\sqrt{4\pi \log n}} = 1 + O\left(\frac{1}{\sqrt{\log n}}\right).$$

Another interesting parameter is the height H_n of binary search trees. It has been shown by Devorye [2] that the expected value of H_n is given by

$$\mathbf{E}H_n \sim c \log n,$$

where $c = 4.31\cdots$ is the solution of the equation $\left(\frac{2e}{c}\right)^c = e$. It is interesting that the variance of H_n is bounded. This fact has been a long standing conjecture has been proved independently by Reed [5] and (a little bit later) by Drmota [3].

In the second part of this lecture we outline a proof of a corresponing property for m-ary search trees (compare with [4]). We also present a conjectural relation for the distribution of the height of m-are search trees. For example, if the limit

$$\lim_{h\to\infty} \frac{y_{h+1}(1)}{y_h(1)}$$

exists, where $y_h(x)$ are polynomials defined by $y_0(x) \equiv 1$, $y'_{h+1}(x) = y_h(x)^2$ $(y_j(0) = 1)$, then there exists a function $\Psi(y)$ such that

$$\mathbf{P}[H_n \leq h] \sim \Psi(n/y_h(1))$$

uniformly for h with $\log n - \log y_h(1) = O(1)$, as $n \to \infty$. It is also very likely that there exist continuous periodic fuctions Δ_1 and Δ_2 with period 1 such that

$$\mathbf{E}H_n = c\log n - \frac{3c}{2(c-1)}\log\log n + \Delta_1\left(c\log n - \frac{3c}{2(c-1)}\log\log n\right) + o(1)$$

and

$$\mathbf{V}H_n = \Delta_2\left(c\log n - \frac{3c}{2(c-1)}\log\log n\right) + o(1)$$

as $n \to \infty$.

References

[1] B. Chauvin, M. Drmota, and J. Jabbour-Hattab, The Profile of Binary Search Trees, *Ann. Appl. Prob.*, submitted.

[2] L. Devroye, A note on the height of binary search trees, *J. Assoc. Comput. Mach.* **33** (1986), 489–498.

[3] M. Drmota, The Variance of the Height of Binary Search Trees, *Theoret. Comput. Sci.*, submitted.

[4] M. Drmota, An Analytic Approach to the Height of Binary Search Trees. II., *preprint.*

[5] B. Reed, How tall is a tree, *preprint.*

An in-depth probabilistic analysis of quicksort

James Allen Fill

The John Hopkins University

Quicksort, the standard sorting procedure in `Unix` systems, is probably the most widely used general-purpose sorting algorithm, and has been the subject of intense analytic and numerical study. I will present the most in-depth probabilistic analysis of the running time of Quicksort to date.

In particular, I will discuss how to extend the *contraction method* of Uwe Rösler and (independently) Ludger Rüschendorf to obtain Berry–Esséen-type results about the rate of convergence to its limiting distribution of the (suitably centered and scaled) number of comparisons required to sort a file of n keys; Wasserstein (or Mallows) and Kolmogorov–Smirnov metrics both play a role in this regard.

The limiting distribution itself (call it F) is a bit nebulous: it is known only as the unique fixed point with finite variance of a certain distributional identity. I will show how to use Fourier analysis to prove that F has an everywhere positive and infinitely differentiable density f, and that each derivative $f^{(k)}$ enjoys superpolynomial decay at $\pm\infty$. In particular, each $f^{(k)}$ is bounded.

I will also discuss how to obtain explicit bounds on the error in (rapid) numerical approximation of F and its derivatives.

If time permits, I will also discuss perfect simulation from F, and/or discuss large deviations, and/or explain how the same sort of program can be carried out for other divide-and-conquer recurrences.

(This is joint work with Svante Janson.)

Some Airy phenomena in analytic combinatorics

Philippe Flajolet

INRIA-Rocquencourt 78150 (France)
Philippe.Flajolet@inria.fr

The Airy function is a classical special function of analysis originally invented to describe some phenomena of optics and whose definition involves the exponential of a cubic. This function has made scattered appearances in the analysis of random discrete structures—e.g., in Louchard's analysis of area under an excursion, in Takács's analysis of path length in trees, in linear probing hashing by Poblete, Viola, *et al.*—giving rise to the "area Airy distribution".

We shall study in detail the occurrence of an Airy law of another brand, the "map Airy distribution", which is the limit distribution governing the size of the largest multiply connected component (the "core") in random planar maps of various types. The problem can be approached by means of a modified form of singularity analysis but also by suitable adjustments of of the saddle point method. Indeed, the presence of nearby and "coalescing" saddle points can be seen to induce a map Airy law under fairly general conditions. Intuitively, two quadratic behaviours coalesce into a cubic behaviour from which the Airy function originates.

We claim this analytic-geometric situation of coalescence to be of some degree of universality. For instance it accounts well for certain phenomena observed in random forests of unrooted trees and in early stages of the random graph evolution under the Erdős-Rényi model (as investigated by Kolchin, Janson, Knuth, Łuczak, Pittel, and others). Related "phase transitions" like in hashing tables (under the linear probing strategy) exhibit an area Airy law whose treatment might be amenable to coalescing saddle point techniques as previously discussed.

(Based on joint work in progress with Cyril Banderier, Gilles Schaeffer, and Michèle Soria and our ICALP'2000 paper.)

Dynamical urn model revisited

Daniele Gardy

Laboratoire PRISM

University of Versailles Saint-Quentin

The random allocation model of a sequence of urns receiving a (specified or random) number of balls has been much studied. When the number of balls is known, the number of urns satisfying some condition generally follows a Gaussian limiting distribution. When the balls are inserted one at a time, one can also often prove that

the limiting process is Gaussian (see e.g. our former work with Drmota and Gittenberger).

We consider here models that allow for deletion of balls. Such cases occur for instance in dynamic databases, and we show how some simple situations can be treated with analytic tools. We then consider whether a systematic approach that would parallel the one for the models restricted to insertions is suitable for insertions/deletions models.

