Abstracts of the 24^{th} *International Meeting on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms*

AofA 2013

May 27-31, 2013, Menorca, Spain

Abstracts of the Keynote Lectures

(in alphabetical order)

What is the information content of an algorithm?

Joachim M. Buhmann
ETH, Switzerland

Abstract

Algorithms are exposed to randomness in the input or noise during the computation. How well can they preserve the information in the data w.r.t. the output space? Algorithms especially in Machine Learning are required to show robustness to input fluctuations or randomization during execution. This talk elaborates a new framework to measure the "informativeness" of algorithmic procedures and their "stability" against noise. An algorithm is considered to be a noisy channel which is characterized by a generalization capacity (GC). The generalization capacity objectively ranks different algorithms for the same data processing task based on the bit rate of their respective capacities. The problem of grouping data is used to demonstrate this validation principle for clustering algorithms, e.g. k-means, pairwise clustering, normalized cut, DBSCAN and dominant set clustering. Our new validation approach selects the most informative clustering algorithm, which filters out the maximal number of stable, task-related bits relative to the underlying hypothesis class. The concept also enables us to measure how many bit are extracted by sorting algorithms when the input and thereby the pairwise comparisons are subject to fluctuations.

Combinatorial Markov Chains

Rudolf Grübel
Leibniz Universität Hannover, Germany

Abstract

By a combinatorial Markov chain we mean a Markov chain $X = (X_n)_{n\in\mathbb{N}}$ that is adapted to a combinatorial family $\mathbb{F}$ in the sense that

$$P(X_n \in \mathbb{F}_n) = 1, \quad P(X_n = y) > 0 \text{ for all } y \in \mathbb{F}_n,$$

where $\mathbb{F}_n$ denotes the set of combinatorial objects that have size n. Such processes typically arise in connection with sequential algorithms that transform an input sequence $\eta_1, \eta_2, \ldots$ into an output sequence $y_1, y_2, \ldots$ if $y_{n+1} \in \mathbb{F}_{n+1}$ depends on y_n and η_{n+1} only.

We review various boundary concepts for combinatorial Markov chains, with emphasis on the interplay with the Analysis of Algorithms: Boundaries may be used to obtain strong limit theorems for functionals of interest, but the existence of an algorithm generating X from i.i.d. input data can also be used to obtain the boundaries.

Several examples will be given, ranging from didactical (Pólya urn) to industrial strength (various trees).

Phase transition in random discrete structures

Mihyun Kang
Technische Universität Graz, Austria

Abstract

The phase transition is a phenomenon observed in mathematics and natural sciences in many different contexts. It deals with a sudden change in the properties of a large structure caused by altering a critical parameter. The phase transition in random discrete structures (e.g. random graphs, random satisfiability problems, Ising/Potts model, percolation) has captured the attention of many scientists in recent years. In particular, the phase transition in random graphs was first discussed in 1959 by Erdős and Rényi in a series of papers. Since their seminal work, the phase transition in random graphs have been extensively studied. This talk will be a gentle introduction to the phase transition in random discrete structures, in particular the phase transition in various random graph models, with focus on proof techniques including analytic, combinatorial, and probabilistic methods.

Concentration inequalities and the entropy method

Gábor Lugosi
ICREA and Pompeu Fabra University, Spain

Abstract
We discuss concentration inequalities that estimate deviations of functions of independent random variables from their expectation. Several methods have been developed for proving such inequalities, such as martingale methods, Talagrand's induction method, or Marton's transportation-of-measure technique. In this talk we focus on the so-called entropy method that is based on some simple information-theoretic inequalities. We present the main steps of the proof technique and discuss various inequalities and some applications.

Analysis of Algorithms and Language Theory

Cyril Nicaud
LIGM, Univ. Paris-Est, France

Abstract
This talk deals with rational languages, which are classical objects of the theory of formal languages. We will especially focus on finite state automata, which are one of the many ways to describe rational languages: automata are simple abstract machines and, as such, are often the best algorithmic representations for these languages.

As often in computer science, the worst case complexities of algorithms on finite state automata are well understood, but little is known about their average or typical behaviors. In this talk, we will look at automata as random combinatorial objects and survey the results obtained in the last decade on random generation and average case analysis of algorithms in this field.

Going After the k-SAT Threshold

Konstantinos Panagiotou
University of Munich, Germany

Abstract

Random k-SAT is the single most intensely studied example of a random constraint satisfaction problem. But despite substantial progress over the past decade, the threshold for the existence of satisfying assignments is not known precisely for any $k \geq 3$. The best current results, based on the second moment method, yield upper and lower bounds that differ by an additive $k \cdot \frac{\ln 2}{2}$, a term that is unbounded in k (Achlioptas, Peres: STOC 2003). The basic reason for this gap is the inherent asymmetry of the Boolean value 'true' and 'false' in contrast to the perfect symmetry, e.g., among the various colors in a graph coloring problem. In this talk we will discuss a new asymmetric second moment method that allows us to tackle this issue. This technique enables us to compute the k-SAT threshold up to an additive $\ln 2 - \frac{1}{2} + O(1/k) \approx 0.19$. Independently of the rigorous work, physicists have developed a sophisticated but non-rigorous technique called the "cavity method" for the study of random CSPs (Mézard, Parisi, Zecchina: Science 2002). Our result matches the best bound that can be obtained from the so-called "replica symmetric" version of the cavity method, and indeed our proof directly harnesses parts of the physics calculations.

