
20th International Meeting on Probabilistic, Combinatorial, and Asymptotic Methods in the Analysis of Algorithms

AOFA'09

June, 14-19 2009

Titles

- ADDARIO-BERRY, Louigi – McGill University: “*Prim’s algorithm, critical percolation, and infinite sub-trees of finite trees*”
- AGUECH, Rafik – Faculté des Sciences de Monastir (Tunisia): “*The Moran Model, height of random walks and analysis of a cache algorithm*”
- BANDERIER, Cyril – LIPN – Univ. Paris 13: “*Average Case Analysis of NP-complete Problems: Maximum Independent Set and Exhaustive Search Algorithms*”
- BARYSHNIKOV, Yuliy – Bell Labs, USA: “*Elliptic curves and frozen regions in fortresses*”
- BOUTTIER, jérémie – CEA, France: “*Metric properties of large random planar maps*”
- CHAUVIN, Brigitte – LMV Univ. Versailles & Inria Rocquencourt: “*Limit distributions for Pólya urns*”
- DAVID, Julien – LIGM, Univ. Paris Est: “*On the average complexity of Moore’s state minimization algorithm*”
- DRMOTA, Michael – TU Wien, Institute of Discrete Mathematics and Geometry: “*(Un)Expected Behavior of the Digital Search Tree Profile*”
- FILL, Jim – The Johns Hopkins University, Department of Applied Mathematics and Statistics: “*The numbers of symbol comparisons used by QuickSelect and by QuickSort (Limiting Distributions)*”
- FLAJOLET, Philippe – Inria, Algorithms: “*The numbers of symbol comparisons used by QuickSelect and by QuickSort (Overview)*”
- FLAJOLET, Philippe – Inria, Algorithms: “*Discrete Buffon Machines and Numbers*”
- FUSY, Eric – LIX, Polytechnique, France: “*A combinatorial method for counting planar graphs*”
- GITTENBERGER, Bernahrd – TU Wien, Institute for Discrete Math. and Geometry: “*On the Moments of the Area below Lattice Paths*”
- GNEDIN, Sasha – Utrecht University: “*Small counts in the infinite occupancy scheme*”
- GRÜBEL, Rudolf – Leibniz Universitaet Hannover, Germany: “*On the subtree size profile of binary search trees*”
- HEUBERGER, Clemens – TU Graz: “*Digit Distribution in Positional Number Systems with Digits Forming an Arithmetic Progression*”
- HOLMES, Susan – Stanford University: “*Random threshold graphs, their limits and applications*”
- HOLMGREN, Cecilia – Uppsala Univ., Sweden: “*Characteristics of Split Trees by use of Renewal Theory*”
- HWANG, Hsein-Kuei – Academia Sinica, Taiwan: “*Psi-series method in random trees and moments of high orders*”
- JACQUET, Philippe – EPI Hipercom, INRIA, France: “*Information propagation speed in DTN*”
- JANSON, Svante – Uppsala Univ., Sweden: “*Some applications of renewal theory to the analysis of algorithms*”
- LLADSER, Manuel – Univ. of Colorado, USA: “*A new approach to population diversity: the extrapolation problem*”
- LOUCHARD, Guy – Univ. Libre de Bruxelles, Belgium: “*Asymptotic independence of Uniform Markov chains*”

- MAHMOUD, Hosam – George Washington University: “Gaussian phases in generalized coupon collection”
- NEBEL, Markus – Univ. of Kaiserslautern, Germany: “Maximum Likelihood Analysis of Algorithms”
- NOY, Marc – Universitat Politècnica de Catalunya: “Asymptotic properties of graphs of given genus”
- ROESLER, Uwe – Univ. of Kiel, Germany: “The Weighted Branching Process and Stochastic Fixed Points”
- SANKOFF, David – Ottawa University, Canada: “Gene order, gene clusters and random graphs”
- SCHAEFFER, Gilles – LIX, Polytechnique, France: “TBA”
- SORIA, Michèle – LIP6/UPMC, France: “Limiting Distribution for Distances in k -trees and other graphs”
- SZPANKOWSKI, Wojciech – Purdue University, USA: “On a non-prefix code and its analysis”
- VALLÉE, Brigitte – CNRS, Caen, France: “The number of symbol comparisons used by QuickSelect and by QuickSort”
- VAN LEEUWAARDEN, Johan – Eindhoven Univ. of Technology and EURANDOM, Netherlands: “Singularity analysis through functional equations for random walks in the quarter plane”
- WILF, Herbert – Univ. of Pennsylvania, USA: “Some new asymptotic problems in partition theory”

Abstracts



ADDARIO-BERRY, Louigi (McGill University): « *Prim’s algorithm, critical percolation, and infinite sub-trees of finite trees* »

(joint work with Simon Griffiths and Ross Kang).