The Width of Galton-Watson Trees

Bernhard Gittenberger

Technical University of Vienna

It is proved that the moments of the width of Galton-Watson trees with offspring variance σ are asymptotically given by $(\sigma\sqrt{n})^p m_p$ where m_p are the moments of the maximum of the local time of a standard scaled Brownian excursion. This is done by combining a weak limit theorem and a tightness estimate. The method is quite general and we state some further applications.

New Upper and Lower Bounds on the Channel Capacity of Read/Write Isolated Memory

Mordecai J. Golin, Xuerong Yong, Yuanping Zhang and Li Sheng

Department of Computer Science
Hong Kong University of Science and Technology

A serial binary $(0,1)$ memory is *read isolated* if no two consecutive positions in the memory may both store 1's. This restriction occurs, for example, in some codes used in magnetic recording and optical recording. A serial binary $(0,1)$ memory that undergoes rewriting is *write isolated* if it satisfies the restriction that no two consecutive positions in the memory can be changed during one rewriting phase. Such restrictions have arisen, for example in the contexts of asymmetric error-correcting ternary codes and of rewritable optical discs

A *read/write isolated memory* (RWIM) is a binary, linearly ordered, rewritable storage medium that obeys both the read and write restrictions In this talk we refine upper and lower bounds for the channel capacity of such a medium. This problem was originally examined by Cohn who proved that C, the channel capacity of the memory, in bits per symbol per rewrite, satisfies

$$0.50913\ldots \leq C \leq 0.56029\ldots.$$

We show how to model the problem as a constrained two-dimensional binary matrix problem and then modify recent transfer matrix techniques for dealing with such matrices to derive improved bounds of

$$0.53500\ldots \leq C \leq 0.55209\ldots.$$

On the distribution of the number of distinct parts size in a random integer composition

Paweł Hitczenko

MCS Department, Drexel University, Philadelphia, PA 19104

In this talk we discuss the asymptotic properties of the distribution function of the number of distinct part sizes in a randomly chosen composition of an integer n. The main focus of earlier work on the subject concentrated on the asymptotics of the expected value of the number of distinct part sizes. Hwang and Yeh [3] used generating function approach and, among other things, derived explicit formulas for the asymptotics of the expectation and the variance. Hwang and Yeh raised the question about the asymptotic behavior of the distribution function of the number of distinct part sizes and suggested that it should be close to a normal random variable.

We will address that question; we show that the limiting distribution is discrete and is close to that of a "discretizied" and randomized extreme value distribution.

Our argument is a mixture of probabilistic and analytical methods. We first argue, using the method developed in [2], that a joint distribution of sizes of parts in a randomly chosen composition of n is close to that of a *fixed* number (roughly $n/2$) of independent, identically distributed (i.i.d.) geometric random variables with parameter $1/2$. We then examine the distribution of the distinct values taken on by $n/2$ i.i.d. geometric random variables. This is accomplished by following arguments used in the analysis of a leader election algorithm [1], [5]. We establish a recurrence relation for the number of distinct values taken by such a sequence. We then use a poissonization argument and the analytical depoissonization technique described in [4] to derive the result for the fixed number of geometric random variables.

1. Fill, J. A., Mahmoud, H., Szpankowski, W., *On the distribution for the duration of a randomized leader election algorithm*, Ann. Appl. Probab. **1**(1996), 1260 – 1283.
2. Hitczenko, P., Savage, C. D. *On the multiplicity of parts in a random composition of a large integer*, preprint, 1999.
3. Hwang, H. K., Yeh, Y. N., *Measures of distinctness for random partitions and compositions of an integer*, Adv. Appl. Math. **19**(1997), 378 – 414.
4. Jacquet, S., Szpankowski, W., *Analytical de–Poissonization and its applications*, Theoret. Comput. Sc. **201** (1998), 1 – 62.
5. Janson, S., Szpankowski, W., *Analysis of the asymmetric leader election algorithm*, Electr. J. Combin., **4** (1997), no. 1, research paper 17, 16pp. **201** (1998), 1–62.

An asymptotic theory for recurrence relations based on minimization and maximization

Hsien-Kuei Hwang

Institute of Statistical Science
Academia Sinica, Taipei

An asymptotic theory is developed for the sequence $f(n)$ defined recursively by $f(n) = \min_{1\le j<n}\{f(j)+f(n-j)\}+g(n)$, when the asymptotic behavior of $g(n)$ is known. Algorithmic applications are discussed. Our tools are general enough and applicable to another sequence $F(n) = \max_{1\le j<n}\{F(j)+F(n-j)+\min\{g(j),g(n-j)\}\}$, also frequently encountered in divide-and-conquer problems.

Four methods and one theorem - another look at hashing with linear probing

Svante Janson
Department of Mathematics
Uppsala University

I will discuss four different methods and their applications, for different ranges of the parameters, to the study of hashing with linear probing. Together they yield complete results on the asymptotic distribution of both the total displacement, or equivalently the average cost of a successful search, and of the average cost of an unsuccessful search. Asymptotics of moments are treated too.

Improved Approximation of MAX-CUT on Graphs of Bounded Degree

Marek Karpinski
University of Bonn

We design a new approximation algorithm for the MAX-CUT problem based on semidefinite relaxation of MAX-CUT, and random hyperplane rounding technique combined with a new local improvement placement method. Let $\alpha \simeq 0.87856$ denote the approximation factor of the best currently known approximation algorithm of Goemans and Williamson for the MAX-CUT problem on general graphs. We show that for graphs of degree at most Δ, our algorithm achieves an approximation factor of at least $\alpha + \epsilon$, where $\epsilon > 0$ is a constant depending only on Δ.
In particular, our algorithm achieves an approximation factor of at least 0.921 for graphs of maximal degree 3, and an approximation factor of at least 0.924 for 3-regular graphs.
(Joint work with U. Feige and M. Langberg.)