This is joint work with Amin Coja-Oghlan.

Combinatorial Systems and Newton iteration

Michèle Soria
LIP6, UPMC, France

Abstract

We present a quadratic iterative Newton method for computing structures defined by systems of recursive combinatorial equations. This iteration transfers to truncations of the corresponding generating series in quasi-optimal complexity. It also transfers to a numerical scheme that converges unconditionally to the values of the generating series inside their disk of convergence. These results provide important subroutines in random generation.

This is joint work with Carine Pivoteau and Bruno Salvy.

---

50 years of Linear Probing Hashing

Alfredo Viola
Universidad de la República, Uruguay

Abstract

Linear probing hashing is at the core of Analysis of Algorithms by its mathematical and historical interest. This key role is clearly presented in the introduction of a wonderful survey by Flajolet and Chassaing in 2003 oriented to french students, that can be summarized as follows:

> "Discrete and continuous mathematics willingly and harmoniously encounter and complement. We would like to illustrate this thesis by presenting a classical problem with several ramifications – the analysis of linear probing hashing. This example is typical of the analysis of algorithms, a topic pioneered by Knuth and which is at the intersection of computer science, combinatorics, and probability theory."

Historical and scientific motivations include *"questions asked by Ramanujan to Hardy in 1913, a summer work in 1962 by Knuth that is at the origin of the analysis of algorithms in computer science, the research in combinatorics done by the statistician Kreweras, several encounters with the model of random graphs by Erdös and Rényi, some complex and asymptotic analysis, trees generated by specific Galton-Watson processes, and, to conclude, a bit of processes like the ineffable Brownian motion!"* All this contributes to a *"very precise understanding of a very simple discrete random problem".*

In this talkk I will survey the history of linear probing hashing, and present some new work in progress.

Abstracts of the Contributed Talks

(in alphabetical order)

Persistence Homology for Brownian Bridges

Yuliy Baryshnikov
Univ. Illinois at Urbana-Champaign, USA

Abstract

Persistence Homology (PH) was introduced in the context of shape reconstruction, but became one of the key tools of the Topological Data Analysis, a novel algorithmic framework of understanding robust patterns in high-dimensional point clouds, large networks and other areas where the ambient space lacks structures enabling the classical approaches like PCA or dimensionality reduction. Persistence Homology is an algebraic apparatus allowing one to relate topological invariants of point clouds at different scales in a systematic way.

Informally, PH in dimension k associates to an increasing filtration of topological spaces (typically given by $M_t := \{f \leq t\}$) a point process on the half-plane $\{t > s\}$, with a point (s, t) corresponding to a k-dimensional homology class emerging at time s and disappearing at the time t.

Understanding the PH on random samples and functions is one of the key problems in the area. One specific class of problems is to analyze the behavior of PH on f being trajectories of Gaussian fields, the archetypal background noise in many applied problems, from cosmology to magnetic resonance imaging.

In this talk I will describe the structure of the PH for one-dimensional

Brownian motion and some of related processes. In particular, we prove a $(t-s)^3$ asymptotic of the PH density near the diagonal.

Algorithms helping Enumeration in Pattern Avoiding Permutations

Miklós Bóna
Univ. Florida, USA

Abstract
We will present an example when an algorithm enables us to answer the usually very difficult question of why a permutation pattern is easier to avoid than another one. We will also consider two different notions of stack sorting, one well-known, the other one not so much, and see how they are connected to pattern avoidance and enumeration. An unexpected symmetry in t-stack sortable permutations will also be shown, together with unimodality results, and questions on log-concavity and the real roots property.

Counting irreducible maps via substitution and bijections

Jérémie Bouttier
Institut de Physique Théorique, CEA Saclay, France

Abstract
Planar maps (connected graphs embedded in the plane) are subject to a constantly renewed interest in enumerative combinatorics, and enjoy applications in algorithmics (random sampling, drawing) and theoretical physics (two-dimensional quantum gravity). In this talk, we consider irreducible maps, i.e. maps whose smallest cycles are non-separating. Irreducible triangulations and quadrangulations were first counted respectively by Tutte and by Mullin-Schellenberg using a substitution approach, and later bijectively by Fusy, Poulalhon and Schaeffer. In this talk, we describe a general approach to enumerating irreducible maps both by substitution and by bijections. Our "master tool" will be the so-called slice decomposition, which consists in cutting a map along shortest paths.

Bootstrap Percolation on Periodic Trees

Milan Bradonjic
Bell Labs, Alcatel-Lucent

Abstract
We study bootstrap percolation with the threshold parameter θ and the initial probability p on infinite periodic trees that are defined as follows. Each node of a tree has degree selected from a finite predefined set of non-negative integers; starting from any node all nodes at the same graph distance have the same degree. We show the existence and derive a closed form relation for the nontrivial critical threshold $p_f \in (0, 1)$ such that for almost every initial configuration: (i) if $p > p_f$ then a periodic tree becomes eventually fully active; (ii) if $p < p_f$ then a periodic tree does not become eventually fully active.