Let G be a finite or infinite connected rooted graph and assign independently to each edge of G a random weight w_e with distribution $\text{Unif}[0, 1]$. The invasion percolation cluster is the tree T grown inside G as follows: start with the root, add the lowest weight edge leaving the root, then at each step thereafter add the lowest weight edge leaving the current tree. When applied to finite weighted graphs this procedure is the well known Prim’s algorithm for growing a minimum weight spanning tree. Less is known about invasion percolation on infinite graphs; it is a subject of current research. We study invasion percolation on Aldous’ Poisson-weighted infinite tree, which may be viewed as a limit of invasion percolation on either the complete graph K_n (as $n \rightarrow \infty$) or the d -ary infinite tree (as $d \rightarrow \infty$). Furthermore, this limit picture, which may be obtained via a Poisson point process on the upper right quadrant $[0, \infty) \times [0, \infty)$, provides a suitable setting for short proofs of many interesting properties. We also define a related stationary process which can be seen as capturing the behaviour of invasion percolation far from its starting point. This behaviour is closely linked with that of critical bond percolation, and so serves as an example of self-organised criticality.



AGUECH, Rafik (Faculté des Sciences de Monastir (Tunisia)): « *The Moran Model, height of random walks and analysis of a cache algorithm* »

Joint work with C. Banderier



BANDERIER, Cyril (LIPN – Univ. Paris 13): « *Average Case Analysis of NP-complete Problems: Maximum Independent Set and Exhaustive Search Algorithms* »

In this talk, we deal with rigorous average case analysis of NP complete problems, relying on some mathematical tools from complex analysis and probability theory. We consider here one of the historical prototype of such problems: Maximum Independent Set (MIS).

For a large class of exhaustive algorithms which always find exactly a MIS, their complexity is directly related to their number of iterations. Under the $\Gamma(n, m)$ and $G(n, p)$ distribution for graphs, we give some fascinating phase transitions between exponential (A^n), superpolynomial ($n^{\ln n}$), and polynomial (n^d) average complexities.

Our approach gives a precise picture of the “complexity landscape” of these algorithms, depending on the average degree (or the ratio vertices/edges) of the graph, and gives access to the location of the “hard regions” where people could then sample their inputs if they want to make some benchmarks (may it be for the worst or average case).

The challenging associated mathematical aspects force us to introduce new analyses (for a large class of recurrences), which will clearly be of interest for many other NP hard problems (typically, optimization problems on graphs).

Direct applications cover graph 3-coloring (which is also a deep motivation for considering our class of exhaustive Tarjan-Chvátal like algorithms), and numerous constraint satisfaction problems.

Joint work with Hsien-Kuei Hwang, Vlady Ravelomanana, Vytas Zacharovas.



BARYSHNIKOV, Yuliy (Bell Labs, USA): *« Elliptic curves and frozen regions in fortresses »*

The diablo tilings of fortresses (or, equivalently, appropriately weighted complete matchings in a family of bipartite graphs) were introduced and studied by Propp and his collaborators. They exhibited frozen-moderate transition similar to that seen in Aztec diamonds, but far more complicated. Using Wulff construction (i.e. entropic approach), Kenyon and Okounkov proved the conjectured by Cohn and Pemantle formula for the frozen region boundary. Here, we give an alternative derivation for the boundary of the frozen region and determine the densities in the moderate region, basing on our with Pemantle recent results on asymptotics of multivariate generating function.



BOUTTIER, j eremie (CEA, France): *« Metric properties of large random planar maps »*

In this talk, we mainly consider planar quadrangulations that form a particular class of maps. When equipped with the graph distance, these may be viewed as metric spaces. We wish to understand their statistical properties under the uniform measure over maps of given size. In the limit of large size, we expect to obtain asymptotic results which are independant of the particular class of maps under consideration, and which characterize their universal scaling limit, the Brownian map.

We shall review a number of metric properties of large random quadrangulations. Their derivation rely on the Schaeffer bijection between quadrangulations and well-labeled trees, or its suitable generalizations, and on a remarkable ”integrability” property which leads to explicit expressions for the associated generating functions.

If time allows, I shall briefly discuss statistical physics models on random maps (such as the Ising model), in which we expect to find different universal scaling limits, but where many steps are missing.

This talk is based on joint work with Emmanuel Guitter and earlier work with Philippe Di Francesco.



CHAUVIN, Brigitte (LMV Univ. Versailles & Inria Rocquencourt): *« Limit distributions for P lya urns »*

The talk focuses on a work in progress together with Nicolas Pouyanne, in order to identify limit distributions for some P lya urns. In a first part, the definition and some historical examples are given (Ehrenfest urn, original P lya urn, ...), with emphasis on P lya urns coming from analysis of algorithms, especially those related to random trees (recursive trees, m -ary search trees, ...). A particular attention is paid to the phase transition which appears in the limit distributions introducing a distinction between “small” and “large” urns.

In the main part, a two colors P lya urn is considered. Assume it is balanced (i.e. the same number S of balls are added at each draw) and assume it is a “large” urn (i.e. σS , the second eigenvalue of the replacement matrix satisfies $S/2 < \sigma S \leq S$). After n drawings, the composition vector has asymptotically a first deterministic term of order n and a second random term of order n^σ . The goal is to catch the limit distribution of this random term.

The method consists in embedding the discrete urn in continuous time, getting thus a multitype branching process. Writing the dislocation equations for this process gives a system of differential

equations on the Fourier transforms of the limit distributions. The resolution is carried out and it turns out that the Fourier transforms are explicitly related to Abelian integrals on a Fermat curve. Some properties of these limit distributions are explored: do they have moments of any order? Are they infinitely divisible? Are they stable? Are they determined by their moments? Do they have a density? Are the tails heavy or not?