On the Expected Number of Leftist Nodes in Simply Generated Trees

Rainer Kemp

Johann Wolfgang Goethe-Universität
Fachbereich Informatik
D - 60054 Frankfurt am Main, Germany

A node x appearing in an ordered tree is said to be a leftist node if in the subtree with root x, the leaf nearest to the root is the leftmost leaf of that subtree. Assuming that all trees belonging to a family of simply generated trees with a specified number of nodes and leaves are equally likely, we present a general approach to the computation of the average number of leftist nodes appearing in such a tree.

If the number of nodes in a simply generated tree with the enumerator $E(z) = z\,\Theta(E(z))$ are specified only, that approach yields exact asymptotical equivalents to the average number $\mathbb{E}[\,\mathsf{X}_n\,]$ of leftist nodes, to the average number $\mathbb{E}[\,\mathsf{Y}_n\,]$ of internal leftist nodes and to the variances $\mathbf{VAR}(\mathsf{X}_n)$ and $\mathbf{VAR}(\mathsf{Y}_n)$ of the corresponding random variables. We find

$$\begin{aligned} \mathbb{E}[\,\mathsf{X}_n\,] &= \mathsf{x}_1\,n + \mathsf{x}_2 + \mathcal{O}\left(\frac{1}{n}\right), \ \mathbb{E}[\,\mathsf{Y}_n\,] \ = \ \mathsf{y}_1\,n + \mathsf{y}_2 + \mathcal{O}\left(\frac{1}{n}\right), \\ \mathbf{VAR}(\mathsf{X}_n) &= \mathsf{v}_\mathsf{x}\,n + \mathcal{O}(1) \ \mathbf{VAR}(\mathsf{Y}_n) \ = \ \mathsf{v}_\mathsf{y}\,n + \mathcal{O}(1), \end{aligned}$$

where

$$\begin{aligned} \mathsf{x}_1 &:= \frac{\rho\xi}{\tau^2},\ \mathsf{x}_2 := \frac{\xi}{\tau^2}\left(\frac{\rho}{2} - \frac{\Theta^{(3)}(\tau)}{[\,\Theta''(\tau)\,]^2}\right) + \frac{\psi}{2\,\tau\,\Theta''(\tau)}\left(\frac{1}{\tau} + \frac{\Theta^{(3)}(\tau)}{\Theta''(\tau)}\right) - \frac{\zeta}{2\,\tau\,\Theta''(\tau)} + \frac{\rho^2\,\eta}{2\,\tau^2}, \\ \mathsf{y}_1 &:= \mathsf{x}_1 - \frac{\rho}{\tau},\ \mathsf{y}_2 := \mathsf{x}_2 - \frac{1}{2\,\tau}\left(\rho + \frac{1}{\tau\,\Theta''(\tau)} - \frac{\Theta^{(3)}(\tau)}{[\,\Theta''(\tau)\,]^2}\right), \\ \mathsf{v}_\mathsf{x} &:= \frac{\rho}{\tau^2}\left(\gamma + \xi + \frac{\xi^2}{\tau^2}\left[\rho - \frac{4}{\tau\,\Theta''(\tau)}\right] - \frac{2\,\xi\,\psi}{\tau}\left[\rho - \frac{2}{\tau\,\Theta''(\tau)}\right] - \frac{\psi^2}{\tau\,\Theta''(\tau)} - \frac{2\,\rho^2\,\xi\,\eta}{\tau^2}\right), \\ \mathsf{v}_\mathsf{y} &:= \mathsf{v}_\mathsf{x} + \frac{\rho}{\tau}\left(1 - \frac{\rho}{\tau} - \frac{1}{\tau^2\,\Theta''(\tau)} - \frac{2\,\xi}{\tau}\left[1 - \frac{\rho}{\tau} - \frac{2}{\tau^2\,\Theta''(\tau)}\right] - \frac{2\,\psi}{\tau^2\,\Theta''(\tau)} + \frac{2\,\rho^2\,\eta}{\tau^2}\right). \end{aligned}$$

Here, $\tau := \min\{x \in \mathbb{R}_+ \,|\, x\,\Theta'(x) = \Theta(x)\}$, $\rho := \frac{\tau}{\Theta(\tau)}$ and $\circ := \sum_{\lambda\geq 1} \circ_\lambda$, $\circ \in \{\xi, \psi, \zeta, \eta\}$, are convergent series with

(a) $\xi_{\lambda+1} = \rho\,[\,\Theta(\xi_\lambda) - 1\,]$, (b) $\psi_{\lambda+1} = \rho\,\Theta'(\xi_\lambda)\,\psi_\lambda$,

(c) $\zeta_{\lambda+1} = \rho\,\Theta''(\xi_\lambda)\,\psi_\lambda^2 + \rho\,\Theta'(\xi_\lambda)\zeta_\lambda$, (d) $\eta_{\lambda+1} = \rho^{-1}\,\xi_{\lambda+1} + \rho\,\Theta'(\xi_\lambda)\,\eta_\lambda$

for $\lambda \geq 1$, where $\xi_1 = \tau$, $\psi_1 = 1$, $\zeta_1 = 0$ and $\eta_1 = 0$.

We illustrate these general results by applying them to various families of simply generated trees with specified nodes, such as t-ary trees, unary–binary trees, unbalanced 2,3–trees, ordered trees, ordered trees without unary nodes and ordered trees with even node degrees. Furthermore, special classes of ordered trees with specified nodes and leaves are considered, too.