This is joint work with Iraj Saniee.

Label patterns in mappings

Marie-Louise Bruner
TU Wien, Austria

Abstract
In this talk, we consider mappings from the set $\{1, 2, \ldots, n\}$ onto itself. Structural parameters of random mappings (respectively of the corresponding functional graphs) have been studied intensively in the literature but label quantities have not yet received any attention.

A combinatorial decomposition of the mapping graph with respect to a specific node allows us to establish recurrence relations for certain interesting label quantities. These recurrence relations involve the corresponding quantities for Cayley trees. A generating functions approach then leads to quasi-linear or linear partial differential equations and yields explicit generating functions for some interesting label quantities. As an example we mention the number of runs and local minima in mappings; both are enumerated by surprisingly simple sequences related to the Stirling numbers of the second kind.

This talk is based on joint work with Alois Panholzer.

Coefficients of positive algebraic functions

Michael Drmota
TU Wien, Austria

Abstract

We study systems of polynomial equations $a_j(x) = P_j(x, a_1(x), ...a_d(x))$, where P_j has only non-negative coefficients, and show that the asymptotic behaviour of the coefficients $[x^n]a_j(x)$ is of very special form. The main problem is to determine the possible structure of the dominating singularities which are either of the form $(1 - x/\rho)^{1/2^k}$ for some $k \geq 1$ or of the form $(1 - x/\rho)^{-m/2^k}$, where $m \geq 1$ and $k \geq 0$.

This is a generalization of the so-called *Drmota-Lalley-Woods-Theorem* that applies only to stronlgy connected systems of equations.

This is joint work with Cyril Banderier.

Quad-K-d Trees

Amalia Duch
Universitat Politècnica de Catalunya, Spain

Abstract

We introduce a new data structure for the storage of multidimensional points, the Quad-K-d tree (or simply qKd tree tree), which is a natural generalization of both point quad trees and K-d trees at the same time. This new general purpose, hierarchical, multidimensional data structure can be tuned by means of insertion heuristics to obtain time and space trade-offs to fit application needs. We explore three such heuristics, and show experimentally –with uniformly and non-uniformly distributed data– their competitive performance. Our results indicate that the qKd tree tree is a flexible data structure that can be tailored to the resource requirements of a given application. Moreover, Quad-K-d Trees might become a formal framework for the analytic study of multidimensional data structures akin K-d trees and quad trees –a challenging open line of interest for the AofA community.

Partitions with Distinct Multiplicities of Parts: On an "Unsolved Problem" Posed by Herbert Wilf

Jim Fill
Department of Applied Mathematics and Statistics, The Johns Hopkins University, USA

Abstract
Wilf's Sixth Unsolved Problem asks for any interesting properties of the set of partitions of integers for which the (nonzero) multiplicities of the parts are all different. We refer to these as *Wilf partitions*. Using $f(n)$ to denote the number of Wilf partitions, we establish lead-order asymptotics for $\ln f(n)$.

This is joint work with Svante Janson and Mark Daniel Ward.

Part Sizes of Smooth Supercritical Compositional Structures

Jason Gao
Carleton U., Canada

Abstract
We define the notion of smooth supercritical compositional structures. Two well-known examples are compositions and graphs of given genus. The "parts" of a graph are the subgraphs that are maximal trees. We show that large part sizes have asymptotically geometric distributions. This leads to asymptotically independent Poisson variables for numbers of various large parts. In many cases this leads to asymptotic formulas for the probability of being gap free and for the expected values of the largest part sizes, number of distinct parts and number of parts of multiplicity k.

The talk is based on a recent joint work with E.A. Bender.

Statistics on restricted lambda-terms

Daniele Gardy
PRiSM, Univ. Versailles Saint-Quentin, France and DMG, TU Wien, Austria

Abstract
Although lambda-terms are the fundamental objects of lambda calculus, their statistical behaviour is still largely unknown. E.g., counting the number of (unrestricted) lambda-terms of specific size is still an open question, as well as their statistical behaviour or probabilistic information on such basic properties as being normalisable (which amounts to forbidding the occurence of a specific pattern).

After recalling the definition of lambda-terms as enriched Motzkin trees (i.e. for each leaf, we may choose to add a pointer to one of its unary node ancestors), and the problems involved in their enumeration (the generating function for unrestricted lambda-terms has null radius of convergence), we turn to restricted classes (bounded number of unary nodes; bounded unary

height, ...). For those classes, although the generating function exhibits an unexpected behaviour, it has non-null radius and we can obtain asymptotic counting results, as well as statistical information on the shape of such terms, most notably about the probability that a random term is normalisable.