Answers:

·πæwɪtɔq ʌɪɾɔɔɔɪɾɪ ʔæʎ ʔɪɾɔɪʂəɪɪɪɪ ɪɔɔɔ ʔɪɪ ʔæʎ ʔsæʎ

The same method is applied to m -ary search trees, in the case when $m \geq 27$, i.e. when the corresponding urn is “large” (meaning that the second eigenvalues of the replacement matrix are large: their real part σ satisfies $1/2 < \sigma \leq 1$). We obtain a system of $m - 1$ integro-differential equations, related to the non linear equation $y^{(m-1)} = y^m$.



DAVID, Julien (LIGM, Univ. Paris Est): « *On the average complexity of Moore’s state minimization algorithm* »

Thanks to good algorithmic properties, deterministic finite automata are often used to represent regular languages in a computer. The minimal automaton of such a language is the unique smallest automaton that recognizes it. In the worst case scenario the best known algorithm computes the minimal automaton in a time complexity bounded by $O(n \log n)$. We prove that another algorithm, due to Moore, which is a lot simpler to implement but often ignored because of its worst case complexity in $O(n^2)$, has an average complexity bounded by $O(n \log n)$.

In collaboration with F. Bassino and C. Nicaud.



DRMOTA, Michael (TU Wien, Institute of Discrete Mathematics and Geometry): « *(Un)Expected Behavior of the Digital Search Tree Profile* »

In collaboration with Wojciech Szpankowski.

A digital search tree (DST) – one of the most fundamental data structure on words – is a digital tree in which keys (strings, words) are stored directly in (internal) nodes. Such trees find myriad of applications from the popular Lempel-Ziv’78 data compression scheme to distributed hash tables. The profile of a DST measures the number of nodes at the same distance from the root; it is a function of the number of stored strings and the distance from the root. Most parameters of DST (e.g., height, fill-up) can be expressed in terms of the profile. However, from the inception of DST, the analysis of the profile has been elusive and it has become a prominent open problem in the area of analysis of algorithms. We make here the first, but decisive step, towards solving this problem. We present a precise analysis of the average profile when stored strings are generated by a biased memoryless source. The main technical difficulty of analyzing the profile lies in solving a sophisticated recurrence equation. We present such a solution for the Poissonized version of the problem (i.e., when the number of stored strings is generated by a Poisson distribution) in the Mellin transform domain. To accomplish it, we introduce a novel functional operator that allows us to express the solution in an explicit form, and then using analytic algorithmics tools to extract asymptotic behavior of the profile. This analysis is surprisingly demanding but once it is carried out it reveals unusually intriguing and interesting behavior. The average profile undergoes several phase transitions when moving from the root to the longest path. At first, it resembles a full tree until it abruptly starts growing polynomially and it oscillates in this range. Our results are derived by methods of analytic algorithmics such as generating functions, Mellin transform, Poissonization and de-Poissonization, the saddle-point method, singularity analysis and uniform asymptotic analysis.



FILL, Jim (The Johns Hopkins University, Department of Applied Mathematics and Statistics): « *The numbers of symbol comparisons used by QuickSelect and by QuickSort (Limiting Distributions)* »

See abstract in common with Brigitte Vallée (and also Philippe Flajolet).



FLAJOLET, Philippe (Inria, Algorithms): « *The numbers of symbol comparisons used by QuickSelect and by QuickSort (Overview)* »

See abstract in common with Brigitte Vallée (and also Jim Fill).



FLAJOLET, Philippe (Inria, Algorithms): « *Discrete Buffon Machines and Numbers* »

Generalizing the well-known Buffon experiment, we ask ourselves which probability distributions can be generated *exactly* from a source of unbiased coin flips. We shall describe and analyse a few simple Buffon machines that can generate geometric, Poisson, and logarithmic-series distributions—these are required to transform continuous Boltzmann samplers into discrete random generators of classical combinatorial structures. Say that a number is Buffon if it is the probability of success of a *simple* probabilistic experiment based on discrete coin flips. We shall also provide human-accessible Buffon machines, which require a dozen coin flips or less, on average, and produce experiments whose probabilities are expressible in terms of numbers such as π , $\log 2$, $\exp(-1)$, $\zeta(3)$.

(Based on joint work with Michele Soria and Maryse Pelletier.)



FUSY, Eric (LIX, Polytechnique, France): « *A combinatorial method for counting planar graphs* »

Giménez and Noy have given recently a complete analytic solution to the problem of counting planar graphs. As a crucial step they calculate implicit analytic expressions for the series counting planar graphs. We present a new method for finding such an expression, which avoids the technical integration steps addressed by Giménez and Noy. The two main ingredients are a bijection due to Bouttier, Di Francesco and Guitter for counting vertex-pointed planar maps, as well as a decomposition grammar (with minus signs) obtained by translating a classical decomposition (formalised by Tutte) of a graph into 3-connected components.



