Mathematisches Forschungsinstitut Oberwolfach

Report No. 50/2014

DOI: 10.4171/OWR/2014/50

Probability, Trees and Algorithms

Organised by
Luc Devroye, Montreal
Ralph Neininger, Frankfurt

2 November – 8 November 2014

Abstract. The subject of this workshop were probabilistic aspects of algorithms for fundamental problems such as sorting, searching, selecting of and within data, random permutations, algorithms based on combinatorial trees or search trees, continuous limits of random trees and random graphs as well as random geometric graphs. The deeper understanding of the complexity of such algorithms and of shape characteristics of large discrete structures require probabilistic models and an asymptotic analysis of random discrete structures. The talks of this workshop focused on probabilistic, combinatorial and analytic techniques to study asymptotic properties of large random combinatorial structures.

Mathematics Subject Classification (2010): 60F05, 60C05, 68W40, 68P10.

Introduction by the Organisers

The asymptotic analysis of random discrete structures that are related to the complexity of fundamental algorithms is an active field both in Mathematics and in Computer Science. In this area various probabilistic and analytic techniques have been developed in the last decades including methods based on martingales, connections to branching random walks, the contraction method, techniques using generating functions, and the method of moments. In Computer Science random trees appear in the performance analysis of data structures, in the context of coding schemes as well as connected to fundamental algorithms such as sorting, searching and selecting. Moreover, in the last years also fascinating connections of these random tree models to coalescent processes, fragmentation theory and other combinatorial stochastic processes have been found. Also their random continuous limits such as random real trees have led to a much refined understanding of these structures and the related complexities. The aims of this workshop were advances

and a deeper understanding in the probabilistic analysis of random discrete structures and their limiting objects and to discuss their connections beyond the realm of models motivated from Computer Science.

The talks covered the following themes:

- Random binary search trees, random recursive trees, other search trees and digital trees (Addario-Berry, Bubeck, Bertoin, Gnedin, Hwang, Janson, Leckey, Neininger, Rösler, Kabluchko)
- Combinatorial tree structures (Duquesne, Gittenberger, Mailler, Sulzbach)
- Random graphs, geometric graphs and their limits (Bhamidi, Grübel, Mitsche, Müller)
- Geometric probabilities (Marckert)
- The smoothing transform (Alsmeyer)
- Formal power series and probability (Fill)

We thank the director, the administration, and the staff of the MFO for their hospitality and great support before and throughout the meeting.

Acknowledgement: The MFO and the workshop organizers would like to thank the National Science Foundation for supporting the participation of junior researchers in the workshop by the grant DMS-1049268, "US Junior Oberwolfach Fellows". Moreover, the MFO and the workshop organizers would like to thank the Simons Foundation for supporting Hsien-Kuei Hwang in the "Simons Visiting Professors" program at the MFO.

Workshop: Probability, Trees and Algorithms

Table of Contents

Abstracts

Asymptotics of Cut-trees

Jean Bertoin

Consider a finite tree $T = (V, E)$, where as usual V denotes the set of vertices and E the set of edges, and imagine that we destroy T by removing its edges one after the other and in a uniform random order (other types of destruction, random or deterministic, can also be considered). The Cut-tree $\mathrm{Cut}(T)$ is a binary rooted tree which encodes main aspects of the destruction of T. The vertex set of $\mathrm{Cut}(T)$ is formed by the set of blocks (connected components) that arise during the destruction. The root of $\mathrm{Cut}(T)$ consists of the initial block V, and each vertex $v \in V$ can be identified as a leaf $\{v\}$ of $\mathrm{Cut}(T)$. The basic structure is that each time a block B is split into two sub-blocks B' and B'' (because an edge in B is removed), then B, B' and B'' are nodes of $\mathrm{Cut}(T)$ such that B' and B'' as the two children of B.

Meir and Moon initiated the study of the number of steps required to isolate a distinguished vertex, if at each time an edge is removed, the block which does not contain the distinguished vertex is discarded forever (in other words, one only takes into account the edge-removals occurring in the block of the distinguished vertex). The number of edge-removals required to isolate a given vertex v in the destruction of T corresponds precisely to the height of the leaf $\{v\}$ in $\mathrm{Cut}(T)$. More generally, the number of cuts required to isolate k vertices $v_1, \ldots, v_k$ coincides with the total length of the cut-tree reduced to its root and the k leaves $\{v_1\}, \ldots, \{v_k\}$, where the length is measured as usual by the graph distance on $\mathrm{Cut}(T)$.

The talk was intended as a survey of some recent developments in the study of the limits of Cut-trees for certain families of large random trees, including Random Recursive Trees, Galton-Watson Trees, and Birthday Trees, and their applications.

References

[1] Bertoin, J. Fires on trees. *Ann. Instit. Henri Poincaré Probab. Stat.* **48** (2012), 909-921.

[2] Bertoin, J. The cut-tree of large recursive trees. To appear in *Ann. Instit. Henri Poincaré Probab. Stat.*

[3] Bertoin, J. and Miermont, G. The cut-tree of large Galton-Watson trees and the Brownian CRT. *Ann. Appl. Probab.* **23** (2013), 1469-1493.

[4] Broutin, N. and Wang, M. Cutting down p-trees and inhomogeneous continuum random trees. Preprint (2014). ArXiv:1408.0144

[5] Broutin, N. and Wang, M. Reversing the cut tree of the Brownian continuum random tree. Preprint (2014). ArXiv:1408.2924

[6] Dieuleveut, D. The vertex-cut-tree of Galton-Watson trees converging to a stable tree. *Ann. Appl. Probab.*(to appear). ArXiv:1312.5525.

[7] Haas, B. and Miermont, G. Scaling limits of Markov branching trees with applications to Galton-Watson and random unordered trees. *Ann. Probab.* **40-6** (2012), 2589-2666.

[8] Meir, A. and Moon, J. W. Cutting down recursive trees. *Mathematical Biosciences* **21** (1974), 173-181.

Maximal clades in random binary search trees

Svante Janson

A *phylogenetic tree*, or a *full binary tree* is a tree where every node has outdegree 0 or 2; nodes with outdegree 0 are called *external* and nodes with outdegree 2 *internal*. By eliminating all external nodes, we get a *binary tree*, and this yields a bijection between phylogenetic trees with $n+1$ external nodes and binary trees with n nodes.

The *clade* of an external node v in a phylogenetic tree is the set of external nodes that are descendants of the parent of v. (This is called a *minimal clade* by [1] and [2].) Note that two clades are either nested or disjoint, and that the set of maximal clades forms a partition of the set of external nodes. We let $F(T)$ denote the number of maximal clades of a phylogenetic tree T. The maximal clades, and the number of them, were introduced by [4], together with a biological motivation, and further studied by [3].

Translated to the corresponding binary tree (i.e., the internal nodes), a clade is thus a node having outdegree at most 1, and a maximal clade is a clade such that all ancestors have outdegree 2.

We consider a random binary search tree $\mathcal{T}_n$ (which corresponds to the Yule–Harding model of a random phylogenetic tree) and the number of maximal clades $X_n := F(\mathcal{T}_n)$ in it. We consider asymptotics as $n \to \infty$.

It was proved by [5] and [3] that

$$\mathbb{E}\, X_n = \mathbb{E}\, F(\mathcal{T}_n) = \alpha n + O(1), \tag{1}$$

where that the mean number of maximal clades $\mathbb{E}\, X_n \sim \alpha n$, where

$$\alpha = \frac{1 - e^{-2}}{4}. \tag{2}$$

Moreover, [3] found also corresponding results for the variance and higher central moments:

$$\mathbb{E}(X_n - \mathbb{E}\, X_n)^2 \sim 4\alpha^2 n \log n, \tag{3}$$

and for any fixed integer $k \geq 3$,

$$\mathbb{E}(X_n - \mathbb{E}\, X_n)^k \sim (-1)^k \frac{2k}{k-2} \alpha^k n^{k-1}. \tag{4}$$

As a consequence of (3)–(4), the limit distribution of $F(\mathcal{T}_n)$ (after centering and normalization) cannot be found by the method of moments. Nevertheless, [3] further proved asymptotic normality, where, unusually, the normalizing uses (the square root of) *half* the variance:

$$\frac{X_n - \mathbb{E}\, X_n}{\sqrt{2\alpha^2 n \log n}} \xrightarrow{\mathrm{d}} N(0,1). \tag{5}$$

We use probabilistic methods to reprove these theorems, together with some further results. In particular, we can explain the appearance of half the variance in (5) as follows:

Fix a sequence of numbers $N = N(n)$, and say that a clade is *small* if it has at most $N+1$ elements, and *large* otherwise. Let X_n^N be the number of maximal small clades, i.e., the small clades that are not contained in any other small clade. It turns out that a suitable choice of N is about $\sqrt{n}$; we have for example the following.

Theorem 1. *Let* $N := \sqrt{n}$. *Then* $\operatorname{Var}(X_n^N) \sim 2\alpha^2 n \log n$ *and*

$$\frac{X_n^N - \mathbb{E}\, X_n^N}{\sqrt{\operatorname{Var} X_n^N}} \xrightarrow{\mathrm{d}} N(0,1). \tag{6}$$

Furthermore, $X_n - X_n^N = o_{\mathrm{p}}\bigl(\sqrt{\operatorname{Var} X_n^N}\bigr)$ *and* $\mathbb{E}\, X_n - \mathbb{E}\, X_n^N = o\bigl(\sqrt{\operatorname{Var} X_n^N}\bigr)$, *so we may replace* X_n^N *by* X_n *in the numerator of* (6). *However,*

$$\operatorname{Var}(X_n - X_n^N) \sim \operatorname{Var}(X_n^N) \sim 2\alpha^2 n \log n. \tag{7}$$

The theorem thus shows that the large clades are rare, and do not contribute to the asymptotic distribution; however, when they appear, the larges clades give a large (actually negative) contribution to X_n, and as a result, half the variance of X_n comes from the large clades. (When there is a large clade, there is less room for other clades, so X_n tends to be smaller than usually.)