Associative and commutative tree representations for Boolean functions

Bernhard Gittenberger
TU Wien

Abstract

Since the 90's, several authors have studied a probability distribution on the set of Boolean functions on n variables induced by some probability distributions on formulas built upon the connectors *And* and *Or* and the literals $\{x_1, \bar{x}_1, \ldots, x_n, \bar{x}_n\}$. These formulas rely on plane binary labelled trees. We extend these results, in particular the relation between the probability and the complexity of a Boolean function, to other models of formulas: non-binary or non-plane labelled trees. This includes the natural tree class where associativity and commutativity of the connectors *And* and *Or* are realised.

This is joint work with Antoine Genitrini, Veronika Kraus and C'ecile Mailler.

Generalized Erdős-Turán laws for the order of random permutation

Alexander Gnedin
Queen Mary, University of London, UK

Abstract

The Erdős-Turán law states that the logarithm of the order (l.c.m. of the cycle lengths) of the uniformly random permutation has asymptotically normal distribution.

Arratia and Tavaré extended this result to permutations with Ewens' distribution.

We consider a wider class of random permutations derived by sampling from stick-breaking partitions of the unit interval.

Like in the classical case, the logarithm of the order of permutation can be approximated by the sum of logarithms of the cycle lengths.

Using methods of the renewal theory we show that, depending on the properties of the stick-breaking factor, normal and other stable distributions can occur in the limit.

Joint work with Alexander Iksanov and Alexander Marynych.

Analysis of the binary asymmetric joint sparse form

Clemens Heuberger
Alpen-Adria-Universität Klagenfurt and TU Graz, Austria

Abstract

We consider redundant binary joint digital expansions of integer vectors. The redundancy is used to minimise the Hamming weight, i.e., the number of non-zero digit vectors. This leads to efficient linear combination algorithms in abelian groups, which are for instance used in elliptic curve cryptography.

If the digit set is a set of contiguous integers containing the zero, a special syntactical condition is known to minimise the weight. We analyse the optimal weight of all non-negative integer vectors with maximum entry less than N. The expectation and the variance are given with a main term and a periodic fluctuation in the second order term. Finally, we prove asymptotic normality. (Joint work with Sara Kropf.)

Bootstrap percolation on Galton-Watson trees

Cecilia Holmgren
Stockholm University, Sweden

Abstract

Bootstrap percolation is a type of cellular automaton which has been used to model various physical phenomena, such as ferromagnetism. For each natural number r, the r-neighbour bootstrap process is an update rule for vertices of a graph in one of two states: 'infected' or 'healthy'. In consecutive rounds, each healthy vertex with at least r infected neighbours becomes itself infected. Percolation is said to occur if every vertex is eventually infected.

Usually, the starting set of infected vertices is chosen at random, with all vertices initially infected independently with probability p. In that case, given a graph G and infection threshold r, a quantity of interest is the critical probability, $p_c(G, r)$, at which percolation becomes likely to occur. In collaboration with Bollobas, Gunderson, Janson and Przykucki I look at infinite trees and, answering a problem posed by Balogh, Peres and Pete, we show that for any $b \geq r$ and for any $\epsilon > 0$ there exists a tree T with branching number $br(T) = b$ and critical probability $p_c(T, r) < \epsilon$. However, this is false if we limit ourselves to the well-studied family of Galton–Watson trees. We show that for every $r \geq 2$ there exists a constant $c_r > 0$ such that if T is a Galton–Watson tree with branching number $br(T) = b \geq r$ then

$$p_c(T, r) > \frac{c_r}{b} e^{-\frac{b}{r-1}}.$$

We also show that this bound is sharp up to a factor of $O(b)$ by giving an explicit family of Galton–Watson trees with critical probability bounded from above by $C_r e^{-\frac{b}{r-1}}$ for some constant $C_r > 0$.

Periodic oscillations of the variance of trie statistics and related structures

Hsien-Kuei Hwang
Academia Sinica, Taiwan

Abstract

I will present analytic tools for dealing with the asymptotics of general trie statistics, which are particularly advantageous for clarifying the asymptotic variance, notably the underlying periodic oscillations. Many concrete examples are discussed for which new Fourier expansions are given. The tools are also useful for other splitting processes with an underlying binomial distribution. This talk is based on joint work with Michael Fuchs and Vytas Zacharovas.

Green Leader Election

Philippe Jacquet
Bell Labs France

Abstract

We consider the problem of leader election in a collision channel where a subset of n transmitters among a total population of N users, contend for

the election of a leader. The problem is classic and in general the performance of a leader election algorithm is given in the average delay to an election. However no analysis has been done in term of the total number of transmissions required for the election. The classic Part and Try algorithm performs an election in an average delay of $\log n$ but with $O(n)$ average number of transmissions. This leads to a far too large energy waste if n counts in millions in a wireless network. We introduce a new leader election algorithm that is much more performant than the part and try algorithm. The algorithm has the property to have a reduced delay in $k \log\log N$ instead of $\log N$. More importantly the algorithm has the property to have a very reduced energy consumption since it requires $O(N^{1/k})$ transmissions instead of $O(N/k)$, per election, k being a tuning parameter.