Limit Laws for Heights in Digital Trees

Charles Knessl
Dept. Mathematics, Statistics & Computer Science
University of Illinois at Chicago
Chicago, Illinois 60607-7045
U.S.A
knessl@uic.edu

Wojciech Szpankowski
Department of Computer Science
Purdue University
W. Lafayette, IN 47907
U.S.A.
spa@cs.purdue.edu

We study the height of *digital trees* (i.e., tries, PATRICIA trie and digital search trees) built from n random strings generated by an unbiased memoryless source (i.e., all symbols are equally likely). In the course of our analysis, we shall identify several natural regions of the *height distribution* and characterize them asymptotically for large n. We find that:

- For *tries*, in the region where most of the probability mass is concentrated, the asymptotic distribution of the height is of extreme value type (i.e., double exponential distribution).
- For *PATRICIA tries* the height behaves quite differently in this region: It exhibits an exponential of a Gaussian distribution (with an *oscillating term*) around the most probable value $k_1 = \lfloor \log_2 n + \sqrt{2 \log_2 n} - \frac{3}{2} \rfloor + 1$. In fact, the asymptotic distribution of PATRICIA height concentrates on one or two points. For most n all the mass is concentrated at k_1, however, there exist subsequences of n such that the mass is on the two points $k_1 - 1$ and k_1, or k_1 and $k_1 + 1$.
- For *digital search trees* in the region where most of the probability mass is concentrated, the asymptotic distribution of the height exhibits an exponential of a Gaussian distribution (with an oscillating term) around the most probable value $k_1 = \lfloor \log_2 n + \sqrt{2 \log_2 n} - \log_2(\sqrt{2 \log_2 n}) + \frac{1}{\log 2} - \frac{1}{2} \rfloor + 1$. More precisely, the asymptotic distribution of a digital search tree is either concentrated on the one point k_1 or the two points $k_1 - 1$ and k_1, which actually proves (slightly modified) Kesten's conjecture.

We derive these results by a combination of analytic methods such as generating functions, Laplace transform, the saddle point method and ideas of applied mathematics such as linearization, asymptotic matching and the WKB method.

Probabilistic Phenomena in Data Compression

I. Kontoyiannis

Department of Statistics
Purdue University
Email: yiannis@stat.purdue.edu

Randomness plays an interesting role in the theory of lossy data compression ("lossy" in that the decoded data is only required to agree with the true data up to within some given level of accuracy). On the one hand data sources are typically modeled by random processes, and on the other hand most of the known constructions of good codes are based on probabilistic arguments – variations of Shannon's "random codes." The interplay between these two sources of randomness gives rise to some interesting phenomena. We will discuss the following:

1. *Codes as probability measures.* At the core of the random coding argument lies a probabilistic result which can be thought of as a natural generalization of the Shannon-McMillan-Breiman theorem. In view of this we can build a useful (and elegant) correspondence between codes and probability measures defined on an appropriate space.
2. *Convergence rates (redundancy) for memoryless sources.* In the simplest case when the data to be encoded comes from a memoryless source (an i.i.d. process), the asymptotic behavior of the optimal codes coincides with that of the random codes, up to and including second order terms. We will explain how this point of view makes it possible to calculate how quickly optimal compression can be approached.
3. *Critical behavior.* For memoryless sources, the above analysis gives very precise information about the pointwise rate of convergence at which the description lengths of optimal compression codes approach Shannon's classical bound of $nR(D)$ bits. This convergence rate is either of order $O(\sqrt{n})$ or $O(\log n)$, where n the length of the uncompressed data. An essentially complete characterization can be given of when the rate is $O(\sqrt{n})$ and when it is $O(\log n)$. Roughly speaking, the rate is $O(\sqrt{n})$ almost always (but not always).

Can small Graphs be colored optimally?

Marek Kubale

The answer to this provocative question is obvious: small graphs can be colored optimally, since all finite graphs can! But the answer to this question is not so obvious, if we restrict ourselves to the most popular sequential coloring algorithms, and by "small graphs" we mean graphs with at most 10 vertices. More formally, let A(G) be the number of colors used by algorithm A to color the vertices of graph G. A graph G is said to be hard-to-color (HC) (resp. slightly HC) if for every (resp. some) implementation of the algorithm A we have A(G) ¿ chi(G), where chi(G) is the chromatic number of G. For a collection of such algorithms graph G is a benchmark, if it is HC for every algorithm in the collection. The study of HC graphs makes it possible to design improved algorithms trying to avoid hard instances as far as possible. Hard-to-color graphs are also good benchmarks for evaluation of existing and future algorithms and provide an alternative way of assessing their quality. In the talk we demonstrate the smallest HC graphs for the best known coloring heuristics in classical applications, as well as when adapted to the chromatic sum coloring and strong coloring of vertices. In addition, we give two collective benchmarks for sequential coloring algoithms.

Distinctness of compositions of an integer: A Probabilistic Analysis

Guy Louchard

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

Compositions of integers are used as theoretical models for many applications. The degree of distinctness of a composition is a natural and important parameter. Here we use as measure of distinctness the number X of distinct parts(or components). In this preliminary approach, we investigate, from a probabilistic point of view, the first empty part, the maximum part size and the distribution of X. We obtain the moments and a numerical estimation of a continuous distribution F, the(discrete) distribution of X being computable from F. We first analyze the classical composition of an integer and next, another compostion: the Carlitz one, where two successive parts are different. We use tools such as Mellin transforms, Markov chain potential theory, limiting hitting times and perturbation analysis.

Evaluating the Average Filtration Efficiency a q-Gram Based Approximate Pattern Matching Algorithm

Pekka Manninen and Erkki Sutinen

Department of Computer Science
University of Joensuu, P.O. Box 111
FIN-80101 Joensuu, Finland
E-mail: {pmanni,sutinen}@cs.joensuu.fi

The approximate pattern matching problem is defined as follows. Given a text $T = T[1 \ldots n]$ and a pattern $P = P[1 \ldots m]$ over an alphabet Σ, and an error level k, give all positions j of text T such that an approximate occurrence of P within a distance of at most k ends at j. Typically, the distance of two strings α and β is given as an edit distance $d(\alpha, \beta)$ which is the minimum number of edit operations, i.e., insertions, deletions, and changes, needed to transform α to β.

The standard solution to the problem needs $O(kn)$ time. This is too much especially for longer texts, and therefore various filtration methods have been developed to search for potential matches. In the following phase, these promising text areas are checked by an accurate, but more expensive method.