For higher moments, the large clades play a similar, but even more extreme, role. Note that (for $n \ge 2$) with probability $2/n$, the root of $\mathcal{T}_n$ has outdegree 1, and then it is the unique maximal clade, and thus $X_n = 1$. Since $\mathbb{E}\, X_n = \alpha n + O(1)$ by (1), we thus have $X_n - \mathbb{E}\, X_n = -\alpha n + O(1)$ with probability $2/n$, and this single exceptional event gives a contribution $\sim (-1)^k 2\alpha^k n^{k-1}$ to $\mathbb{E}(X_n - \mathbb{E}\, X_n)^k$, which explains a fraction $(k-2)/k$ of the moment (4); in particular, this explains why the moment is of order n^{k-1}.

For proofs and further details, see [6].

References

[1] Michael G. B. Blum and Olivier François, Minimal clade size and external branch length under the neutral coalescent. *Adv. in Appl. Probab.* **37** (2005), no. 3, 647–662.

[2] Huilan Chang and Michael Fuchs, Limit theorems for patterns in phylogenetic trees. *J. Math. Biol.* **60** (2010), no. 4, 481–512.

[3] Michael Drmota, Michael Fuchs and Yi-Wen Lee, Limit laws for the number of groups formed by social animals under the extra clustering model. (Extended abstract.) *Proceedings, 2014 Conference on Analysis of Algorithms, AofA '14 (Paris, 2014)*, *DMTCS Proceedings*, 2014.

[4] Eric Durand, Michael G. B. Blum and Olivier François, Prediction of group patterns in social mammals based on a coalescent model. *J. Theoret. Biol.* **249** (2007), no. 2, 262–270.

[5] Eric Durand and Olivier François, Probabilistic analysis of a genealogical model of animal group patterns. *J. Math. Biol.* **60** (2010), no. 3, 451–468.

[6] Svante Janson, Maximal clades in random binary search trees. Preprint, 2014. arXiv:1408.6337

On the Number of Unary-Binary Structures with Bounded Unary Height

Bernhard Gittenberger
(joint work with Olivier Bodini, Danièle Gardy, Zbigniew Golębiewski)

We consider rooted tree-like structures which are composed of unary and binary nodes. The unary height of a vertex is the number of unary nodes on the path connecting it to the root. First we study the asymptotic number of Motzkin trees (=unary-binary trees) with constraints on the number of unary nodes or with bounded unary height. Then we perform the analogous investigation on a similar structure, namely λ-terms with bounded number of nestings for the quantifyer λ. The latter can be represented by a Motzkin tree (as their skeleton) where further edges are added according to some restrictions.

The analysis is done by a generating function approach and singularity analysis. The generating function for the number of Motzkin trees with unary height less than k can be written as nested square-root with polynomial expressions under the square-roots and k levels of nesting. Obviously, the function becomes singular if one of the square-roots vanishes. We show that there is only one singularity at the circle of convergence. At the singular point the outermost radical vanishes whereas the other radicals do not. Thus, the dominant singularity ρ_k of the generating function $T_k(z)$ is of squareroot type, i.e. the Puiseux expansion of $T_k(\rho_k)$ is a power series in $\sqrt{1-\frac{z}{\rho_k}}$, and using a transfer lemma of the Flajolet-Odlyzko type [4] we can obtain the asymptotic number of Motzkin trees with n nodes and bounded unary height:

$$[z^n]T_k(z) \sim \frac{\sqrt{1+4\rho_k^2}}{4\rho_k^{n+1}\, n\sqrt{\pi n}}.$$

Next we look at balanced Motzkin trees, i.e. Motzkin trees with the property that all leaves have unary height k. Again the generating function $B_k(z)$ happens to be a nested squareroot of polynomial expressions. This time there are exactly two singularities on the circle of convergence, $z=-1/2$ and $z=1/2$. At the negative one, the innermost radical vanishes whereas the other do not. This singularity is therefore again of squareroot type. At $z=1/2$ all $k+1$ radicals vanish simultaneously. Hence we have a singularity of type $\left(1-\frac{z}{\rho_k}\right)^{1/2^{k+1}}$ which dominates the squareroot singularity at $z=-1/2$. We obtain

$$[z^n]B_k(z) \sim 2^{n+2^{-k-1}}\, \frac{n^{-1-2^{-k-1}}}{\Gamma(-2^{-k-1})}.$$

Whereas most recursively defined combinatorial structures exhibit a squareroot type singularity, other types are possible and have been characterized in [2]. This result offers an example for "unusual" singularity types.

Since λ-terms are built from skeleton which is a Motzkin tree, we similarly obtain generating function of nested squareroot shape. For the class of λ-terms with unary height bounded by k we get $k+1$ levels of nesting. But this time we

observe rather strange phenomena: For $k = 0$ we are faced with Catalan trees. There is one level of nesting and this squareroot determines the singularity. For $k = 1$ we have two levels of nesting and at the singularity both of them vanish simultaneously, thus yielding a fourth root type singularity. For $k = 2, 3, \ldots, 7$ only the second radicand (counted from inside out) vanishes at the singularity. So we have a squareroot singularity in these cases. At $k = 8$ it is the second and the third radical determine the singularity, and then for $k = 9, 10, \ldots$ it is the third radical only. At $k = 135$ there are again 2 radicals involved, the 3rd and the 4th one. The sequence of special values where a phase transition occurs can be identified as the doubly exponential sequence $N_i = u_i^2 - u_i + i$ with $u_0 = 0$ and $u_{i+1} = u_i^2 + i + 1$.

We can show that, if $N_j < k < N_{j+1}$, $k \neq N_j$, then the singularity of $H_k(z)$ originates from the j-th radicand and is of square root type. Otherwise, the singularity of $H_k(z)$ originates from both the j-th and $(j+1)$-th radicand and is of 4th root type. Its value is then $\frac{-1+\sqrt{1+4(N_j-j)}}{4(N_j-j)} = \frac{1}{2u_j}$

The first values of the sequence N_j are $1, 8, 135, 21760, 479982377, \ldots$

By means of the theory of doubly exponential sequences developped by Aho and Sloane [1] we can represent it in the form

$$N_j = u_{j+1} - u_j - 1, \qquad u_j = \lfloor \chi^{2^j}, \quad \text{and } \chi = \lim_{j\to\infty} u_j^{1/2^j} \approx 1.36660956$$

By means of further analysis and transfer lemmas we obtain

(i) If $N_j < k < N_{j+1}$, then there exists a constant h_k such that, with $\tilde{\sigma}_{j,k}$ the unique real positive zero of $R_{j,k}$

$$[z^n]H_k \sim h_k n^{-3/2}(\sigma_{j,k})^{-n}.$$

(ii) If $k = N_j$,

$$[z^n]H_{N_j} \sim h_{N_j} n^{-5/4}(2u_j)^n,$$

where

$$h_{N_j} = \frac{\gamma_j^{\frac{1}{4}}}{4\sqrt{2}\,\Gamma(3/4)\,\rho_j^{\frac{1}{4}}\,2^{N_j-j}\,\sqrt{\prod_{i=1}^{N_j-j}\lambda_i}},$$

The reason for this rather strange and unexpected behaviour seems to be the following observation: In a typical large λ-term with bounded unary height k, the leaves are concentrated in the last $\log\log k$ "unary depth levels" of the tree. It seems that in those levels, the number of leaves is normally distributed with mean and variance $\mu \sim cn, \qquad \sigma^2 \sim dn$. In the lower levels the number of leaves is $O(1)$.

For $k = N_j$, the next level w.r.t. unary depth enters. The number of leaves in the newly coming level has expectation $\Theta(\sqrt{n})$ A detailed analysis of this phenomenon is ongoing research.

Concerning the constants h_k in the asymptotic main terms, we observe the following numerical values:

$$h_1 = 0.24261...;$$
$$h_8 = 9.31888... \cdot 10^{-5};$$
$$h_{135} = 8.56995... \cdot 10^{-157}.$$

It can be shown that h_k is indeed a sequence of superexponential decay. Note that therefore, for large k, the asymptotic main term is barely invisible in simulations, since the actual asymptotic behaviour shows up no sooner than for very large values of n.

Those phenomena are also one reason why Boltzmann samplers (see [3]) face serious difficulties in generating random λ-terms of large size.

References

[1] Alfred V. Aho and Neil J. A. Sloane. Some Doubly Exponential Sequences. Fibonacci Quarterly, Vol. 11 (1970), pp. 429-437.

[2] Cyril Banderier and Michael Drmota. Coefficients of algebraic functions: formulae and asymptotics. Proceedings 25th International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC 2013), DMTCS Proc. AS, pp. 1065-1076, 2013.

[3] Philippe Duchon, Philippe Flajolet, Guy Louchard, and Gilles Schaeffer, Boltzmann samplers for the random generation of combinatorial structures, Combinatorics, Probablity, and Computing **13** (2004), no. 4–5, 577–625, Special issue on Analysis of Algorithms.

[4] Philippe Flajolet and Andrew Odlyzko. Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics*, 3(2):216-240, 1990.

High degrees in recursive trees

Louigi Addario-Berry

(joint work with Laura Eslava)

1. Statement of results

Let T_n be a random recursive tree (RRT) on n vertices. Let $\deg(i)$ denote the in-degree of vertex $i \in [n]$ in T_n[1]. Devroye and Lu [1] showed that the maximum degree Δ_n of T_n satisfies $\Delta_n / \log_2 n \to 1$ a.s. The focus in this work is the distribution of the vertices with degree near the maximum. For any $d \in \mathbb{Z}$, let

$$X_n(d) = |\{i \in [n] : \deg(i) = \lfloor \log_2 n \rfloor + d\}|.$$

We show that there exists a Poisson point process $\mathcal{P}$, and suitable projections $\mathcal{P}_\varepsilon$ of $\mathcal{P}$ into $\mathbb{Z}$ approximate $\{X_n(d)\}_{d\in\mathbb{Z}}$ when n is large. The limiting distribution of the number of degrees of constant order, $Y_n = |\{i \in [n] : \deg(i) = d\}|$, $d \geq 0$; was first study in [7]. Subsequent studies can be found in [8] and [3].