Weighted random staircase tableaux

Svante Janson
Uppsala University, Sweden

Abstract

We study a relatively new combinatorial structure called staircase tableau. We consider a general model of random staircase tableaux in which symbols that appear may have arbitrary positive weights. Under this general model we derive a number of results, including limiting laws for the number of appearances of the symbols. This generalizes earlier results for specific values of the weights. Furthermore, our generality allows us to analyze the structure of random staircase tableaux and we obtain several results in this direction.

One of the tools we use is generating functions. This leads to a two-parameter family of polynomials that are generalizations of the classical Eulerian polynomials.

There are also connections to an urn model studied by a number of researchers, including Philippe Flajolet.

Joint work with Paweł Hitczenko.

Towards the distribution of the size of the largest non-crossing matchings in random bipartite graphs

Marcos Kiwi
U. Chile

Abstract

We consider the following question: When a randomly chosen regular bipartite multi-graph is drawn in the plane in the "standard way", what is the distribution of its maximum size planar matchings (sets of non-crossing disjoint edges)? The problem is a generalization of the Longest Increasing Sequence (LIS) problem (also called Ulam's problem).

We present combinatorial identities which relate the number of r-regular bipartite multi-graphs with maximum planar matching (maximum planar subgraph) of at most d edges to a signed sum of restricted lattice walks in $\mathbb{Z}^d$, and to the number of pairs of standard Young tableaux of the same shape and with a "descend-type" property. Our results are obtained via generalizations of combinatorial proofs through which Gessel's identity can be derived (an identity that is crucial in the derivation of a bivariate generating function associated to the distribution of LISs, and key to the analytic attack on Ulam's problem).

For the very special $d = 2$ case, work by Banderier and Flajolet, allow us to obtain generating functions for some of the quantities of interest. A problem this work leaves wide open is how to handle the $d > 2$ case.

This is joint work with Martin Loebl (Charles U.)

Analysis of urn models with multiple drawings

Markus Kuba

FH Technikum Wien, Austria

Abstract

We discuss urn models, where at each discrete time step m balls are drawn at random from the urn. Balls are added/removed to the urn according to the inspected multiset of balls, generalizing the well known classical model, case m = 1, where only one ball is drawn. An analysis of a Polya urn model, and a Friedman urn will be presented. The connection to a few other models is sketched, in particular, a related result for a diminishing urn model.

This talk is mainly based on joint work with May-Ru Chen, Hosam Mahmoud, and Alois Panholzer.

Sum of positions of records in random permutations: A precise analysis

Guy Louchard

Université Libre de Bruxelles, Belgium

Abstract
This statistic has been the object of recent interest in the litterature. Using the saddle point method, we obtain from the generating function of the sum of positions of records in random permutations and Cauchy's integral formula, asymptotic results in central and non-central regions. In the non-central region, we derive asymptotic expansions generalizing some results by Kortchemski. In the central region, we obtain a limiting distribution related to Dickman's function.

Some Fast Buffon Machines

Jérémie Lumbroso
GREYC, Université de Caen

Abstract
Buffon machine are a general framework introduced by Flajolet, Pelletier and Soria (2011) to simulate discrete probability laws (Bernoulli of $1/\pi$, geometric of $\exp(-1)$, Poisson of $3/27...$) exactly without ever making any arithmetic operations: all computations are "done probabilistically" through a sequence of coin flips.

We will briefly recall this framework, and will extend it with a few additional constructions which either optimize existing laws, or allow for the simulation of new ones. The algorithms are elegant and amusing, but may also be competitively efficient, as we may show with the discrete uniform or Poisson laws.

This a work in progress with Michèle Soria.

Some Node Degree Properties of Series-Parallel Graphs Evolving Under a Stochastic Growth Model

Hosam M. Mahmoud
The George Washington University, USA

Abstract
We introduce a natural growth model for directed series-parallel (SP) graphs and look at some of the graph properties under this stochastic model.

Specifically, we look at the degrees of certain types of nodes in the random SP graph. We examine the degree of a pole and will find its exact distribution, given by a probability formula with alternating signs.

We also prove that, for a fixed value s, the numbers of nodes of outdegree $1, \ldots, s$ asymptotically have a joint multivariate normal distribution.

Polya urns will systematically provide a working tool.

Smoothing Equations for Large Pólya Urns

Cecile Mailler
Univ.Laboratoire de Mathmatiques de Versailles, Versailles Saint-Quentin, France

Abstract

This talk will focus on large two-colour Pølya urns. From the study of the asymptotic behaviour of such an urn arises a random variable denoted by W. The underlying tree structure of the urn permits to see W as the solution in law of a fixed point equation, from which we can deduce information about its moments, or about the existence of a density. This work can be done on the discrete urn itself, or on its continuous time embedding. Though the two variables W (arisen from discrete or continuous time) are different, they are related by connexions, which often permit to translate results from one W to the other. This work is a collaboration with Brigitte Chauvin and Nicolas Pouyanne.