A particular type of filtration methods is based on q-grams, i.e., substrings of length q [1]. Basically, the idea is to search for q-grams of pattern P in T. If a certain window of T contains a sufficient amount of these q-grams, it will be considered a potential approximate match.

We evaluate the distribution of the number of potential approximate matches, produced by the q-gram filtration, with five different methods:

1. experimentally;
2. deriving the distribution from the expectation and variance [3];
3. using the estimates for three different intervals in the neighborhood of the expectation [3];
4. using the exact values from generating functions [3]; and
5. using motif statistics [2].

References

[1] P. Jokinen and E. Ukkonen. Two algorithms for approximate string matching in static texts. *Proc. 16th International Symposium on Mathematical Foundations of Computer Science*, Lecture Notes in Computer Science 520, 240–248. Springer-Verlag, 1991.

[2] P. Nicodme, B. Salvy, and P. Flajolet. Motif statistics. *Proc. 7th Annual European Symposium on Algorithms ESA'99*, Lecture Notes in Computer Science, vol. 1643, 194–211, Prague 1999.

[3] M. Regnier, W. Szpankowski. On the approximate pattern occurrence in a text. *Proc. Compression and Complexity of SEQUENCE'97*, IEEE Computer Society, 253–264, Positano 1997.

Urn Models with Non-square Addition Matrix

Hosam Mahmoud

The George Washington University

We discuss urn models progressing according to rules based on drawing fixed-size sets of balls. A non-square addition matrix corresponds to the scheme. (The classical case considers picking the balls one at a time and corresponds to a square addition matrix.) Under appropriate conditions concerning row sums and balancing effects, one gets Gaussian limit laws for the number of balls of a particular color (under suitable norming) from underlying martingale differences. An application in random circuits serves as an example.

A generating tree approach to schedule printing devices

Donatella Merlini

Dip. Sistemi e Informatica - Univ. Firenze

In this talk we present the following problem: suppose we have p slow devices, as for example p printers, and try to describe their behaviour. We can image that their operating time is divided into discrete slots, which, for the sake of simplicity, we consider all of the same length. For each printer, we can think that a time slot corresponds to printing a single page. If each device operates for at most n time slots (i.e., each printer outputs at most n pages), we are interested in studying some parameters, as for example the average waiting time of the users and the average lenght of the printers queue, when the p printers work at the same time. To do so, we introduce a model based on some combinatorial objects, which we call p-histograms, and use the combinatorial properties of these objects to study such parameters. This study is based on an interesting relation between p-histograms, generating trees and $(p+1)$-ary trees. In particular, we find solutions which depend on the counting generating function for $(p+1)$-ary trees.

Parking with density

Jean-Francois Marckert

Institut Eli Cartan

France

We consider a parking with n places $\{p_1, \cdots, p_n\}$ and $m(n) = \alpha n$ cars $(0 < \alpha < 1)$. The i^{th} car **chooses** a place $p_{l(i)}$ among the n places (independently from the other cars). Then, it **parks** on the first empty place among $p_{l(i)}, p_{l(i)+1}, p_{l(i)+2}, \cdots$ (with the convention $p_{n+1} = p_1$, that is the parking is a circle).

In the usual parking, $l(i)$ is a random variable uniformly distributed on $1, \cdots, n$ (one can write $l(i) = \lceil nU_i \rceil$, with the U_i uniform). Here, we consider that

$$l(i) = \lceil nX_i \rceil$$

where the X_i are i.i.d. random variables with continuous density f on $[0, 1]$ (different from uniform density on [0,1]).

- One shows a threshold phenomenon for the apparition of a giant block of consecutive occupied place in the parking.
- One describes the size of the largest blocks in the sub-critical and supercritical cases.
- One gives a precise asymptotic for the total displacement of the cars in the parking.

On the Horton-Strahler Number for Uniform Random Tries

Markus E. Nebel

Johann Wolfgang Goethe-Universität
Fachbereich Informatik
Frankfurt am Main
Germany

In this talk we consider generalized extended binary trees which resemble tries in the uniform model. We prove that under this combinatorial point of view the Horton-Strahler number of a trie with α internal nodes and β keys is of asymptotic size

$$\frac{4^{2\beta-\alpha}\log(\alpha)(2\beta-1)(\alpha+1)(\alpha+2)\binom{2\alpha+1}{\alpha-1}}{8\sqrt{\pi}\alpha^{3/2}\log(2)(\beta-1)\beta\binom{2\beta}{\beta}^2},$$

$\rho := \frac{\alpha}{\beta}$ fixed, $\alpha \to \infty$, on the average. A similar result is shown for tries with α internal nodes but with an arbitrary number of keys.
Furthermore, we shall present results related to the average number of critical nodes in uniform random tries. A node is called critical if it is responsible for an increase of the Horton-Strahler number, it is called p-critical if addititonally its Horton-Strahler number is p. It will be proved that the average number of p-critical nodes in a random trie with α internal nodes and β keys is asymptotically given by

$$4^{-p}\beta + \frac{1}{6} - \frac{1}{6}4^{-p} + \mathcal{O}(\frac{1}{\alpha}),$$

$\rho := \frac{\alpha}{\beta}$ fixed, $\alpha \to \infty$. For tries with α internal nodes and an arbitrary number of leaves the corresponding expectation is equal to

$$\frac{2}{3}4^{-p}\alpha - \frac{1}{12}4^{-p+1} + \frac{1}{6} + 4^{-p} + \mathcal{O}(\frac{1}{\alpha}),\ \alpha \to \infty.$$

We will also present results related to the average number of critical nodes in a trie of given size and to the average path-length (number of edges) from a p-critical node to its critical successors with a Horton-Strahler number $p-1$.

Sampling from the Quicksort limit distribution

Ralph Neininger

University of Freiburg

The weak limit of the scaled number of comparisons needed by the quicksort algorithm to sort n randomly permuted items is known to be determined implicitly by a distributional fixed point equation. We give an algorithm for exact random variate generation from this distribution. This is joint work with L. Devroye and J.A. Fill.