GITTENBERGER, Bernahrd (TU Wien, Institute for Discrete Math. and Geometry): « *On the Moments of the Area below Lattice Paths* »

We study the moments of the area below directed lattice paths, i.e., walks on the integers with jumps from an a priori given finite set of integers. The focus is on enumerative results obtained by means of generating functions and the kernel method. First we study excursions and meanders of simple paths, where the roots of the kernel have a somewhat simpler structure, and give full asymptotics of average and variance. Then we present some investigations on more complicated walks which is work in progress. It is possible to obtain representations for the generating functions of all factorial moments in terms of certain roots of the kernel.



GNEDIN, Sasha (Utrecht University): « *Small counts in the infinite occupancy scheme* »

The paper is concerned with the classical occupancy scheme with infinitely many boxes, in which n balls are thrown independently into boxes $1, 2, \dots$, with probability p_j of hitting the box j , where $p_1 \geq p_2 \geq \dots > 0$ and $\sum_{j=1}^{\infty} p_j = 1$. We establish joint normal approximation as $n \rightarrow \infty$ for the numbers of boxes containing r_1, r_2, \dots, r_m balls, standardized in the natural way, assuming only that the variances of these counts all tend to infinity. The proof of this approximation is based on a de-Poissonization lemma. We then review sufficient conditions for the variances to tend to infinity. Typically, the normal approximation does not mean convergence. We show that the convergence of the full vector of r -counts only holds under a condition of regular variation, thus giving a complete characterization of possible limit correlation structures.

Joint work with A. D. Barbour.



GRÜBEL, Rudolf (Leibniz Universität Hannover, Germany): *« On the subtree size profile of binary search trees »*

For random trees T generated by the binary search tree algorithm from uniformly distributed input we consider the subtree size profile, which maps $k \in \mathbb{N}$ to the number of nodes in T that root a subtree of size k . Complementing earlier work by Devroye (1991), by Feng, Mahmoud and Panholzer (2008), and by Fuchs (2008), we obtain results for the range of small k -values and the range of k -values proportional to the size n of T . In both cases emphasis is on the process view, i.e. the joint distributions for several k -values. Regarding the dynamics of the tree sequence we find a qualitative difference between the asymptotic behaviour of the small and large range of the profile.

The talk is based on joint work with Florian Dennert, and with Steve Evans and Anton Wakolbinger.



HEUBERGER, Clemens (TU Graz): *« Digit Distribution in Positional Number Systems with Digits Forming an Arithmetic Progression »*

We consider digit expansions to a positive integer base $b \geq 2$ with positive digits forming an arithmetic progression, whose common difference is coprime to the base. Since in general, this is not sufficient to represent all positive integers, the most significant digit is allowed to belong to a certain larger set of possible digits. We do not require zero to be an admissible digit.

The motivation for studying these digit expansions comes from a quite surprising answer to a graph theoretic optimisation problem: For a given number n of vertices, the unique graph with maximum degree 3 with the largest number of independent vertex subsets can be described by using a binary expansion of $n + 1$ using the digits 1 and 4 (and additionally, the digit 2 is allowed in the most significant position). For higher maximum degrees, higher bases and other arithmetic progressions arise.

The obvious question that comes to mind is how the digits are distributed, among, say, the first N integers. To answer this, the Mellin-Perron technique is used. But an even simpler question which isn't really a question for the standard b -ary system, say, must be addressed: the length of the representation! In general, the length is not a monotone function anymore, and a larger number may have a shorter representation!

In base b , the length of the standard representation (index of the leading digit) of n is given by $\lfloor \log_b n \rfloor$. Similar formulæ are given in our case, but with several terms, all expressed with floor functions.

We give explicit formulæ for length and digits, we also study the averages and obtain asymptotic results involving periodic functions that are completely described via their Fourier coefficients.

(based on joint work with H. Prodinger and St. Wagner)



HOLMES, Susan (Stanford University): *« Random threshold graphs, their limits and applications »*

I will present a limit theory of large threshold graphs and apply this to a variety of models for random threshold graphs. The results give a nice set of examples of this special class of graphs which have an underlying variable, very useful in applications. Graphs have important applications in modern systems biology and social sciences. Edges are created between interacting genes or people who know each other. However graphs are not objects which are naturally amenable to simple statistical analyses, there is no natural average graph for instance. Being able to predict or replace a graph by hidden (statisticians call them latent) real variables has many advantages. This talk studies such a class of graphs, that sits within the larger class of interval graphs itself a subset of intersection graphs.

This is joint work with Persi Diaconis and Svante Janson.



HOLMGREN, Cecilia (Uppsala Univ., Sweden): « *Characteristics of Split Trees by use of Renewal Theory* »

We investigate some characteristics of random split trees introduced by Devroye; split trees include for example binary search trees, m -ary search trees, quadrees, median of $(2k + 1)$ -trees, simplex trees, tries and digital search trees. More precisely: We introduce renewal theory, and use this theory to prove several results about split trees. A split tree of cardinality N is constructed by distributing N “balls” (which often represent some “key numbers”) in a subset of vertices of an infinite tree. One of our main result is to give a relation between the deterministic number of balls N and the random number of vertices n . Devroye has shown that there is a central limit law for the depth of the last inserted ball so that most vertices are close to $\frac{\ln N}{\mu} + \mathcal{O}(\sqrt{\ln N})$, where μ is some constant depending on the type of split tree; we sharpen this result by finding an upper bound for the expected number of vertices with depths $\geq \frac{\ln N}{\mu} + \ln^{0.5+\epsilon} N$ or depths $\leq \frac{\ln N}{\mu} - \ln^{0.5+\epsilon} N$ for some arbitrary constant $\epsilon > 0$. We also find the first asymptotic of the variances of the depths of the balls in the tree.