Notation. Given a Poisson Point Process $\mathcal{N}$ on $\mathbb{R}$ and $A \subset \mathbb{R}$, we write $N(A) = |N \cap A|$. Let $\varepsilon_n = \log_2 n - \lfloor \log_2 n \rfloor$ for any $n \in N$. Let $\mathcal{P}$ be a Poisson point

[1]Write $[n] = \{1, 2, \ldots, n\}$.

process on $\mathbb{R}$ with intensity $f(x) = \ln 2 \cdot 2^{-x}$. For any $\varepsilon \in [0,1]$, let $\mathcal{P}_\varepsilon$ be a Point process in $\mathbb{Z}$ obtained by setting $P_\varepsilon(d) = P[d-\varepsilon, d+1-\varepsilon)$ for all $d \in \mathbb{Z}$.

Theorem 1. *For $n \in \mathbb{N}$ let μ_n be the law of $\{X_n(d)\}_{d\in\mathbb{Z}}$ and λ_ε be the law of $\mathcal{P}_\varepsilon$. Then*

$$dist(\mu_n, \lambda_{\varepsilon_n}) \to 0 \tag{1}$$

in the sense of finite dimensional distributions.

Goh and Schmutz [2] first studied the limiting distribution of $\Delta_n - \log_2 n$, they used singularity analysis of generating functions. Theorem 1 in [2] is, in fact, a corollary of Theorem 1 above. The following strengthens that result.

Theorem 2. *Uniformly on integers $-\infty \le d \le 2\ln n$,*

$$\mathbf{P}(\Delta_n \ge \lfloor \log_2 n \rfloor + d) = (1 - \exp\{-2^{-d+\varepsilon_n}\})(1+o(1)). \tag{2}$$

Finally, we also obtain asymptotic normality for $P_n(d)$ when d tends to $-\infty$ slowly enough. Write $\sigma_n = \sqrt{2^{d+\varepsilon_n}}$.

Theorem 3. *Let $d = d(n) \to -\infty$ and $d = o(\ln n)$, then for any sequence n_k with $\varepsilon_{n_k} \to \varepsilon \in [0,1)$,*

$$\frac{P_{n_k}(d) - 2^{d+\varepsilon_{n_k}}}{\sigma_{n_k}} \to N(0,1).$$

The proofs are basically obtained using the method of moments for any finite collection $X_n(d_1), \ldots, X_n(d_k)$ (see [6], section 6.1). In particular, we need to understand

$$\mathbf{P}(\deg(i_j) = m_j,\ j \in [k]) \tag{3}$$

for all $i_1, \ldots, i_k \in [n]$ and $m_j - \lfloor \log_2 n \rfloor \in \mathbb{Z}$.

The standard model for RRT defines T_n so that every vertex $i \in [n]$ attaches to a uniformly chosen vertex $k \in [i-1]$. In particular, $\deg(i)$ is a sum of Bernoulli variables; $\deg(i) = \sum_{j=i+1}^{n} \mathbf{I}_{\{j\to i\}}$. For any fixed $i = i(n)$, the exact distribution of $\deg(i)$ is derived in [4], and its limiting distribution in [5]. However, results on the joint distribution of $\{\deg(i)\}_{i\in[n]}$, in particular in the regime $i = \theta(\ln n)$, have proved challenging; in part due to the lack of symmetry of the model. To make the analysis of (3) easier, we use Kingman's coalescent to generate trees whose degree sequence is exchangeable and has the same joint distribution as $\{deg(i)\}_{i\in[n]}$.

2. The perspective of Kingman's coalescent

Kingman's n-coalescent consists of a sequence of forests $F_1, \ldots, F_n$ defined as follows. Independently for each $1 \le i \le n-1$ let (a_i, b_i) be a uniformly random element of $\{(a,b);\ 1 \le a < b \le n+1-i\}$ and let ξ_i be Bernoulli(1/2)-distributed and independent of $\{a_i, b_i\}$. The forest F_i consists of $n+1-i$ rooted trees $T_1^{(i)}, \ldots, T_{n+1-i}^{(i)}$, we order the trees according to the smallest-labelled vertex, so in particular $1 \in T_1^{(i)}$ for all $i \in [n]$.

Let F_1 be the forests consisting of n isolated vertices $1, 2, \ldots, n$. Given F_i, we construct F_{i+1} by adding an edge e_i between the roots of $T_{a_i}^{(i)}$ and $T_{b_i}^{(i)}$. If $\xi_i = 1$,

direct e_i towards $T^{(i)}_{a_i}$; otherwise e_i is directed towards $T^{(i)}_{b_i}$. If $\xi_i = 1$ we say that that ξ_i *favours* the vertices of $T^{(i)}_{a_i}$, and otherwise that it favours the vertices of $T^{(i)}_{b_i}$. The forest F_{i+1} consists of the new tree and the remaining $n-1-i$ unaltered trees from F_i.

Let $L = (L_e, e \in E(T^{(n)}_1))$ denote the edge labels in $T^{(n)}_1$. A counting argument shows that $(T^{(n)}_1, L)$ is uniform over trees with vertices labelled $1, \dots, n$ and with a decreasing edge labelling using $1, \dots, n-1$.

Fact. *Viewed as rooted labelled trees, $T^{(n)}_1$ and T_n have the same law.*

As a consequence, in studying degrees we may use $T^{(n)}_1$ in place of T_n; one advantage of this is that degrees in $T^{(n)}_1$ are exchangeable.

3. Degrees in $T^{(n)}_1$

Abusing notation, let $\deg(i)$ be the in-degree of vertex i in $T^{(n)}_1$. We will sketch the proof that for any $k \in \mathbb{N}$, and $d = m - \lfloor \log_2 n \rfloor \in \mathbb{Z}$,

$$\mathbf{P}(\deg(i) = m,\ i \in [k]) = \mathbf{P}(\deg(1) = m)^k (1 + o(1)) = 2^{-k(m+1)}(1 + o(1)). \tag{4}$$

This will imply that the law of $X_n(d)$ is approximately $\mathrm{Po}(2^{-(d-\varepsilon_n)})$ as $n \to \infty$.

Let $S = \{s \in [n-1],\ 1 \in \{a_s, b_s\}\}$ be the times at which the tree containing vertex 1 is chosen to be merged. It follows that, at step $s \in S$, vertex 1 increases its degree only if for all $i \in S \cap [s]$, ξ_i favours 1. To the contrary, if at time $s \in S$, vertex 1 is the root of $T^{(s)}_1$ and ξ_s does not favours 1, then vertex 1 is no longer a root in $T^{(j)}_1, j > s$; and its degree is fixed for the remainder of the process. Therefore,

$$\begin{aligned}\mathbf{P}(\deg(1) = m) &= 2^{-m}\mathbf{P}(|S| = m) + 2^{-(m+1)}\mathbf{P}(|S| > m) \\ &= 2^{-(m+1)}(2\mathbf{P}(|S| = m) + \mathbf{P}(|S| > m)).\end{aligned}$$

It turns out that $\mathbf{P}(|S| > m) = 1 + o(1)$ for $m \le 2\ln n$, implying (4) for $k = 1$. The proof of this follows from the observation that if $S = \{s_1, s_2, \dots\}$ (with $s_1 < s_2 < \dots$), then $n - s_j$ can be approximated by $nB_1 \cdots B_j$, where B_i are independent and distributed as Beta(1,2). This approach is similar to that in [9] in which the study of the height of a binary search tree relies on approximating the size of a subtree of a given node at depth j as $nU_1 \cdots U_j$ where U_i are independent uniform variables in $(0, 1)$.

For $k \ge 2$, (4) says that the degrees of k vertices behave essentially independently. The main obstacle for this to occur is the following. Suppose that at a given time s, the degrees of vertices 1 are 2 are less than m and that the trees selected to be merged are precisely those containing vertex 1 and 2 as their roots. Then ξ_s can only favour one of the two vertices and consequently at least one of them will not achieve the desired degree m in $T^{(n)}_1$. The key to (4) is to prove that such bad scenarios will not occur with high probability.

Define $\tau_{i,j} = \min\{s \in [n-1],\ a_s = i, b_s = j\}$. And let $\tau_k = \min\{\tau_{i,j} \in [n-1],\ i,j \le k\}$. Then any possible collisions between vertices in $[k]$ must occurs after step τ_k. It is not difficult to see that $\tau_k = n - O_p(1)$.

Now, let $D_{i,s}$ be the event that vertex i has degree exactly m in F_s. Then (4) is obtained by showing that, for suitable s, with $n-s \to \infty$, $\mathbf{P}(\cap_{i\in[k]} D_{i,s}) = 1+o(1)$. That means that with high probability, vertices in $1,\dots,k$ attain the desired degree even before step τ_k.

References

[1] Devroye, Luc and Lu, Jiang *The strong convergence of maximal degrees in uniform random recursive trees and dags.* In Random Structures Algorithms **7** (1995), pages 1–14.

[2] Goh, William and Schmutz, Eric *Limit distribution for the maximum degree of a random recursive tree.* In J. Comput. Appl. Math. **142** (2002), pages 61–82.

[3] Janson, Svante *Asymptotic degree distribution in random recursive trees.* In Random Structures and Algorithms **26** (2005), pages 69–83.

[4] Javanian, Mehri and Vahidi-Asl, Mohammad Q. *Note on the outdegree of a node in random recursive tree.* In Journal of Applied Mathematics and Computing **13** (2003), pages 99–103.

[5] Kuba, Markus and Panholzer, Alois *On the degree distribution of the nodes in increasing trees.* In Journal of Combinatorial Theory, Series A **114**(2007), pages 597–618 .

[6] Janson, Luczak and Rucinski "Random Graphs". Wiley Interscience series in Discrete Mathematics and Optimization (2000) 335pp.

[7] Na, Hwa Sung and Rapoport Anatol *Distribution of nodes of a tree by degree.* In Mathematical Biosciences **6** (1970), pages 313–329.