Distributions of Valuations on Trees

Michel Nguyen-The

We consider arithmetical expression trees. The result of the expression is called the *valuation* of the tree. Defining X_n as the random variable equal to the valuation of a tree of size n (number of internal nodes), we study the limit distribution of X_n, for n going towards ∞, in particular cases. We will only consider binary trees, and hence binary operators.

We first constrain the leaves (external nodes) to be labelled by integers belonging to a finite set I, and the internal nodes to belong to be labelled by $+$ and $-$, or by $-$ only. The case with operators $+$ and $-$ easily turns out to be Gaussian. The case with operators $-$ also proves to be Gaussian, but more work is necessary to obtain the result. In fact, trees with $-$ at internal nodes and 1 at leaves involves the classical Narayana numbers, linked to other combinatorial objects. A few combinatorics explains how they appear.

We then consider trees with operators $+$ and $\times$. Calculating the first moments is not difficult, but as there is no concentration of distribution, studying the distribution proves to be harder.

We may then briefly mention the case of trees with max and $+$ operators.

Polynomials over Finite Fields Free from Large and Small Degree Irreducible Factors and Cryptographical Applications

Daniel Panario

Technical University of Vienna

We study the number $N_q(n, m_1, m_2)$ of polynomials of degree n over a finite field fq with all irreducible factors of degree bigger than m_2 and less than or equal to m_1. Applying the saddle point method, we obtain estimates for $N_q(n, m_1, m_2)$ in the range $m_1 = o(n)$ and m_2 slowly tending to infinity, which have the flavor of the ones for the corresponding problem for integers.

Our results have applications in computational number theory and cryptography, and include the smooth polynomials studied by Odlyzko and others. We briefly comment on the application of our results to the study of a generalized version of the index calculus method for the discrete logarithm problem in fq, when $q = p^n$, p is a small prime and $n \to \infty$.

[Joint work with Theo Garefalakis.]

More results about non-crossing trees

Alois Panholzer

Technical University of Vienna

Let P_n be a fixed set of points, conventionally ordered counter-clockwise, forming the vertices of a convex polygon. Define a non-crossing tree as a connected acyclic graph with vertex set P_n whose edges are straight line segments that do not cross.

A systematic approach to enumerate non-crossing structures (trees, forests, graphs, ...) was given by Flajolet and Noy. It relies on generating functions, symbolic methods, singularity analysis and singularity perturbation.

In this talk, we present some bijective results. So we construct a new bijection, to establish the fact, that the number N_n on non-crossing trees of size n satisfies $N_{n+1} = T_n$, where T_n enumerates ternary trees of size n. Also we construct a bijection that explains' the recursion $3(3n-1)(3n-2)T_{n-1} = 2n(2n+1)T_n$ for ternary trees. This bijection can be extended easily to the case of t-ary trees. Considering the vertices labelled from 1 to n and the root of the tree is vertex 1, we also study the number of ascendants and descendants of the vertex with label j.

(Joint work with Helmut Prodinger.)

Combinatorics, probability and geometry or how to analyze range queries in K - dimensional data structures

Martinez Parra

Department de Llenguatges i Sistems Informatics,
Universtat Politecnica de Catalunya, Barcelona

In this talk I'll present some on-going work on the analysis of orthogonal range searches in K-d-trees. The cost of partial matches in K-d-trees (and several other spatial data structures) can be nicely formulated as a characteristic of random K-d-trees as shown by the seminal work by Flajolet and Puech (1986) and the works of many other authors thereafter. Such a formulation makes possible their study using algebraic and analytical tools.

The situation is quite different for orthogonal range searches and nearest neighbor queries in K-d-trees. Most current analysis make use of a combination of geometric and probabilistic arguments. Many of these assume that the search tree is almost balanced (from our past experience, this may lead to somewhat incorrect conclusions). The most relevant results to this respect are due to Devroye and his coauthors (1999, 2000) where probabilistic methods and the relation between range search and partial match are wisely exploited. Furthermore, in these works the K-d-trees are assumed to be random, not almost balanced.

In my talk I'll present an alternative model, which also exploits the close relationship between partial match and range search, yielding new lower and upper bounds on the expected cost of range search in random relaxed K-d-trees. I'll also present a combinatorial approximate model whose main characteristic is that it reduces the problem to purely combinatorial terms.

This is joint work with Amalia Duch.

Old and New Results on Alternating Sums

Helmut Prodinger

Mathematics Department
University of the Witwatersrand
P.O. Wits
2050 Johannesburg
South Africa

Alternating sums (weighted by binomial coefficients) appear in many places, notably when analyzing digital tree structures and protocols for broadcast communication. Asymptotic techniques to handle them include the Mellin transform and Rice's method (from the calculus of finite differences). What Knuth once called the "hardest asymptotic nut that he ever had to crack" is quite well understood today, especially thanks to a survey paper by Flajolet and Sedgewick about the subject.

Yang has proposed several algorithms for broadcast communication, for example selecting the maximum. His analysis (even after adding a coauthor) is very preliminary, and so I joined forces with P. Grabner to analyze these algorithms. This is done in the spirit of Rice's integrals, but leads to some integrals that are quite challenging to evaluate numerically. The recursions related to this class of problems resemble familiar ones known from binary search trees and quicksort/quickselect, but with a logarithmic toll function, something I have never seen before. Thus, with A. Panholzer, we are currently investigating such recursions in some detail.

Gaussian polynomials are the q-versions of binomial coefficients. Indeed, there is also a q-version of Rice's formula. A nice application of it is to easily rederive a formula of Dilcher (which implies another formula, due to Hernandez, in a very straightforward way). Dilcher's paper is related to an analysis of a random graph model, by Andrews, Crippa, and Simon. Research on this subject is still very much in progress.