HWANG, Hsein-Kuei (Academia Sinica, Taiwan): « *Psi-series method in random trees and moments of high orders* »

An unusual and surprising lacunary asymptotic expansion is derived of the probability that two randomly chosen binary search trees are identical. A similar quantity also arises in the analysis of phylogenetic trees. The approach we use to attack the nonlinear differential equation is new in the analysis of algorithms and relies on the so-called psi-series method. Such an expansion turns out to be the tip of an iceberg and we show that similar lacunary-type asymptotic expansions also appear in other random trees of logarithmic height and in the asymptotics of higher moments of some problems in discrete probability and statistical physics. (This talk is based on joint work with Hua-Hua Chern and Conrado Martinez.)



JACQUET, Philippe (EPI Hipercom, INRIA, France): « *Information propagation speed in DTN* »

We analyze the performance of epidemic routing in large-scale intermittently connected networks, under a random geometric graph model and for different mobility parameters (such as the random-waypoint, random walk and Brownian motion models). We assume that packet transmission and propagation are instantaneous within a fixed radio range. We derive a generic scaling law on the delay which provides us with lower bounds: the average delay from a source to a destination and the average broadcast delay are both at least proportional to the square root of n times the ratio of the radio range over node speed, where n is the number of nodes in the network for a given normalized density. This result disproves the conjecture of a delay proportional to $\log n/n$ times meeting rate which was obtained from previous models based on an Erdos-Renyi graph model evolution. The average meeting rate is proportional to the node speed times radio range divided by the node density.

The analysis is based on a continuous space-time information path density upper-bound and segmentation, Poissonization and dePoissonization over the number of nodes, and Laplace transform singularity analysis.



JANSON, Svante (Uppsala Univ., Sweden): « *Some applications of renewal theory to the analysis of algorithms* »

I will show how some well-known results from renewal theory (including results by myself) can be used in AofA. Examples include applications to random digital search trees and to Khodak codes.



LLADSER, Manuel (Univ. of Colorado, USA): « *A new approach to population diversity: the extrapolation problem* »

Consider an urn with colored balls but with an unknown composition i.e. you do not know what are the specific colors in the urn nor their relative proportions. To learn about its composition imagine sampling balls with replacement out of the urn. The urn models a variety of situations of practical interest. For instance, the different colors in the urn could represent different solutions to a particular binding site problem in an RNA pool, or the different species of bacteria present in a section of soil. In rich environments like these even the most exhaustive data cannot saturate sampling and the traditional tools to estimate alpha-diversity (i.e. the total number of species or other taxonomic units present in the environment) provide inaccurate results. This motivates to reformulate the problem in terms of the fraction of the population that belongs to unseen classes. During my talk I will show that this last quantity cannot be predicted once the sample size has been fixed. Instead, using randomized sample sizes, one can construct conservative as well as exact confidence intervals. Extensions of the basic idea will allow us to predict other measures of diversity e.g. beta-diversity and gamma-diversity.

This research has been partially funded by the NIH grant 3R01GM048080-13S1 and the NSF grant DMS-0805950.



LOUCHARD, Guy (Univ. Libre de Bruxelles, Belgium): « *Asymptotic independence of Uniform Markov chains* »

Let $\{X_k\}_{k \geq 0}$ be a discrete time Markov chain on the state space of positive integers. We assume that the chain is ergodic, and denote its stationary distribution by $\{\pi(j)\}_{j > 0}$. Let $P(i, j)$, $(i, j > 0)$, denote the transition probabilities. Our focus is on the chains that meet a *uniformity* condition: the series $\sum_{j > 0} P(i, j)$ converges uniformly for $i > 0$. For the uniform chains, one should expect that, for a large j , the consecutive visits to $D = [j, \infty)$ occur at time moments far removed from each other. If so, the chain should have plenty of time to “stationarize” between the visits to D ; specifically, for the repeat visits the distribution of a visited state should be close to $\{\pi(j)/\pi(D)\}_{j \in D}$, $\pi(D) := \sum_{j \in D} \pi(j)$, and the interarrival times should be close to being independent, geometrically distributed with parameter $\pi(D)$. Our goal is to fully confirm this longstanding conjecture.

This result will allow us to approximate the visit (“hitting”) times and the visit (“hit”) locations for such a remote D during a time interval $[0, m]$ by an allocation model: at each step a ball lands in a box in D^c with probability $\pi(D^c)$, and in a box $j \in D$ with probability $\pi(j)/\pi(D)$, independently of other balls. (Of course, for this to make sense, $m \rightarrow \infty$ and moreover $\pi^{-1}(D) = O(m)$.)