[8] Mahmoud, Hosam M. and Smythe, R.T. *Asymptotic joint normality of outdegrees of nodes in random recursive trees.* In Random Structures and Algorithms **6** (1992), pages 255–266.

[9] Reed, Bruce *The height of a random binary search tree.* In *J. of the ACM* **50** (2003), pages 306-332.

Hyperbolic Random Geometric Graphs

Tobias Müller

(joint work with Michel Bode, Erik Broman, Nikolaos Fountoulakis, Johan Tykesson)

Continuum percolation goes back to the work of E.N. Gilbert [2] in 1961. In the most simple set version of the model, we take a constant intensity Poisson process on the plane, and obtain an infinite graph by joining pairs of points by an edge if their distance is less than one.

The term *random geometric graph* usually refers to finite version of continuum percolation. Such a finite version can for instance be obtained by discarding all points outside of some large disk D. Equivalently, because of the scaling properties of the Poisson process, we can also consider a Poisson process of larger intensity on the unit disk and join two of its points by an edge if their distance is less than some (small) number $r > 0$. Sometimes the variant is considered in which a fixed number n of points is sampled from the unit disk and two points are joined if their distance is less than r. Results are usually easily transferred between the two settings.

Studying finite random geometric graphs allows us to ask some questions, such as "what is the probability that the graph is connected" or "what is the probability that it has a Hamilton cycle", that do not seem to make much sense, or have trivial answers, in the infinite case. Hafner [3] was the first to study such (finite) random geometric graphs. They have since become the focus of considerable research effort and very precise answers are now known to many of the natural questions for this model. Penrose's monograph [5] contains a good overview of the results prior to 2003.

A natural question is what happens if we define this model not on the ordinary euclidean plane but instead on the hyperbolic plane. For the infinite version of the model, continuum percolation, this was already considered by Tykesson [6] who showed that a strikingly different behaviour occurs. In euclidean continuum percolation, there is a "critical intensity" $\lambda_c > 0$ such that if the intensity λ of the underlying Poisson process is at most λ_c then all components of the graphs are finite, while if $\lambda > \lambda_c$ then there is a unique infinite component. As shown by Tykesson, in hyperbolic continuum percolation there are two important special values of the intensity: λ_c and λ_u with $0 < \lambda_c < \lambda_u < \infty$. If $\lambda < \lambda_c$ then all components are finite, if $\lambda > \lambda_u$ then there is a unique infinite component, but if $\lambda_c < \lambda < \lambda_u$ then there are infinitely many infinite components.

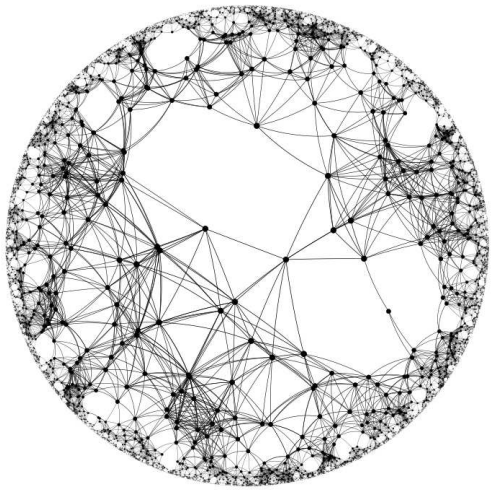

FIGURE 1. Computer simulation of continuum percolation on the hyperbolic plane, shown in the Poincaré disk model, with intensity parameter $\lambda = 5$.

In the talk I presented some recent work, and work in preparation, on (finite) random geometric graphs in the hyperbolic plane. In the hyperbolic plane we do not have the easy scaling relations at our disposal that exist in the euclidean

setting and there are different, inequivalent ways in which one might define a finite random geometric graphs on the hyperbolic plane.

In an ongoing joint work with Erik Broman and Johan Tykesson we consider one possible and natural definition: We take a constant intensity Poisson process on the hyperbolic plane, join two points by an edge if the distance is less than one, and we discard all points outside of a disk of radius R. We then consider the number of points in the largest component, for large R.

Another way to define a random geometric graph in the hyperbolic plane was recently introduced by Krioukov et al. [4]. Here we randomly pick N points at random from a disk of radius R on the hyperbolic plane, and we join them if their distance is less than R (the radius of the disk the points live on). It turns out that if the number of points N is proportional to the square root of the (hyperbolic) area of the disk then one gets a model with constant average degree, a power-law degree sequence and a positive clustering coefficent. In two joint works with Michel Bode and Nikolaos Fountoulakis we studied the largest component and the connectivity threshold in this model. The transition from a connected to a disconnected graph exhibits behaviour that is very different from all other random graph models in the literature as far as we are aware.

References

[1] M. Bode, N. Fountoulakis, T. Müller, *The probability that the hyperbolic random graph is connected*, preprint, available from *http://www.math.uu.nl/ Muell001/Papers/BFM.pdf*

[2] E.N. Gilbert, *Random plane networks*, J. Soc. Indust. Appl. Math. **9** 1961 533–543.

[3] R. Hafner, *The asymptotic distribution of random clumps*, Computing (Arch. Elektron. Rechnen), **10**:335–351, 1972.

[4] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat and M. Boguná, *Hyperbolic Geometry of Complex Networks*, Phys. Rev. E **82** (2010), 036106

[5] M.D. Penrose, *Random Geometric Graphs*, Oxford University Press, 2003

[6] J. Tykesson, *The number of unbounded components in the Poisson Boolean model of continuum percolation in hyperbolic space*, Electron. J. Probab. **12** (2007), no. 51, 1379–1401

Metric structure and universality of scaling limits: Maximal components of critical random graphs

Shankar Bhamidi

(joint work with Nicolas Broutin, Sanchayan Sen, Xuan Wang)

Over the last few years a wide array of random graph models have been postulated, including the configuration model, inhomogeneous random graph models and dynamic random graph processes such as bounded size rules. Most of these models come with a parameter t (usually related to the edge density) and a (model dependent) critical time t_c which specifies when a giant component emerges. There is evidence to support that under moment conditions, the nature of this emergence is universal and look like the classical Erdős-Rényi random graph, in the sense of the critical scaling window and (a) the sizes of the components in this window

(all maximal component sizes scaling like $n^{2/3}$) and (b) the structure of components (rescaled by $n^{1/3}$) converge to random fractals related to the continuum random tree. Till date, (a) has been proven for a few models using different techniques while (b) has been proven for only two models, the classical Erdős-Rényi random graph and the rank-1 inhomogeneous random graph. The aim of this paper is to develop a general program for proving such results. The program requires three main ingredients: (i) in the critical scaling window, components merge approximately like the multiplicative coalescent (ii) the scaling exponents of the susceptibility functions are the same as the Erdős-Rényi random graph and (iii) control on expected distances between random points for "large components" in the barely subcritical regime. We show that these apply for a wide class of random graph models including the configuration model and random r-regular graphs, inhomogeneous random graphs modulated via a finite kernel κ and random graphs created via bounded size rules. Thus these models all belong to the domain of attraction of the classical Erdős-Rényi random graph. As a by product we also get results for component sizes for a general class of inhomogeneous random graphs in the barely subcritical and critical regime.

In order to prove the main universality result, the key heavy lifting is done via an earlier work joint with Sanchayan Sen and Xuan Wang [1] where we studied the special case of the rank one random graph. In this paper we established a connection between this model and tilted versions of a famous class of random tree models called **p**-trees; this class of random trees have earlier arisen in describing the entrance boundary of the additive coalescent. In this paper assuming high enough moments convergence of the maximal components viewed as metric spaces was established in the stronger l^4 metric as formulated by Addario-Berry, Broutin and Goldschmidt.

References

[1] Shankar Bhamidi, Sanchayan Sen, Xuan Wang, *Continuum limit of critical inhomogeneous random graphs*, Arxiv preprint (2014).

On rigidity, orientability and cores of random graphs with sliders

Dieter Mitsche

(joint work with Julien Barré, Marc Lelarge)

Consider a set of points, some of them allowed to move freely in the Euclidean plane, and some constrained to move on fixed lines, called sliders. The free points have two degrees of freedom, the points attached to sliders have only one. Now, add bars between pairs of these points; a bar fixes the length between the two endpoints. The points and bars form a *framework*. A framework is said to be rigid if it cannot be deformed (but can possibly be translated and rotated on the plane); equivalently, it is rigid if the distance between any pair of points, connected by a bar or not, is fixed. Characterizing the rigidity of a framework is very difficult in general. In the absence of sliders, a celebrated theorem by Laman [2] ensures

that for a generic framework, its rigidity properties only depend on its underlying graph, where points are vertices and bars are edges: the geometry does not enter. We will call vertices of *type 1* (resp. *type 2*) if they are (resp. are not) connected to a slider. Consider now a percolation setting: take a set of n vertices, a fraction q of which are type 2, and add edges randomly. The questions are: When does a giant (that is: including a finite fraction of the vertices) rigid structure emerge? What is its size? When edges are added uniformly at random between pairs of vertices, the resulting graph is an Erdös-Rényi random graph $\mathcal{G}(n, c/n)$. In this case and for $q = 1$ (no slider), Kasiviswanathan et al. [1] showed that the threshold for a giant rigid component is $c \simeq 3.588$, and that the transition is discontinuous: above the threshold, the giant rigid component always includes a finite fraction of all n vertices. We generalize this setup by allowing for sliders, and mention now all results proved in the full paper.

1. Statements of Results

Let G be a graph (V, E) with $|V| = n$ and $|E| = m$. Vertices are either of type 1 or of type 2, and for $i \in \{1, 2\}$, n_i denotes the number of vertices of type i, so that $n = n_1 + n_2$. Subgraphs are typically denoted by G' with $n_i(G')$ vertices of type $i \in \{1, 2\}$, $n(G') = n_1(G') + n_2(G')$ vertices in total and $m(G')$ edges. When the context is clear, we use the notations: $n' = n(G')$, $n'_i = n_i(G')$ and $m' = m(G')$.