The height of the Lyndon tree

Lucas Mercier
Institut Elie Cartan, France

Abstract

We consider the set $\mathcal{L}_n$ of n-letters long Lyndon words on the alphabet $\mathcal{A} = \{0, 1\}$. For a random uniform element L_n of the set $\mathcal{L}_n$, the binary tree $\mathfrak{L}(L_n)$ obtained by successive standard factorization of L_n and of the factors produced by these factorization is the *Lyndon tree* of L_n. We prove that the height H_n of $\mathfrak{L}(L_n)$ satisfies

$$\lim_n \frac{H_n}{\ln n} = \Delta,$$

in which the constant Δ is solution of an equation involving large deviation rate functions related to the asymptotics of Eulerian numbers ($\Delta \simeq 5.092\ldots$). The convergence is the convergence in probability of random variables.

This is joint work with Philippe Chassaing.

On unbalanced Pólya urns: Analytic Combinatorics strikes again

Basile Morcrette
Inria – Université Pierre et Marie Curie

Abstract

Take an urn containing balls of two colors, say black and white. At each time, pick a ball at random, put it back in the urn and add black and/or white balls with regards to some replacement rule. This is known as a Pólya urn process. Starting in 2005, the work of Flajolet *et al.* has shown that analytic combinatorics is a powerful tool to encode and study the behavior of such models in the *balanced* case, i.e. when the total number of added balls at each time is constant.

We lead the first steps of a general analytic-combinatorial approach of *non necessarily balanced* urns. In this case, any urn process is represented by some generating function verifying a partial differential equation. We show it is possible to access an explicit form of the generating function when some first order differential equation is solvable.

With our method, it is even possible —on some examples, such as Knuth's strings— to automatically obtain an expression of the probability generating function.

This work, inspired by Philippe Flajolet, is currently in progress. We will present preliminary results; but more advanced topics, such as limit laws, are not yet attained.

Work in collaboration with Philippe Dumas.

Average Case and Distributional Analysis of Java 7's Dual Pivot Quicksort

Markus E. Nebel
Kaiserslautern University, Germany

Abstract

In this talk we discuss a new Quicksort variant which recently became the standard sort of Oracle's Java 7. Compared to classic Quicksort it uses two pivots and proved superior to classic Quicksort (based on experiments) even if previous theoretical considerations discouraged the use of multiple pivots.

By performing a classical average-case analysis for the expected number of comparisons, we are able to (partly) explain the new algorithm's success. Furthermore, we show lower bounds for improvements possible by optimizing the new partitioning algorithm as well as results on the distribution of costs derived by the contraction method. This is joint work with Ralph Neininger and Sebastian Wild.

Pólya urns via the contraction method

Ralph Neininger
J.W. Goethe University, Frankfurt a.M., Germany

Abstract

An approach to analyze the asymptotic behavior of Pólya urns is proposed based on the contraction method. For this a combinatorial discrete time embedding of the evolution of the composition of the urn into random rooted trees is developed. A decomposition of the trees leads to a system of recursive distributional equations which capture the distributions of the numbers of balls of each color. Ideas from the contraction method are used to study such systems of recursive distributional equations asymptotically. This approach is applied to a couple of concrete Pólya urns that lead to limit laws with normal limit distributions, with non-normal limit distributions and with asymptotic periodic distributional behavior.

Based on a joint paper with M. Knape, available at

`http://arxiv.org/abs/1301.3404`

At the time of the talk I hope to know more than what is in the paper.

Why Clump Analysis helps understanding DNA evolution?

Pierre Nicodème
LIPN - Univ. Paris 13, France

Abstract

Recent work studied the waiting time needed for a k-mer (word of length k) to appear in a sequence $S(t)$ of length n of DNA under a model of evolution, assuming that the k-mer is not present at time t=0 in the sequence $S(0)$. More specifically, the waiting time is the inverse of the probability $P(n)$ that the k-mer occurs at time $t = 1$ in the sequence $S(1)$ while not occurring

in $S(0)$. Behrens and Vingron (2010), with some simplifying hypotheses, computed by inclusion-exclusion on the sequence $S(1)$ the waiting times for $n = 1000$ and all 5-, 6-, .., 10-mers under a Bernoulli and a Markov model. Behrens et al. (2012) used a squared DNA alphabet of size 16 and an automata that handles simultaneous traversals of $S(0)$ and $S(1)$; they computed in full generality the waiting times for the same set of k-mers (k from 5 to 10). We use a heuristics that allows considering only the sequence $S(0)$; we can then use clump analysis (Bassino et al. -2006), either by formal languages decompositions, or by automata. This leads to two generating functions; the quotient of their Taylor coefficients of order n approximates the probability $P(n)$. Finally an easy analysis of singularity leads to a quasi-linear behaviour of $P(n)$ for a very large range of values of n, and, as an immediate consequence, a hyperbolic behaviour of the waiting time. Previous work of Nicodème et al (Motif Statistics - 2002) allows to extend the methods to the Markov case.

Phase Transition of Symmetric Inhomogeneous Random Graphs

Elie de Panafieu
LIAFA, Univ. Paris 7, France

Abstract
Inhomogeneous graphs were introduced by [Söderberg,2002] as graphs where vertices come in different types and the probability of realizing an edge depends on the types of its terminal vertices. They were then extensively studied in [Bollobás,Janson,Riordan,2007].

This general model encodes many polynomial decision problems and links the phase transition corresponding to the birth of the complex component of inhomogeneous graphs with the probability of satisfiability of the decision problems.