Let us now see the evolution of the Markov process, for m steps, as an urn model, where we throw m balls into urns, (each visit to k throws a ball into urn k), and we are interested in random variables(RV) related to large j , as $m \rightarrow \infty$, for instance

- first empty urn
- last non-empty urn
- number of non-empty urns (measure of distinctness)
- number of urns with exactly i balls (i fixed)
- number of gaps, i.e. number of empty urns between urn 1 and the last non-empty urn

In our favorite examples, they are all related to $j = \log m + \mathcal{O}(1)$ or $j = \log m + 2 \log \log m + \mathcal{O}(1)$. Assume that we can asymptotically use $\pi(j)$ instead of $P(i, j)$ i.e. we consider m iid RV with distribution $\pi(j)$. It is known (by Poissonization-dePoissonization) that this is asymptotically equivalent to independent urns, where the number of balls in urn j is given by a Binomial $B[m\pi(j), m\pi(j)(1 - \pi(j))]$.

The motivation for this problem comes from the study of two Combinatorial structures: the column-convex animals (cca) and the Carlitz compositions . These structures have been probabilistically analyzed in several previous papers with tools such that complex analysis, singularity analysis, hitting times analysis, Markov chains, moments of extreme-value related distributions functions. With the asymptotic independence property, we can now obtain, in an almost mechanical way, old and new results (old ones were usually derived by complicated machinery).

Joint work with Boris Pittel.



MAHMOUD, Hosam (George Washington University): « *Gaussian phases in generalized coupon collection* »

A generalized coupon collection problem is considered in which a random number of purchases occurs at each stage of collecting n , a large number of coupons. We address the question of how many different coupons have been collected after $k=k(n)$ draws, as n tends to infinity. We identify three phases of $k(n)$: The sublinear, the linear, and the superlinear. In the sublinear phase we see $o(n)$ different coupons, and with true randomness in the number of purchases, under the appropriate centering and scaling, a Gaussian distribution is obtained across the entire phase. However, if the number of purchases is fixed, a degeneracy comes into the picture and normality holds only at the higher end of this phase. In the case of a number of purchases with a fixed range, the small number of different coupons collected in the sublinear phase is upgraded to a number in need of centering and scaling to become normally distributed in the linear phase with a different normal distribution of the type that appears in the usual central limit theorems. The Gaussian results are obtained via martingale theory. We say a few words in passing about the high probability of collecting nearly all the coupons in the superlinear phase. It is our aim to present the results in a way to explore the critical transition at the “seam lines” between different Gaussian phases, and between these phases and other nonnormal phases.



NEBEL, Markus (Univ. of Kaiserslautern, Germany): « *Maximum Likelihood Analysis of Algorithms* »

We present a new approach for an average-cases analysis of algorithms and data structures that supports a non-uniform distribution of the inputs and is based on the maximum likelihood training of stochastic grammars. The approach is exemplified by analyzing heapsort and quicksort comparing the results derived to known ones obtained by traditional techniques. Investigating traditional settings like the random permutation model we rediscover well-known results formerly derived by pure analytic methods; changing to biased data yields original results which proves our method useful. All but one steps of our analysis can be automated on top of a computer-algebra system. Thus our new approach eases the effort required for an average-case analysis exceptionally allowing for the consideration of realistic input distributions with unknown distribution functions at the same time.

As a by-product our approach yields an easy way to generate random combinatorial objects according to various probability distributions.



NOY, Marc (Universitat Politècnica de Catalunya): « *Asymptotic properties of graphs of given genus* »

We show that the number of labelled graphs on n vertices embeddable in an orientable surface of genus g is asymptotically given by

$$c(g)n^{5(g-1)/2-1}\gamma^n n!,$$

where $c(g)$ is a constant depending on g and $\gamma = 27.22688\dots$ is the exponential growth constant of planar graphs. We also show that several basic parameters, like number of edges, number of blocks, and size of the largest block, have a limit distribution law which does not depend on the genus. The proofs use the enumeration of maps on surfaces (Bender, Canfield, Richmond and others), the enumeration of labelled planar graphs (Giménez, Noy), and embedding results based on the face-width parameter (Robertson, Vitray). Similar results hold for graphs embeddable in non-orientable surfaces.

This is joint work with Guillaume Chapuy, Eric Fusy, Omer Giménez, and Bojan Mohar.



ROESLER, Uwe (Univ. of Kiel, Germany): « *The Weighted Branching Process and Stochastic Fixed Points* »

We give a (personal) overview on the Weighted Branching Process and Fixed Point Equations, starting from the beginning to recent results and work in progress. Beside the mathematical aspect we present examples from various fields, like analysis of algorithms, population genetics, biology, financial mathematics and more.



SANKOFF, David (Ottawa University, Canada): « *Gene order, gene clusters and random graphs* »

Comparisons of chromosomal gene order from different genomes may reveal various similarities. These may, however, be purely coincidental, especially in large genomes. We are thus motivated to study the chance similarities among random genomes as a baseline for testing the significance of the evolutionary relationship among real genomes. For example, definitions of conserved clusters in two genomes differ in how much emphasis they place on common gene content versus how similar are the orders of the common genes. We investigate the notion of generalized gene adjacency, a clustering criterion that adjusts the trade-off between content and order. The connected components of the intersecting generalized adjacency graphs from two genomes become the gene clusters. These definitions require imposing values on some threshold parameters. We suggest a way of determining "natural" values for the clustering parameters in terms of maximizing sensitivity of a weighting scheme on pairs of genes at different distances in two random genomes. This leads surprisingly to questions in the theory of record times of random variables. We also compare the percolation behaviour of the random adjacency graphs to other kinds of random graph.