Let G be a graph with $n = n_1 + n_2$ vertices and m edges. G is **sparse** if for all subgraph $G' \subseteq G$ on $n' = n'_1 + n'_2 \geq 2$ vertices and m' edges, we have:

$$m' \leq 2n'_2 + n'_1 + \min(0, n'_1 - 3) = 2n' - \max(n'_1, 3).$$

A graph is **rigid** if it contains a spanning subgraph which is minimally rigid. A **rigid block** in G is defined to be a vertex-induced rigid subgraph. A **rigid component** of G is an inclusion-wise maximal block. We show the following:

Theorem 1. *The set of all minimally rigid graphs with n_1 vertices of type 1, n_2 vertices of type 2 forms the set of bases of a matroid, whose ground set is the set of edges of the complete graph on $n = n_1 + n_2$ vertices.*

Our goal from now on is to investigate rigidity, orientability and cores as a function of $q \in [0, 1]$. Since $q = 0$ is closely related to the emergence of the giant connected component, we are thus interested in situations interpolating between standard connectivity percolation and rigidity percolation as studied in [1].

Consider in the following a random graph $G \in \mathcal{G}(n, c/n)$ with $c > 0$. Denote a graph chosen in this way by $G(n, c/n)$. Let $Q(x, y) = e^{-x} \sum_{j \geq y} \frac{x^j}{j!}$. Throughout this note, for fixed $q \in [0, 1]$, each vertex independently from all others gets type 1 with probability $1 - q$ and type 2 with probability q. We define the function $c^*(q)$ as follows: for $q \leq 1/2$, we set $c^*(q) = \frac{1}{1-q}$; for $q > 1/2$, let $\xi^* = \xi^*(q)$ be the unique positive solution to $\xi \frac{(1-q)Q(\xi,1)+qQ(\xi,2)}{(1-q)Q(\xi,2)+2qQ(\xi,3)} = 2$, and set in this case then $c^*(q) = \frac{\xi^*}{(1-q)Q(\xi^*,1)+qQ(\xi^*,2)}$.

Theorem 2. *For $G = G(n, c/n)$ as above, and let $q \in [0,1]$. Let $R_n(q,c)$ (resp. $R_n^C(q,c)$) be the number of vertices covered by the largest rigid component (resp. largest connected rigid block) of G.*

- *For $c > c^*(q)$, there is a giant rigid component in G, i.e., there exists $\alpha = \alpha(q,c) > 0$ such that $\mathbb{P}\left(\frac{R_n(q,c)}{n} \geq \alpha\right) \to 1$ as $n \to \infty$.*
- *For $c < c^*(q)$, there is no giant rigid component in G, i.e., $\forall \alpha > 0$, $\mathbb{P}\left(\frac{R_n(q,c)}{n} \geq \alpha\right) \to 0$ as $n \to \infty$.*

The above results also hold true for $R_n^C(q,c)$. Moreover, for $c > c^*(q)$, a.a.s., there is one unique giant rigid component (resp. one unique giant connected rigid block).

We also show that the transition as c varies and q is held fixed is continuous for $q \leq 1/2$ and discontinuous if $q > 1/2$. More precisely, we have the following:

Theorem 3. *Let $G = G(n, c/n)$ as above.*

- *The transition is discontinuous for $q > 1/2$: let $q > 1/2$; there is $\alpha(q) = \alpha > 0$ such that for any $c > c^*(q)$, $\lim_{n\to\infty} \mathbb{P}\left(\frac{R_n(q,c)}{n} \geq \alpha\right) = 1$.*
- *The transition is continuous for $q \leq 1/2$: let $q \leq 1/2$; for any $\alpha > 0$, $c > \frac{1}{1-q}$ but $c \to \frac{1}{1-q}$, $\lim_{n\to\infty} \mathbb{P}\left(\frac{R_n(q,c)}{n} \geq \alpha\right) = 0$.*

We also relate rigidity with orientability and cores. A graph is **1.5-orientable** if there exists an orientation of the edges such that type 1 vertices have in-degree at most 1 and type 2 vertices have in-degree at most 2.

Theorem 4. *Let $G = G(n, c/n)$ be as above and let $q \in [0,1]$. We have:*

(a) *if $c < c^*(q)$, G is 1.5-orientable a.a.s.*
(b) *if $c > c^*(q)$, G is not 1.5-orientable a.a.s.*

For a graph with type 1 and type 2 vertices, the **2.5-core** is the largest induced subgraph with all type 1 vertices with degree at least 2 and all type 2 vertices with degree at least 3. Note that this definition coincides with the 2-core (resp. 3-core) if the graph contains only type 1 vertices (resp. type 2). In our probabilistic setting, it turns out that for a fixed q, the 1.5-core appears at a value $\tilde{c}(q) \leq c^*(q)$.

Let $Q(x,y)$ as before. Define $\tilde{c}(q) = \inf_{\xi>0} \frac{\xi}{(1-q)Q(\xi,1)+qQ(\xi,2)}$, and let $\tilde{\xi}(q,c)$ be the largest solution to $\xi = c(1-q)Q(\xi,1) + cqQ(\xi,2)$.

Theorem 5. *Let $G = G(n, c/n)$ be as above. Let Core be the 2.5-core of G, $n_1(Core)$ (resp. $n_2(Core)$) be the number of nodes of type 1 (resp. 2) in the core and $m(Core)$ be the number of edges in the core. Let $q \in [0,1]$. We have*

(a) *if $c < \tilde{c}(q)$ and $q > 0$, then a.a.s. the 2.5-core has $o_p(n)$ vertices.*
(b) *if $c > \tilde{c}(q)$, then a.a.s. the 2.5-core is not empty and $n_1(Core)/n \to (1-q)Q(\tilde{\xi}(q,c)),2)$, $n_2(Core)/n \to qQ(\tilde{\xi}(q,c)),3)$, and $2m(Core)/n \to \tilde{\xi}(q,c)\left((1-q)Q(\tilde{\xi}(q,c)),1) + qQ(\tilde{\xi}(q,c)),2)\right)$.*

Starting from the 2.5-core, one constructs a larger subgraph as follows: add recursively type 1 vertices which are linked by one edge with the current subgraph,

and type 2 vertices which are linked by two edges with the current subgraph. The resulting subgraph is called the **2.5+1.5-core**. Note that this definition coincides with the 2+1-core (resp. 3+2-core) if the graph contains only type 1 vertices (resp. type 2).

Furthermore, we also compute the threshold and the size of the 2.5+1.5-core. This proves a conjecture in [1] on the 3+2-core.

Theorem 6. *Let $G = G(n, c/n)$ as above. Let $Core+$ be the $2.5 + 1.5$-core of $G(n, c/n)$, and $n(Core+)$ the number of vertices inside the $2.5 + 1.5$-core. Let $q \in [0, 1]$ and let $c > \tilde{c}(q)$. Then a.a.s., $n(Core+)/n = (1 + o(1)) \left(1 - e^{-\tilde{\xi}} + q\tilde{\xi}e^{-\tilde{\xi}}\right)$.*

References

[1] S. P. Kasiviswanathan, C. Moore and L. Théran. *The rigidity transition in random graphs*, *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms (SODA) 2011*, (2011), 1237–1252.

[2] G. Laman. *On graphs and rigidity of plane skeletal structures*, *J. Engrg. Math.* **4**, (1970), 331–340.

About Sylvester's question.

Jean-François Marckert

Place n points $z_1, ..., z_n$ at random independently and under the uniform distribution in a compact convex set K of the plane (with non empty interior). Let P_n^K denote the probability that the z_i's are in a convex position, that is the vertices of a convex polygon.

Bläschke proved in 1918 that, for any K, for $n = 4$,

$$P_n^T \leq P_n^K \leq P_n^\circ$$

where T designates the triangle and $\circ$ the disk. It is worth mentioning that affine transformations conserve both convex sets and uniform distributions so that the particular position, shape, area of the triangle T or those of the disk do not matter.

Valtr [5] (1995) showed that if $\square$ is a square (or a non flat parallelogram) then, for $n \geq 1$,

$$P_n^\square = \left(\frac{\binom{2n-2}{n-1}}{n!} \right)^2,$$

and in a second paper, [6] (1996) he proved that if T is a (non flat) triangle then, for $n \geq 1$,

$$P_n^T = \frac{2^n (3n-3)!}{(n-1)!^3 (2n)!}.$$

Buchta [3] goes further and gives an expression for $P_{n,m}^\square$ and $P_{n,m}^T$ as a finite sum of explicit terms where $P_{n,m}^K$ designates the probability that the convex hull has

size m. Bárány (1999) [1] has shown that

$$\lim_{n\to+\infty} n^2 (P_K^n)^{1/n} = e^2 A^3(K)/4 \tag{1}$$

where $A^3(K)$ is the supremum of the *affine perimeter* of all convex sets $S \subset K$. For the disk one gets

$$\log(P_D^n) = -2n\log n + n\log\left(2\pi^2 e^2\right) + o(n).$$

The first part of the talk was devoted to the computation of $P_{n,m}^\circ$. This new result can be found in [4]. The idea, consists in establishing a sort of decomposition of the computation of $P_{n,m}^\circ$. For this instead of $P_{n,m}^\circ$, we compute the probability $F_n(\theta)$ that n points are taken in the segment $S(\theta)$ with angle-θ (the intersection of a disk and an half-plane) taken together with the two extreme points of the special border of the segment, are in a convex position.

Then $P_n^\circ = \lim_{\theta\to 2\pi^-} F_n(\theta)$. The rest of the the computation consists in establishing a functional equation linking $F_n(\theta)$ with the $(F_k(\nu), k < n, \nu \in (0, 2\pi))$. This is done by a decomposition of $F_n(\theta)$ relying on the smallest segment $S(\nu)$ with special border the same as $S(\theta)$ containing all the points in $S(\nu)$. Now for any $n \geq 0$, $\theta \in (0, 2\pi)$ define

$$L_{n,m}(\theta) \quad = \quad \frac{B_{n,m}(\theta)(\theta - \sin(\theta))^n \sin(\theta/2)}{n!} \tag{2}$$

Hence

$$L_0(\theta) = \sin(\theta/2),\ L_1(\theta) = \sin(\theta/2)(\theta - \sin(\theta)). \tag{3}$$

Theorem.