New accurate results on this phase transition for a sub-class of inhomogeneous random graphs are derived using tools from analytic combinatorics. As applications, we present a new proof of an already known result from [Pittel,Yeum,2010] on the probability of 2-coloriability, and new results on the probability of satisfiability of quantified 2-XOR-formulas, already investigated by [Creignou,Daudé,Egly,2007].

This is joint work with Vlady Ravelomanana.

Analytic Combinatorics in Several Variables

Robin Pemantle
University of Pennsylvania, USA

Abstract

The analytic framework for estimating coeffiecients of a generating function is the same in many variables as in one variable: evaluate Cauchy's integral by manipulating the contour into a "standard" position. That being said, the manipulation of contours in several complex variables is a very different story. The geometry in several variables can be much more complicated.

This talk, drawing on the recent book of the same title, surveys analytic methods for extracting asymptotics from multivariate generating functions. I will try to give an idea of the main pieces of the puzzle. In particular, I will try to explain in pictures the roles of Morse theory, complex algebraic geometry and hyperbolicity in the asymptotic evaluation of integrals, and hence of coefficients of generating functions.

This is joint work with Mark C. Wilson.

On the probability of planarity of a random graph near the critical point

Vlady Ravelomanana
LIAFA, Univ. Paris 7, France

Abstract

Consider the uniform random graph $G(n, M)$ with n vertices and M edges. Erdős and Rényi (1960) conjectured that the limit

$$\lim_{n\to\infty} \Pr\{G(n, \tbinom{n}{2}) \text{ is planar}\}$$

exists and is a constant strictly between 0 and 1. Łuczak, Pittel and Wierman (1994) proved this conjecture and Janson, Łuczak, Knuth and Pittel (1993) gave lower and upper bounds for this probability. In this paper we determine the exact probability of a random graph being planar near the critical point $M = n/2$. For each λ, we find an exact analytic expression for

$$p(\lambda) = \lim_{n\to\infty} \Pr G(n, \tfrac{n}{2}(1 + \lambda n^{-1/3}))isplanar.$$

In particular, we obtain $p(0) \approx 0.99780$. We extend these results to classes of graphs closed under taking minors. As an example, we show that the probability of $G(n, \frac{n}{2})$ being series-parallel converges to 0.98003. For the sake of completeness and exposition we reprove in a concise way several basic properties we need of a random graph near the critical point.

This is joint work with Marc Noy and Juanjo Rué.

Implicit Species at the Basis of Analytic Combinatorics

Bruno Salvy
INRIA, (France)

Abstract
The automation of the symbolic method developed in Flajolet & Sedgewick's "Analytic Combinatorics" requires algorithms to decide whether a system of combinatorial equations is meaningful or not. We establish a dictionary between Joyal's species theory and the symbolic method. From there, simple algorithms follow. We then extend those to the case of ordered structures and integral operators. This is joint work with Carine Pivoteau and Michèle Soria.

Analytic Combinatorics for the Masses

Robert Sedgewick
Princeton, USA

Abstract
In the 21st century, most professors are still teaching in virtually the same way they were taught and their teachers were taught, stretching back at least to the 19th century. This situation is likely to change, and soon. Technology is transforming (if not threatening to overwhelm) higher education, as MOOCs and online content become widely available. University students seeking to learn a topic who now have little if any choice are about to be presented with a vast array of choices. What student would not want to swap a tired professor writing slowly on a chalkboard for a well-produced series of videos and associated content, given by a world leader in the field? This imminent change raises a host of fascinating and far-reaching questions. This talk will address a few of these questions, but focus on one in particular: Can an advanced subject such as Analytic Combinatorics be taught effectively online?

When Means Bound Variances: Concentration for Recursively Determined Random Values

Mike Steele
Wharton, University of Peensylvania, USA

Abstract

In many algorithmic contexts there is a "value" that is determined by a recursion of some kind. It is often the case that the asymptotic behavior of the mean of the value can be found without too much difficulty, but, even in problems where the mean is understood with some precision, one often has little further information about the distribution of the realized value. The main goal of this talk is to describe a rich class of problems where the variance is bounded by small multiple of the mean. This class includes many problems from the theory of combinatorial optimization where the means are determined by a Bellman equation (or an analogous recursion). En route, one finds a bounded difference martingale, so useful concentration inequalities come at no extra cost.

Average Redundancy of the Shannon Code for Markov Sources

Wojciech Szpankowski
Purdue, USA

Abstract

It is known that for memoryless sources, the average and maximal redundancy of fixed–to–variable length codes, such as the Shannon and Huffman codes, exhibit two modes of behavior for long blocks. It either converges to a limit or it has an oscillatory pattern, depending on the irrationality or rationality, respectively, of certain parameters that depend on the source. In this paper, we extend these findings, concerning the Shannon code, to the case of a Markov source, which is considerably more involved. While this dichotomy, of convergent vs. oscillatory behavior, is well known in other contexts (including renewal theory, ergodic theory, local limit theorems and large deviations of discrete distributions), in information theory (e.g., in redundancy analysis) it was recognized relatively recently. To the best of our knowledge, no results of this type were reported thus far for Markov sources. We provide a precise characterization of the convergent vs. oscillatory behavior of the Shannon code redundancy for a class of irreducible, periodic and aperiodic, Markov sources. These findings are obtained by analytic methods, such as Fourier/Fejér series analysis and spectral analysis of matrices.