On Three Classes of Tries with Adaptive Multi-Digit Branching

Yuriy A. Reznik

RealNetworks, Inc.,
Seattle, WA 98121

In this talk we present an overview of recent research on a modification of digital trees (*tries*) with *adaptive multi-digit branching*. Such tries can dynamically adjust degrees of their nodes by selecting the number of digits to be processed per each lookup. While we do not specify any particular practical method for selecting degrees of nodes, we identify three possible classes of such algorithms, and provide a preliminary analysis of their expected behavior in a memoryless (*Bernoulli*) model.

First, we consider tries that select degrees of nodes based on their levels (distance from the root node) and (possibly) parameters of source distribution (or their estimates produced at a run time). Such selection (in a *Bernoulli model*) yields multi-digit tries with nodes of the same degree across the levels. We call these structures *level-partitioned tries* or *LP-tries*, and develop algorithms of building such trees when upper bounds for the expected depth or size of the tree are given. In all these cases, we obtain *LP-tries* of relatively simple structure (e.g. a multi-digit root node followed by 4-ary or binary nodes), and provide their precise asymptotic analysis.

In the second case, we study tries that select degrees of nodes based on the number of strings inserted in sub-tries they originate as well as (estimated) parameters of source distribution. Such tries (we call them *Adaptive Multi-Digit Tries* or *AMD-tries*) appear to be much more difficult to analyze. We provide recurrent expressions for their parameters, and derive few (rather coarse) bounds for the expected depths and sizes of these tries for given parameters of the source. In spite of many obvious shortcomings, our analysis allows us to establish attainability of the constant ($O(1)$) depth and linear ($O(n)$, n - number of strings) size of such tries in the symmetric and asymmetric cases, as well as some other interesting results.

In the last case, we consider tries that select degrees of their nodes based on both their levels and the number of strings they contain. This is the most complex case for the analysis in the group, and we only provide the recurrent expressions for their parameters and illustrate various difficulties we encounter in their analysis.

We will accompany our presentation with a brief survey of several known types of adaptive trie structures (such as: *distributive partitioning* [1, 2], *N-trees* [3, 4, 5], *LC-tries* [6, 7, 8], etc.), and will show various experimental results.

References

[1] W. Dobosiewitz, Sorting by Distributive Partitioning, Information Processing Letters, 7, 1, (1978) 1-6

[2] W. Dobosiewitz, The Practical Significance of DP Sort Revisited, Information Processing Letters, 8, 4 (1979) 170-172

[3] G. Ehrlich, Searching and Sorting Real Numbers, J. Algorithms, 2 (1981) 1-14

[4] M. Tamminen, Analysis of N-Trees, Information Processing Letters, 16, 3 (1983) 131-137

[5] M. Tamminen, Two Levels Are as Good as Any, J. of Algorithms, 6 (1985) 138-144.

[6] A. Andersson, and S. Nilsson, Improved Behaviour of Tries by Adaptive Branching, Information Processing Letters, 46 (1993) 295-300.

[7] S. Nilsson, Radix Sorting and Searching, Ph.D. thesis, Department of Computer Science, Lund University (1996)

[8] S. Nilsson, M. Tikkanen, Implementing a Dynamic Compressed Trie, Proc. 2nd Workshop on Algorithm Engineering (WAE '98), Saarbruecken, Germany, August 20-22, 1998.

[9] Yu. Reznik, On Compact Level Partitioning of Digital Trees, Mathematical Machines and Systems (ISSN 1028-9763), v.1, 1998, pp.38-45.
[10] Yu. Reznik, Some Results on Tries with Adaptive Branching, Proc. 6th Annual International Computing and Combinatorics Conference (COCOON'2000), Bondi Beach , Sydney , Australia, July 26-28, 2000.

A New Ultimate Convex Hull Algorithm in the plane

Uwe Roesler

Mathematisches Seminar,
C.A. - Universitaet Kiel, Germany

Given n points, what is the convex hull of these? In the talk we present an algorithm better than any other known one for that problem. The best results on the runtime are (asymptotically) of the form constant $n \ln h$ with n the number of points and h the number of extremes. The point of the talk are better constants. Simulations confirm that our algorithm is at least twice as fast as others.

This is joint work with David Kravitz and Bill Steiger.

Multidimensional tries: Asymptotic normality of the costs of almost every partial match query

W. Schachinger

Dept. of Statistics and Decision Support Systems
University of Vienna
Brünnerstr. 72, A-1210 Wien, Austria

We investigate costs $C_N(q)$ of partial match retrievals in K-d tries, where both the N-element set of keys and the query q are distributed with respect to the Bernoulli model. We are mainly interested in properties of the sequence $(C_N(q))_{N \geq 0}$ satisfied for almost every fixed q. Denoting by $c_N(q)$ and $v_N(q)$ expectation and variance of $C_N(q)$ for fixed q, and by c_N the expectation of $c_N(q)$, two such properties are

1) $\dfrac{C_N(q) - c_N(q)}{\sqrt{v_N(q)}} \xrightarrow{d} \mathcal{N}(\prime, \infty)$,

2) $\lim\limits_{N \to \infty} \dfrac{c_N(q)}{c_N} = 0$, in the asymmetric Bernoulli model.

In the proof we make use of central limit theorems for martingale difference arrays.

Random planar maps and triangulations

Gilles Schaeffer

Loria-CNRS, Nancy (France)
Gilles.Schaeffer@loria.fr

A triangulation is a planar graph embedded in the plane in such a way that all faces are triangles. No edge can be added in such a graph without breaking the planarity, so that they are sometimes also called maximal planar graphs. Up to homeomorphisms of the plane there are finitely many triangulations with a given number of vertices. We consider the uniform distribution on these finitely many triangulations with n vertices and call them random triangulations.

Triangulations (and more generally embeddings of planar graphs with arbitrary face degrees, called planar maps) have been studied since the 60's by enumerative combinatorists following W. Tutte. From there a theory of random (embeddings of) planar graphs has emerged but some natural questions, like the average diameter of a random triangulations are still open.

We shall consider random sampling algorithms for various families of planar maps. Experiments conducted with them suggest for instance that the diameter of random triangulations with n vertices grows like $n^{1/4}$ and obey a limit law.