(*i*) For any $n \geq 1$,

$$P_n^\circ = \lim_{\substack{t\to 2\pi \\ t<2\pi}} B_{n-1}(t).$$

(*i'*) For any $n \geq 2$,

$$P_n^\circ \quad = \quad \frac{(n-2)!}{2^{n-2}\pi^{n-1}} \int_0^{2\pi} \sum_{k=0}^{n-2} L_k(\phi) L_{n-2-k}(2\pi - \phi) d\phi$$

(*ii*) For any $\theta \in (0, 2\pi)$ and any $n \geq 1$,

$$\frac{L_n(\theta)}{2} = \int_0^\theta \frac{\sin(\theta/2)^{2n+1}}{\sin(\phi/2)^{2n+1}} \int_0^\phi \sum_{k=0}^{n-1} L_k(\eta) L_{n-1-k}(\phi - \eta) d\eta d\phi. \tag{4}$$

Analogous results can be obtained for $P_{n,m}^\circ$:

(*iii*) For any $\theta \in (0, 2\pi)$ any k, and any $l \geq k+1$, $L_{k,l}(\theta) = 0$. For any $\theta \in (0, 2\pi)$, any $n \geq 1$ and any $1 \leq m \leq n$

$$\begin{aligned} \frac{L_{n,m}(\theta)}{2} \quad &= \quad \int_0^\theta \int_0^\phi \frac{\sin(\theta/2)^{2n+1}}{\sin(\phi/2)^{2n+1}} \sum_{\substack{n_1+n_2+n_3=n-1 \\ m_1+m_2=m-1}} \frac{(\sin(\eta) + \sin(\phi-\eta) - \sin(\phi))^{n_3}}{n_3!} \\ &\times \quad L_{n_1,m_1}(\eta) L_{n_2,m_2}(\phi - \eta) d\eta d\phi \end{aligned}$$

An alternative form can be given using

$$\sin(\eta) + \sin(\phi - \eta) - \sin(\phi) \quad = \quad 4\sin(\frac{\phi - \eta}{2})\sin(\phi/2)\sin(\eta/2).$$

(iii') For any $n \geq 2$ and any $1 \leq m \leq n$

$$P^{\circ}_{n,m} \quad = \quad \frac{(n-2)!}{2^{n-2}\pi^{n-1}} \int_0^{2\pi} \sum_{\substack{n_1+n_2=n-1 \\ m_1+m_2=m-1}} L_{n_1,m_1}(\phi) L_{n_2,m_2}(2\pi - \phi) d\phi.$$

(iv) For any $n \geq 1$,

$$P^{\circ}_{n,m} = \lim_{\substack{t \to 2\pi \\ t < 2\pi}} B_{n-1,m-1}(t).$$

A second part of the talk has been devoted to the proof of the fact that for any K, for $n = 5$,

$$P^{K}_{n} \leq P^{\circ}_{n}.$$

Since the result has not been submitted for the moment, I prefer not discuss here the main idea of this part.

References

[1] I. Bárány. Sylvester's question: The probability that n points are in convex position. *Ann. Probab.*, 27(4):2020–2034, 1999.

[2] W. Blaschke. Über affine geometrie xi: Lösung des 'vierpunktproblems' von sylvester aus der theorie der geometrischen wahrscheinlichkeiten. *Ber. Verh. Sachs. Akad. Wiss. Leipzig Math.-Phys*, 69:436–453, 1917.

[3] C. Buchta. On the number of vertices of the convex hull of random points in a square and a triangle. *Anz. Österreich. Akad. Wiss. Math.-Natur. Kl.*, 143:3–10, 2009/10.

[4] JF. Marckert, *Probability that n random points in a disk are in convex position*, arXiv:1402.3512.

[5] P. Valtr. Probability that n random points are in convex position. *Discrete & Computational Geometry*, 13:637–643, 1995.

[6] P. Valtr. The probability that n random points in a triangle are in convex position. *Combinatorica*, 16(4):567–573, 1996.

The Leading Root of a Formal Power Series

$$f(x,y) = \sum_{n=0}^{\infty} a_n(y)\, x^n,$$

with Connections to Probability

James Allen Fill

(joint work with Alan D. Sokal)

This abstract concerns the "leading root" (unique power-series root) $x_0(y)$ of a formal power series $f(x,y) = \sum_{n=0}^{\infty} a_n(y)\, x^n$, where the series $a_n(y)$ have nonzero constant terms for $n = 0, 1$ and for $n \geq 2$ satisfy modest smallness conditions such as $a_n(y) = O(y^{\alpha(n-1)})$ for some $\alpha > 0$. Problems of this type arise frequently in combinatorics, statistical mechanics, number theory, and analysis.

A prominent example is the "deformed exponential function" (DE)

$$f(x,y) = \sum_{n=0}^{\infty} \frac{y^{n(n-1)/2}}{n!} x^n.$$

For this example, let $U(y) := -x_0(y)$. Extensive numerical computations lead to the conjecture that $U(y)$ has all strictly positive coefficients; and, even more strongly, that $F(y) := 1 - [1/U(y)]$ has all strictly positive coefficients after the vanishing constant term. This is just the proverbial tip of the iceberg, in that similar positivity properties, both for the leading root and for approximations to the leading root used for efficient computation of its coefficients, are conjectured for wide families of examples that include DE.

There are almost no proofs available yet for these conjectures, but there are exceptions. There are also several connections of these problems with probability. Here are two obvious connections:

(i) **Connection with Poisson distribution:** If $N \sim \mathrm{Poi}(u)$, then

$$f(-u,y) = e^u \, \mathbf{E}[(-1)^N y^{N(N-1)/2}].$$

(ii) **Connection with log-normal distribution:** Let $W \sim \mathrm{N}(\mu, \sigma^2 = v)$, where $\mu \in \mathbf{R}$ and $0 < v < \infty$, and consider the log-normal random variable $L := e^W$. Then L has the following moments:

$$\mathbf{E}\, L^n = \exp\left(n\mu + \frac{1}{2} v n^2\right).$$

The moment generating function ϕ of L is finite only at the origin, but *formally*

$$\phi(t) = \mathbf{E}\, e^{tL} = \sum_{n=0}^{\infty} \mathbf{E}\, L^n \, \frac{t^n}{n!} = f(t \exp(\mu + \tfrac{1}{2}v), \exp(\tfrac{1}{2}v)).$$

We cite just a few of the many applications of the DE function:

- Statistical mechanics: partition function of single-site lattice gas [5]
- Combinatorics: generating function for Tutte polynomials on K_n [6]; acyclic digraphs [1]; inversions of trees [4]; hashing with linear probing [2]; ...
- Functional-differential equation: $F'(x) = F(yx)$, where $' = \partial/\partial x$ [3]
- Complex analysis: Whittaker and Goncharov constants [8]

To touch on the connection with Tutte polynomials, we note (see [7] and [6]) for example that with

$$C_n(v) := \sum_m c_{n,m} v^m,$$

where $c_{n,m}$ is the number of connected graphs with n (labeled) vertices and m edges, we have

$$\sum_{n=1}^{\infty} \frac{x^n}{n!} C_n(v) = \log f(x, 1+v).$$

Generalizing (DE), consider the formal power series

$$f(x) = 1 + x + \sum_{n=2}^{\infty} a_n x^n$$

where $\mathbf{a} = (a_n)_{n=2}^{\infty}$ are indeterminates. Following a suggestion of David Wagner, we call this the *generic formal power series* with leading term $1 + x$. We now think that this is the right way to approach the leading-root problem, i.e., treating $(a_n)_{n=2}^{\infty}$ as indeterminates; one can *later* specialize $a_n \to a_n(y)$ as one wishes.

This extended abstract focuses on one of several methods for computing (to any specified number of terms) the leading root $x_0(y)$ of the generic formal power series with leading term $1 + x$, namely, by use of Waring polynomials. We define the *Waring polynomials* $P_n(\mathbf{a})$ *of the first kind* in (the first n of) the indeterminates $\mathbf{a} = (a_i)_{i \geq 1}$ as follows:

$$\begin{aligned} P_n(\mathbf{a}) &= (-1)^{n-1} n\, [t^n] \log\Bigl(1 + \sum_{i=1}^{\infty} a_i t^i\Bigr) \\ &= n \sum_{\mathbf{m}:\|\mathbf{m}\|=n} (-1)^{\|\mathbf{m}\|-|\mathbf{m}|} \frac{(|\mathbf{m}|-1)!}{m_1! \cdots m_n!} \prod_{i=1}^{n} a_i^{m_i} \\ &= n \sum_{\mathbf{m}:\|\mathbf{m}\|=n} \frac{(|\mathbf{m}|-1)!}{m_1! \cdots m_n!} \prod_{i=1}^{n} [(-1)^{i-1} a_i]^{m_i}. \end{aligned}$$

Here $\mathbf{m} = (m_1, m_2, \ldots, m_n)$ and $|\mathbf{m}| := \sum_i m_i$ and $\|\mathbf{m}\| := \sum_i i m_i$.

This formula, or rather its immediate implication expressing the nth power-sum symmetric function in a collection of variables in terms of the elementary symmetric functions, was found (and proved by induction) by the English mathematician Edward Waring in the 1700s. His second formula goes the other way around, and there are similar formulas involving the complete homogeneous symmetric functions h_n. The Waring formula is familiar to probabilists because it can be used to express cumulants in terms of moments.