SCHAEFFER, Gilles (LIX, Polytechnique, France): « *TBA* »



SORIA, Michèle (LIP6/UPMC, France): « *Limiting Distribution for Distances in k -trees and other graphs* »

This talk examines the distances between vertices in k -trees and other families of graphs. The starting point is a correspondence between these graphs and varieties of trees that can be specified in terms of combinatorial specifications. Distances in the graphs are transformed into parameters of trees that can be studied via generating functions and analytic combinatorics. For example, we show a Rayleigh limiting distribution for distances to a random vertex in a random k -tree: in a k -tree on n vertices, the proportion of vertices at distance $d = x\sqrt{n}$ from a random vertex is asymptotic to $(c_k^2 x)/\sqrt{n} \times \exp(-(c_k^2 x^2)/2)$, where $c_k = kH_k$.

Joint work with A. Darrasse and O. Bodini



SZPANKOWSKI, Wojciech (Purdue University, USA): « *On a non-prefix code and its analysis* »

In this talk, we discuss one-to-one codes whose average length is smaller than the source entropy in defiance of the lower bound of Shannon. One-to-one codes are "one shot" codes that assign a distinct codeword to source symbols and are not necessarily prefix codes (more generally, uniquely decodable). First, we analyze a block code of length n generated for a binary memoryless source, and prove that the average redundancy is

$$-\frac{1}{2} \log_2 n + C + F(n) + o(1)$$

where C is a constant and either $F(n) = 0$ if $\log_2(1-p)/p$ is irrational (p is the probability of generating a "0") or otherwise $F(n)$ is a fluctuating function as the code length increases. Then we partially extend this result to non-binary codes which turns out to be more interesting and analytically more challenging. These findings were derived by a combination of analytic tools such as precise evaluation of Bernoulli sums, the saddle point method, and theory of distribution of sequences modulo 1.



VALLÉE, Brigitte (CNRS, Caen, France): « *The number of symbol comparisons used by QuickSelect and by QuickSort* »

Philippe Flajolet (Overview), Brigitte Vallée (Average-case analysis), and Jim Fill (Limiting distributions).

Consider n independent and identically distributed keys, each represented as an infinite string of symbols from a linearly ordered finite or countably infinite alphabet. Furthermore, assume that each time two keys are compared, they are compared symbol-by-symbol until differing corresponding symbols are found. We analyze two algorithms: (i) **QuickSort** and (ii) **QuickQuant**(α), i.e., **QuickSelect** when it is used to find the α -quantile. Let S_n be the total number of symbol comparisons used by the algorithm. Under a mild tamedness condition on how each key is generated probabilistically, we obtain two results: a precise asymptotic analysis of the mean number $E[S_n]$ for both algorithms, and, in the case of **QuickQuant**(α), an explicit limiting distribution for S_n/n . Fill is currently working with Ph.D. student Patrick Bindjeme on a limiting distribution in the case of **QuickSort**.



VAN LEEUWAARDEN, Johan (Eindhoven Univ. of Technology and EURANDOM, Netherlands): « *Singularity analysis through functional equations for random walks in the quarter plane* »

We present a novel technique to derive asymptotic expressions for rare event probabilities for random walks in the quarter plane. For concreteness, we study a tandem queue with Poisson arrivals, exponential service times and coupled processors. The service rate for one queue is only a fraction of the global service rate when the other queue is non empty; when one queue is empty, the other queue has full service rate. The bivariate generating function of the queue lengths gives rise to a functional equation. In order to derive asymptotic expressions for large queue lengths, we combine the kernel method for functional equations with boundary value problems and singularity analysis. Joint work with Fabrice Guillemin (Orange).



WILF, Herbert (Univ. of Pennsylvania, USA): « *Some new asymptotic problems in partition theory* »

Suppose the partitions of an integer n are restricted in the set of parts that can be used, and in the set of multiplicities of those parts. We'll speak about quantifying the influence of these restrictions on the asymptotic growth of the partition function, and will raise some unsolved problems.