These algorithms are probabilistic and have expected running time $cn^{5/3}$ to generate a random triangulation with n vertices. The analysis involves the study of a combinatorial parameter of random maps, the size of the core, the distribution of which obeys the "map Airy law"as described by Philippe Flajolet in his talk.

(Based on my article at STOC'99 and on joint work in progress with Cyril Banderier, Philippe Flajolet and Michéle Soria).

Finding Independent Sets of Guaranteed Size in Uniform Hypergraphs

Hadas Shachnai

Department of Computer Science
The Technion, Haifa 32000, Israel

Let $H = (V, E)$ be a k-uniform hypergraph, with δ_v the degree of any $v \in V$. We use the probabilistic method to derive the lower bound of

$$\alpha_k(H) = \sum_{v \in V} \frac{1}{\binom{\delta_v + 1/(k-1)}{\delta_v}}$$

on the size of a maximal *Independent Set (IS)* in H. Our proof yields efficient randomized parallel algorithm for finding in H an IS of expected size $\alpha_k(H)$, which is the best possible, for any $2 \leq k \leq n$ and any set of vertex degrees $\{\delta_v\}$. We show that our algorithm can be derandomized to yield an NC algorithm, for any k-uniform hypergraph in which the maximal degree is polylogarithmic in n.

Stochastic Analysis of shell sort

Robert Smythe

Department of Statistics,

George Washington University

We analyze the Shell Sort algorithm under the usual random permutation model. Using empirical distributions functions, we recover Louchard's result that the running time of the 1-stage of (2,1)-Shell Sort has a limiting distribution given by the area under the absolute Brownian bridge. The analysis extends to (h,1)-Shell Sort where we find a limiting distribution given by the sum of areas under correlated absolute Brownian bridges. A variation of (h,1)-Shell Sort which is slightly more efficient is presented and its asymptotic behavior analyzed.

Singularity Composition Schemas

Michele Soria

LIP6 - Universite Paris 6
8 rue du Capitaine Scott, Paris

In a combinatorial subsitution $C[M]$, the size of the underlying C-structure is described by the functional composition $C(uM(z))$. Such a composition schema is called "critical" when the singularity of the outer function C is equal to the singular value of the inner function M. In a critical schema, a predominant role is played by the nature of the singularity of $M(z)$. Using Singularity Analysis, we show that if M has a singular expansion in $(1-z)^e$, the limiting distribution of the composition schema is of hypergeometric type in the case $e < 1$, and Airy type in the case $e = 3/2$. Some applications to forests, functional graphs and planar maps will be derived.

Average bit-complexity of Euclidean Algorithms

Brigitte Vallee

We obtain new results regarding the precise average bit-complexity of five algorithms of a broad Euclidean type. We develop a general framework for analysis of algorithms, where the average-case complexity of an algorithm is seen to be related to the analytic behaviour in the complex plane of the set of elementary transformations determined by the algorithms. The methods rely on properties of transfer operators suitably adapted from dynamical systems theory and provide a unifying framework for the analysis of an entire class of gcd-like algorithms.

Some open problems on the Combinatorial Structure of Linear Probing Hashing with Buckets

Alfredo Viola

Instituto de Computation,
Universidad de la Republica,
Julio Herrera y Reissig 565,
Montevideo, Uruguay

It has recently been presented several combinatorial relations relating linear probing hashing with other problems leading to Airy phenomena (like graph connectivity, tree inversions or area ander excursions). When buckets are considered, analyzing the construction cost of hash tables under the linear probing strategy is considerably harder. In this talk we will present partial results in this direction, and some open problems that we are working on.

Average Case Complexity of Continuous Problems

Henryk Woźniakowski

Department of Computer Science, Columbia University, and
Institute of Applied Mathematics, University of Warsaw

There is a branch of computational complexity that studies complexity of continuous problems for which only partial information is available. This branch is called *information-based* complexity and uses the real model of computation. Typical problems are infinite-dimensional for which the inputs are values of multivariate functions of d variables. Examples include multivariate integration and approximation, solution of ordinary or partial differential equations, integral equations, optimization and solving non-linear equations. Such problems can be solved only approximately with error ε. The computational complexity is studied in different settings including worst case, average case, probabilistic and randomized settings.

In this tutorial talk we concentrate on average case complexity for multivariate continuous problems. We discuss tractability issues. The problem is *tractable* if its complexity is polynomial in d and ε^{-1}. We present recent results on tractability of such linear functionals as multivariate integration for the class of continuous functions equipped with the classical Wiener sheet measure.

The Number of Spanning Trees in Circulant Graphs

Yuanping Zhang, Xuerong Yong and Mordecai Golin
Department of Computer Science,
The Hong Kong University of Science and Technology

Let $1 \leq s_1 < s_2 < \cdots < s_k$. The *undirected circulant graph*, $C_n^{s_1,s_2,\cdots,s_k}$, has n vertices labelled $0,\ 1,\ 2,\ \cdots,\ n-1$, with each vertex i $(0 \leq i \leq n-1)$ adjacent to the $2k$ vertices $i \pm s_1,\ i \pm s_2,\ \cdots,\ i \pm s_k \bmod n$. The special case $C_n^{1,2}$ is known as *the square of the cycle* of n vertices.

In this paper we develop a method for determining, given $1 \leq s_1 < s_2 < \cdots < s_k$, the exact number of spanning trees in the circulant graph $C_n^{s_1,s_2,\cdots,s_k}$ as a function of n. For the square of the cycle case it was already known (proven by Kleitman and Golden and Baron, Prodinger, Tichy, Boesch and Wang) that the number of spanning trees is nF_n^2, where F_n are the *Fibonacci* numbers. In this talk we generalize this formula and show that for any given $1 \leq s_1 < s_2 < \cdots < s_k$, the number of spanning trees of $C_n^{s_1,s_2,\cdots,s_k}$ satisfies a recurrence relation of a special form.