A general connection between the leading root $x_0(y)$ of the generic formal power series and the Waring polynomials can be established. When specialized to (DE), for example, this connection establishes that

$$[y^k]\, F(y) \;=\; [y^k]\, F_n(y) \qquad \text{whenever } k < n, \tag{1}$$

where

$$F_n(y) := 1 - P_n\left(1, \frac{y^{2(2-1)/2}}{2!}, \ldots, \frac{y^{n(n-1)/2}}{n!}\right)^{1/n}. \tag{2}$$

So we can compute F by computing the F_n's (or P_n's). But how can we compute the P_n's efficiently? The answer is this: by establishing a polymer interpretation and then a simple recurrence relation that follows from it; due to space limitations, we omit further details. Not only does it appear to be true that $F(y)$ has all strictly positive coefficients after the vanishing constant term; it even appears that the

same is true for each of the series $F_n(y)$! For $n = 2$ and $n = 3$, one can prove that F_n is a probability generating function (pgf). We know a (non-probabilistic) technique to prove that F_4 and F_5 (and perhaps F_n for other fairly small values of n) have nonnegative coefficients. But what about general F_n? Is F_n the pgf of some random variable one can describe (easily)?

Returning to the generic setting, specialize $U(\mathbf{a}) = -x_0(\mathbf{a})$ to the *alternating-signs case* by fixing an integer $p \geq 2$ and setting $a_2 = \ldots a_{p-1} = 0$ and $a_i = (-1)^i c_i$ for $i \geq p$. Then for $\beta \geq -(2p-1)$, it is not hard to prove that the power series $\big[\xi_0(\mathbf{a})^\beta - 1\big]\big/\,\beta$ in the variables $\mathbf{c}$ has nonnegative coefficients. Even in this simple alternating-signs case, analogous results for approximations devised using Waring polynomials [as at (1)–(2) for (DE)] seem elusive. Recently we have made some initial progress by obtaining such a result in the very special univariate case where c_i is set to zero for all $i \geq p + 1$. The proof uses, among other things, rather nontrivial interlacing-roots arguments.

References

[1] E.A. Bender, L.B. Richmond, R.W. Robinson and N.C. Wormald, The asymptotic number of acyclic digraphs. I., Combinatorica **6**, 15–22 (1986).

[2] P. Flajolet, P. Poblete and A. Viola, On the analysis of linear probing hashing, Algorithmica **22**, 490–515 (1998).

[3] J.K. Langley, A certain functional-differential equation, J. Math. Anal. Appl. **244**, 564–567 (2000).

[4] C.L. Mallows and J. Riordan, The inversion enumerator for labeled trees, Bull. Amer. Math. Soc. **74**, 92–94 (1968).

[5] A.D. Scott and A.D. Sokal, The repulsive lattice gas, the independent-set polynomial, and the Lovász local lemma, J. Stat. Phys. **118**, 1151–1261 (2005), cond-mat/0309352 at arXiv.org.

[6] A.D. Scott and A.D. Sokal, Some variants of the exponential formula, with application to the multivariate Tutte polynomial (alias Potts model), Séminaire Lotharingien de Combinatoire **61A**, article 61Ae (2009), arXiv:0803.1477 [math.CO] at arXiv.org.

[7] W.T. Tutte, On dichromatic polynomials, J. Combin. Theory **2**, 301–320 (1967).

[8] J. Waldvogel, Zero-free disks in families of analytic functions, in *Approximation Theory*, Lecture Notes in Mathematics #1287 (Springer-Verlag, Berlin, 1987), pp. 209–228.

Decomposition of Lévy trees along their diameter.

Thomas Duquesne

(joint work with Minmin Wang)

We consider the diameter of Lévy trees that are random compact metric spaces obtained as the scaling limits of Galton-Watson trees. Lévy trees have been introduced by Le Gall and Le Jan [3] and they generalise Aldous' Continuum Random Tree [2] that corresponds to the Brownian case. We first characterize the law of the diameter of Lévy trees and we prove that it is realized by a unique pair of points. We prove that the law of Lévy trees conditioned to have a fixed diameter $r \in (0, \infty)$ is obtained by glueing at their respective roots two independent size-biased Lévy trees conditioned to have height $r/2$ and then by uniformly re-rooting the resulting

tree; we also describe by a Poisson point measure the law of the subtrees that are grafted on the diameter. This decomposition relies on a similar one for Lévy trees along the geodesic realizing their height that has been obtained by Abraham & Delmas [1] As an application of this decomposition of Lévy trees according to their diameter, we characterize the joint law of the height and the diameter of stable Lévy trees conditioned by their total mass; we also provide asymptotic expansions of the law of the height and of the diameter of such normalized stable trees, which generalizes the identity due to Szekeres [4] in the Brownian case.

References

[1] R. Abraham, J-F. Delmas *Williams' decomposition of the Lévy continuum random tree and simultaneous extinction probability for population with neutral mutations*, Stochastic Process. Appl. **119** (2009), 1124–1143.

[2] D. Aldous, *The continuum random tree. I*, Annals of Probability **19** (1991), 1-28.

[3] J-F. Le Gall, Y. Le Jan, *Branching processes in Lévy processes*, Annals of Probability **26** (1998), 213–252.

[4] G. Szekeres, *Distribution of labelled trees by diameter*, In *Combinatorial Mathematics, X (Adelaide, 1982* Lecture Notes in Math. Springer Berlin **1036** (1983), 392–397.

On random recursive triangulations of the disk

Henning Sulzbach

(joint work with Nicolas Broutin)

The study of random triangulations of polygons and the disk in the plane was initiated by Aldous [4, 5] in the context of his seminal works [1, 2, 3] on the continuum random tree arising as scaling limits for various classes of trees of height of order $\sqrt{n}$. In this talk, I presented results on related objects, namely random recursive laminations and triangulations of the disk and their dual trees as introduced by Curien and Le Gall [8]. Here, on the unit circle (considered as the unit interval identifying $s = 0$ and $s = 1$), in each step, one chooses two points uniformly at random and connects them by a chord provided the latter does not intersect any chord already present. For a simulation of the process, see Figure 1. Curien and Le Gall mainly studied the lamination L_n as a subset of the unit disk. Exploiting the theory of fragmentation processes, compare, e.g. Bertoin's book [6], among other results, they prove the following theorem.

Theorem 1 (Curien, Le Gall)**.** *There exists a random continuous function Z on $[0,1]$, such that $\overline{\bigcup_{n\geq 1} L_n}$ is a triangulation of the disk encoded by the function Z. In other words, points $0 \leq s \leq t \leq 1$ are connected by a chord if and only if $Z(t) = Z(s) = \inf\{Z(u) : u \in [s,t]\}$. For a simulation of Z, see Figure 2.*

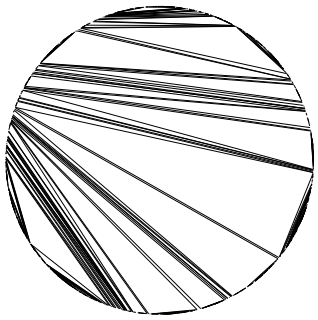

FIGURE 1. L_{2000}

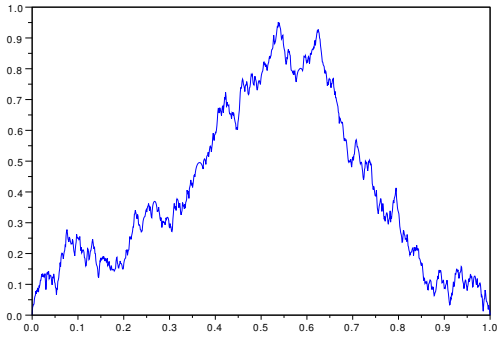

FIGURE 2. The limit Z

1. Main results

In the talk, I discussed the following strengthening of the theorem above concerning the tree T_n which is dual to the lamination L_n rooted at the node assigned to the fragment covering $s = 0$. For the theory of real trees encoded by continuous functions and the Gromov-Hausdorff topology on the space of compact metric spaces, we refer to [9] and [11].

Theorem 2 (Broutin, Sulzbach [7])**.** *For the contour process $D_n(s), s \in [0,1]$, of the tree T_n, where $D_n(s)$ denotes the depth of the node covering s, we have, almost surely and uniformly in $s \in [0,1]$*

$$n^{-\beta/2} D_n(s) \to Z(s), \quad \beta = \frac{\sqrt{17}-3}{2}.$$

Almost surely, with respect to the Gromov-Hausdorff distance we have, $n^{-\beta/2} T_n \to \mathcal{T}_Z$ where $\mathcal{T}_Z$ the real tree encoded by Z.

The proof of the theorem relies on a recursive decomposition of the quantity $D_n(s)$. The result follows from an application of the contraction method worked out in the space of regular functions, compare [12] for an elaborate approach to this matter. The process Z can be characterized by a stochastic fixed-point equation on the space of continuous functions reflecting the distributional recursivity.

2. A homogeneous model

In a related model, in each step, a fragment in the lamination of the disk is chosen uniformly at random and split by a chord again inserted uniformly at random. In contrast to the first model, this lamination process is not biased towards producing small fragments, see Figure 3 for an illustration. Denoting L_n^h the lamination process, T_n^h its dual tree and $D_n^h(s)$ its contour process, we have the following result.

Theorem 3 (Broutin, Sulzbach [7]). *Almost surely,*

$$n^{-1/3} D_n^h(s) \to H(s), \quad n^{-1/3} T_n^h \to \mathcal{T}_H$$

where the first convergence is uniform in $s \in [0,1]$ *and the second with respect to the Gromov-Hausdorff distance. The process* H *is continuous, see Figure 4 for a simulation.*

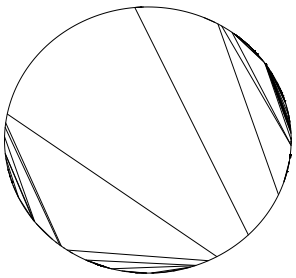

FIGURE 3. L_{2000}^h

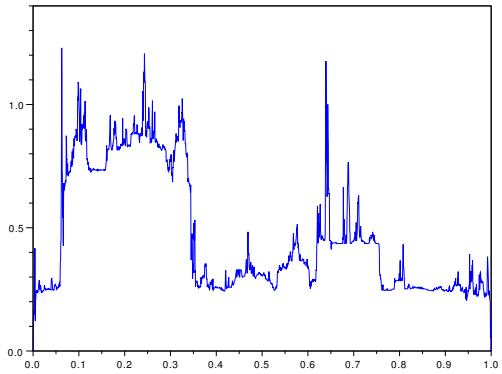

FIGURE 4. The limit H

3. Fractal dimensions

The fractal dimension offers further insight to the structure of fractal objects often arising in the context of an underlying recursive construction. We focus on Hausdorff and Minkowski dimension. For a comprehensive account to this matter, we refer to Falconer's book [10]. For the continuum Random tree $\mathcal{T}_e$, it is well-known that, both for Hausdorff and Minkowski dimension, we have, almost surely,

$$\dim(\mathcal{T}_e) = 2.$$

This result is a consequence of the fact that standard Brownian Motion (and Brownian excursion) has Hölder continuous paths with exponent $1/2 - \varepsilon$ for any $\varepsilon > 0$. We can prove a similar result here: almost surely,

$$\dim(\mathcal{T}_Z) = \beta^{-1} = 1.78\ldots, \quad \dim(\mathcal{T}_H) = 3.$$

Curien and Le Gall [8] proved that the Hausdorff dimension of limiting triangulation $\overline{\bigcup_{n\geq 1} L_n}$ in both models equals $1 + \beta$.