Participants

ADDARIO-BERRY, Louigi – *McGill University*
AGUECH, Rafik – *Faculté des Sciences de Monastir (Tunisia)*
AKHAVI, Ali – *Univ. de Caen, France*
BANDERIER, Cyril – *LIPN – Univ. Paris 13*
BARRERA, Javiera – *Universidad Adolfo Ibanez, Chile*
BARYSHNIKOV, Yuliy – *Bell Labs, USA*
BASSINO, Frédérique – *LIPN – Univ. Paris 13*
BODINI, Olivier – *UMPC, Paris*
BOURDON, Jérémie – *LINA, Nantes et INRIA Rennes, France*
BROUTIN, Nicolas – *Inria, Algorithms*
CESARATTO, Eda – *Universidad Nac. de Gral. Sarmiento, Argentina*
CHASSAING, Philippe – *Institut Elie CARTAN, Nancy, France*
CHAUVIN, Brigitte – *LMV Univ. Versailles & Inria Rocquencourt*
CICHON, Jacek – *Univ. of Wroclaw, Institute of Computer Science, Poland*
CLÉMENT, Julien – *CNRS, Caen, France*
CREIGNOU, Nadia – *LIF, Marseille, France*
DARRASSE, Alexis – *LIP6 - UPMC, Paris*
DAUDÉ, Hervé – *Université d'Aix-Marseille 1, France*
DAVID, Julien – *LIGM, Univ. Paris Est*
DEVROYE, Luc – *McGill University*
BOUTTIER, jérémié – *CEA, France*
DRMOTA, Michael – *TU Wien, Institute of Discrete Mathematics and Geometry*
DUCHI, Enrica – *LIAFA, Paris 7*
DUCHON, Philippe – *LaBRI, Univ. Bordeaux 1*
FAWZI, Omar – *McGill University*
FILL, Jim – *The Johns Hopkins University, Department of Applied Mathematics and Statistics*
FLAJOLET, Philippe – *Inria, Algorithms*
FUSY, Eric – *LIX, Polytechnique, France*
GAFNI, Ayla – *INRIA*
GARDY, Daniele – *Univ. de Versailles St-Quentin, France*
GENITRINI, Antoine – *Univ. de Versailles St-Quentin, France*
GIAMBRUNO, Laura – *Università di Palermo, Italy*
GITTENBERGER, Bernahrd – *TU Wien, Institute for Discrete Math. and Geometry*
GNEDIN, Sasha – *Utrecht University*
GOLEBIEWSKI, Zbigniew – *Univ. of Wroclaw, Institute of Computer Science, Poland*
GRÜBEL, Rudolf – *Leibniz Universitaet Hannover, Germany*
GÉRIN, Lucas – *LIAFA, France*
HEUBERGER, Clemens – *TU Graz*
HOLMES, Susan – *Stanford University*
HOLMGREN, Cecilia – *Uppsala Univ., Sweden*
HWANG, Hsein-Kuei – *Academia Sinica, Taiwan*
JACQUET, Philippe – *EPI Hipercom, INRIA, France*
JANSON, Svante – *Uppsala Univ., Sweden*
KALPATHY, Ravi – *The George Washington Univ., Washington D.C.*
KENNICHE, Hichem – *LIPN, Paris-Nord, France*
DIACONIS, Persi – *Stanford University*
KLONOWSKI, Marek – *Wroclaw University of Technology, Poland*
KNOPFMACHER, Arnold – *Univ. of the Witwatersrand, South Africa*
KRAUS, Veronika – *TU Wien - Institute of Discrete Mathematics and Geometry*
LAVAUULT, Christian – *LIPN – univ. Paris 13, France*
LE BARS, Jean-Marie – *Univ. de Caen, GREYC, France*
LLADSER, Manuel – *Univ. of Colorado, USA*
LOUCHARD, Guy – *Univ. Libre de Bruxelles, Belgium*
LUMBROSO, Jérémie – *Inria, Algorithms*
MADRITSCH, Manfred – *Univ. Caen, GREYC, France*
MAHMOUD, Hosam – *George Washington University*
MAIRESSE, Jean – *Cnrs, Univ. Paris 7, France*
MARCKERT, Jean-François – *LaBRI, Univ. Bordeaux 1*
MARTINEZ, Conrado – *Univ. Politècnica de Catalunya*
MAUDUIT, Christian – *Marseille, France*
MOLINERO, Xavier – *Universitat Politècnica de Catalunya*
MORENO, Eduardo – *Universidad Adolfo Ibanez, Chile*
MUNAGI, Augustine – *Univ. of the Witwatersrand, South Africa*
NEBEL, Markus – *Univ. of Kaiserslautern, Germany*
NEININGER, Ralph – *Goethe Univ. Frankfurt, Germany*
NICODÈME, Pierre – *LIX, CNRS, France*
NOY, Marc – *Universitat Politècnica de Catalunya*
PANARIO, Daniel – *Carleton University, Canada*
PANHOLZER, Alois – *TU Wien, Institute of Discrete Mathematics and Geometry*
PITTEL, Boris – *The Ohio State University, USA*
PIVOTEAU, Carine – *LIP6, Paris, France*
POUYANNE, Nicolas – *Univ. de Versailles Saint Quentin, France*
PRODINGER, Helmut – *Univ. of Stellenbosch, South Africa*
RAVELOMANANA, Vlady – *LIPN – univ. Paris 13, France*
ROESLER, Uwe – *Univ. of Kiel, Germany*
ROUSSEL, Olivier – *LIP6, Paris, France*
SALVY, Bruno – *Inria, Algorithms*
SANKOFF, David – *Ottawa University, Canada*
SCHAEFFER, Gilles – *LIX, Polytechnique, France*
SEDGEWICK, Bob – *Princeton Univ., USA*
SORIA, Michèle – *LIP6/UPMC, France*
SPRUGNOLI, Renzo – *Dip. Sistemi e Informatica - Università' di Firenze*
SZPANKOWSKI, Wojciech – *Purdue University, USA*
VALLÉE, Brigitte – *CNRS, Caen, France*
VAN LEEUWAARDEN, Johan – *Eindhoven Univ. of Technology and EURANDOM, Netherlands*
VERA, Antonio – *CNRS/INRIA/Nancy-Université, France*
VIOLA, Alfredo – *Universidad de la Republica, Montevideo, Uruguay*
WARD, Mark – *Purdue University, USA*
WILF, Herbert – *Univ. of Pennsylvania, USA*