This is joint work with Neri Merhav (Technion, Israel).

Typical depth of a Digital Search Tree built on a general source

Brigitte Vallée
GREYC, CNRS and University of Caen, France

Abstract
The digital search tree (dst) plays a central role in compression algorithms, of Lempel-Ziv type. This structure can be viewed as a mixing of a digital structure (the trie) with a binary search tree. Its probabilistic analysis is thus involved, even in the case when the text is produced by a simple source (a memoryless source, or a Markov chain). After the seminal paper of Flajolet and Sedgewick (1986) which deals with the memoryless unbiased case, papers of Jacquet, Louchard, Prodinger Szpankowski, Tang (between 1990 and 2005) dealt with general memoryless sources or Markov chains, and performed the analysis of the main parameters of Digital Search Trees (DST) –namely, internal path length, profile, typical depth– . Here, we are interested in a more realistic analysis, when the words are emitted by a general source. There exist previous analyses of text algorithms or digital structures that have been performed for general sources, for instance for tries (Clément, Flajolet, Vallée), and (Cesaratto, Vallée), or for basic sorting and searching algorithms (Clément-Fill-Flajolet-Vallée) and (Clément-NguyenThi-Vallée). However, the case of digital search trees has not yet been considered, and this is the main subject of the talk. The first steps of the work were performed with Philippe Flajolet, during November/December 2010.

Compositions, canonical trees, acyclic digraphs and their common structural properties

Stephan Wagner
Stellenbosch University, South Africa

Abstract
The aim of this talk is to exhibit the structural similarities between several seemingly unrelated combinatorial objects, and to show how they reflect on the analytic level. Among these objects are: canonical t-ary trees, partitions and compositions of 1 into unit fractions, compositions of integers, and acyclic digraphs.

"Canonical" t-ary trees are rooted t-ary trees, i.e., each vertex has 1 or t children, with the additional property that the vertex degrees on each level decrease from left to right. They are in bijection with partitions of 1 into unit fractions whose denominators are powers of t as well as equivalence classes of compact prefix-free t-ary codes. Their limit as $t \to \infty$ are compositions of integers, and they can actually themselves be modeled as compositions with certain restrictions. Some of their structural parameters (e.g., the height) have limit distributions that are quite different from other tree families such as simply generated families of trees, but that have analogues in the world of compositions. As it turns out, acyclic digraphs, i.e., digraphs without directed cycles, form another family of combinatorial objects with similar traits that also occurs frequently in computer science.

The similarities between all these combinatorial classes will be demonstrated by looking at various structural parameters, such as height, width, path length, profile, etc. The talk is based on joint work with Clemens Heuberger and Daniel Krenn.

Resolution of Thomas Ward's Question and Steven Finch's Conjecture. Precise Asymptotic Analysis of an Integer Sequence Motivated by the Dynamical Mertens' Theorem for Quasihyperbolic Toral Automorphisms

Mark D. Ward
Purdue University, USA

Abstract

We analyze the first-order asymptotic growth of the sequence

$$a_n = \int_0^1 \prod_{j=1}^{n} 4\sin^2(\pi j x)\, dx.$$

This sequence is a weighted average of the number of orbits in a particular quasihyperbolic automorphism of a $2n$-torus, which has applications to ergodic and analytic number theory. The combinatorial structure of the sequence is also of interest, as the "signed" number of ways in which 0 can be represented as $\sum_{\substack{-n\le j\le n\\ j\neq 0}} \epsilon_j j$, with $\epsilon_j \in \{0,1\}$. Our result answers a question of Thomas Ward (no relation to the author) and confirms a conjecture of Steven Finch.

Analysis of Quickselect under Yaroslavskiy's Dual-Pivoting Algorithm

Sebastian Wild
Univ. Kaiserslautern, Germany

Abstract

There is excitement within the algorithms community about a new partitioning algorithm introduced by Yaroslavskiy. This algorithm renders Quicksort slightly faster than its speed under classic partition methods. We show that this improved performance in Quicksort is not sustained in Quickselect (a variant of Quicksort for finding order statistics).

This is joint work with Hosam Mahmoud and Markus Nebel.

Diagonal asymptotics for products of combinatorial classes

Mark C. Wilson
University of Auckland, New Zealand

Abstract

We generalize and improve recent results by Bóna and Knopfmacher and by Banderier and Hitczenko concerning the joint distribution of the sum and number of parts in tuples of restricted compositions. Specifically, we generalize the problem to general combinatorial classes and relax the requirement that the sums of the composition be equal. We extend the main explicit results to enumeration problems whose counting sequences are Riordan arrays. We give an alternative method for computing asymptotics in the supercritical schema case of Flajolet and Sedgewick, avoiding explicit diagonal extraction. We claim that this method is more computationally efficient.