We conclude that in comparison to the limiting triangulations, the dual trees $\mathcal{T}_Z$ and $\mathcal{T}_H$ are arguably the more interesting objects incorporating the geometry of their discrete counterparts.

References

[1] D. Aldous, *The continuum random tree. I*, Ann. Probab. **19** (1991), 1–28.
[2] D. Aldous, *The continuum random tree. II: an overview*, Cambridge, In Stochastic analysis, Cambridge University Press, Cambridge **19** (1991), 23–70.
[3] D. Aldous, *The continuum random tree. III*, Ann. Probab. **21** (1993), 248–289.
[4] D. Aldous, *Triangulating the circle, at random*, Amer. Math. Monthly **101** (1994), 223–233.
[5] D. Aldous, *Recursive self-similarity for random trees, random triangulations and Brownian excursion*, Ann. Probab. **22** (1994), 527–545.
[6] J. Bertoin, *Random fragmentation and coagulation processes*, Cambridge University Press, Cambridge (2006)
[7] N. Broutin, H. Sulzbach, *The dual tree of a recursive triangulation of the disk*, Ann. Probab. **43** (2015), 738–781.
[8] N. Curien, J.-F. Le Gall, *Random recursive triangulations of the disk via fragmentation theory*, Ann. Probab. **39** (2011), 2224–2270.
[9] S. N. Evans, *Probability and real trees*, Lecture Notes in Mathematics, **1920** Lectures from the 35th Summer School on Probability Theory held in Saint-Flour, July 6–23, 2005, Springer, Berlin (2008), xii+193.
[10] K. J. Falconer, *Fractal geometry: Mathematical Foundations and Applications*, John Wiley & Sons Ltd., Chichester (1990).
[11] J.-F. Le Gall, *Random trees and applications*, Probab. Surv. **2** (2005), 245–311
[12] R. Neininger, H. Sulzbach, *On a functional contraction method*, to appear in Ann. Probab. (2014)

The smoothing transform in random environment

Gerold Alsmeyer

Given an infinite vector $(C, T_1, T_2, ...)$ of random variables taking values in $\mathbb{R} \times \mathbb{R}_{\geq} \times \mathbb{R}_{\geq} \times ...$, the classical smoothing transform is a map $\mathcal{S}$ that maps a distribution Q on $\mathbb{R}$ to

$$\mathcal{S}Q \;=\; \mathcal{L}\left(\sum_{k\geq 1} T_k X_k + C\right),$$

where $X_1, X_2, ...$ are iid with common law Q and independent of $(C, T_1, ...)$ and $\mathcal{L}(X)$ means the law of a random variable X. Q is a fixed point of $\mathcal{S}$ if $\mathcal{S}Q = Q$, thus

$$X \;\stackrel{d}{=}\; \sum_{k\geq 1} T_k X_k + C$$

if X is a copy of the X_n and $\stackrel{d}{=}$ denotes equality in distribution. The homogeneous case occurs if $C = 0$. There are many examples of random recursions related to branching processes, random trees and algorithms which feature the fixed point of a smoothing transform as a limiting distribution. Here are two well-known examples:

- If W denotes the a.s. limit of a simple supercritical Galton-Watson process $(Z_n)_{n\geq 0}$ with $Z_0 = 1$ and offspring mean m, thus $W = \lim_{n\to\infty} m^{-n} Z_n$, then $W \stackrel{d}{=} \sum_{k=1}^{Z_1} m^{-1} W_k = \sum_{k\geq 1} m^{-1} \mathbf{1}_{\{Z_1 \geq k\}} W_k$, where $W, W_1, ...$ are iid and independent of Z_1.

- If X denotes the distributional limit of the normalized number of key comparisons the algorithm `Quicksort` needs to sort a list of n items, then it satisfies the fixed-point equation $X \stackrel{d}{=} UX_1 + (1-U)X_2 + g(U)$, where U has a uniform law on $(0,1)$, $g(U) = 1 + 2U \log U + 2(1-U)\log(1-U)$ and X, X_1, X_2 are iid and independent of U, see [5]

Many articles have dealt with the problem of describing all fixed points of $\mathcal{S}$ under varying assumptions on $(C, T_1, ...)$, see [4, 1, 2, 3] and the references therein.

The purpose of this talk is to give a brief introduction of how to define the smoothing transform in the presence of a (here iid) random environment. As an instructive example, consider the above supercritical Galton-Watson process $(Z_n)_{n\geq 0}$, but now in iid random environment. More precisely, let $\mathbf{e} = (\mathbf{e}_n)_{n\geq 0}$ be a sequence of iid random offspring distributions with random offspring means $m(\mathbf{e}_n)$. Supercriticality mandates that $\mathbb{E}\log m(\mathbf{e}_0) > 0$. Given $\mathbf{e} = \xi = (\xi_n)_{n\geq 0}$, ξ_n then denotes the common offspring distribution of all individuals in generation n. Put $m_n(\mathbf{e}) := m(\mathbf{e}_0 \cdot ... \cdot m(\mathbf{e}_{n-1})$ for $n \geq 1$. Then $m_n(\mathbf{e})^{-1} Z_n$, $n \geq 0$, is again a nonnegative martingale and thus a.s. convergent to a limit $W = W(\mathbf{e})$. The recursive structure being the same as in the case of a constant environment, it is not surprising that W still satisfies the distributional equation stated above, viz. $W \stackrel{d}{=} \sum_{k=1}^{Z_1} m^{-1} W_k$. However, due to the environment, the W_k on the right-hand side are no longer independent, but conditionally independent given $\mathbf{e} = \xi$ with common distribution given by the law of $W([\xi]_1)$, where $[\xi]_n = (\xi_n, \xi_n, ...)$ for $n \geq 1$.

We proceed to a general definition of the smoothing transform in iid random environment and confine to the homogeneous case. Let E be the set of distributions on $[0,\infty)^\infty$ endowed with the Borel σ-field $\mathcal{E}$ generated by the total variation norm. Further let $\mathbf{e} = (\mathbf{e}_n)_{n\geq 0}$ be a sequence of iid random elements taking values in $(E, \mathcal{E})$ and let $\mathcal{P}(E^\infty, [0,\infty))$ denote the set of stochastic kernels from $(E^\infty, \mathcal{E}^\infty)$ to $[0,\infty)$ endowed with the Borel σ-field. Finally, let $T = (T_n)_{n\geq 1}$ be a sequence of nonnegative random variables such that the conditional law of T given $\mathbf{e} = \xi$ equals ξ_0. Then the (homogeneous) smoothing transform $\mathcal{S}_T$ associated with T is defined as the mapping which sends a kernel $P \in \mathcal{P}(E^\infty, [0,\infty))$ to the conditional law of $\sum_{k\geq 1} T_k X_k$ given $\mathbf{e}$, when $X_1, X_2, ...$ are conditionally iid given $\mathbf{e} = \xi$ with common law $P([\xi]_1, \cdot)$. In other words,

$$\mathcal{S}_T P(\xi, \cdot) \;=\; \mathbb{P}\left(\sum_{k\geq 1} T_k X_k \in \cdot \,\middle|\, \mathbf{e} = \xi \right).$$

If $\mathcal{S}_T P = P$, then P is a fixed point of $\mathcal{S}_T$.

After these settings it is easily seen that the conditional law of W as defined above for the Galton-Watson process in random environment constitutes a fixed point of $\mathcal{S}_T$ with $T_n = \mathbf{1}_{\{Z_1 \geq n\}} m(\mathbf{e}_0)^{-1}$ for $n \geq 1$. As another example, where such fixed points occur, we will briefly discuss the limits of normalized weighted branching processes in iid random environment which have been studied in some

detail by Kuhlbusch [6]. In particular, conditions are given there for these limits to be nondegenerate.

References

[1] G. Alsmeyer, J. D. Biggins, and M. Meiners. The functional equation of the smoothing transform. *Ann. Probab.*, 40(5):2069–2105, 2012.

[2] G. Alsmeyer and M. Meiners. Fixed points of inhomogeneous smoothing transforms. *J. Difference Equ. Appl.*, 18(8):1287–1304, 2012.

[3] G. Alsmeyer and M. Meiners. Fixed points of the smoothing transform: two-sided solutions. *Probab. Theory Related Fields*, 155(1-2):165–199, 2013.

[4] J. D. Biggins and A. E. Kyprianou. Fixed points of the smoothing transform: the boundary case. *Electron. J. Probab.*, 10:no. 17, 609–631 (electronic), 2005.

[5] U. Rösler. A limit theorem for "Quicksort". *RAIRO Inform. Théor. Appl.*, 25(1):85–100, 1991.

[6] D. Kuhlbusch. On weighted branching processes in random environment. *Stochastic Process. Appl.*, 109(1):113–144, 2004.

Radix Sort on Markov Sources

Kevin Leckey

(joint work with Ralph Neininger, Henning Sulzbach, Wojciech Szpankowski)

Radix sort is one of the most famous non-comparative integer sorting algorithms. The are several variants of this algorithm including sorting algorithms for numbers in the unit interval and words (in lexicographical order). Both tasks are covered by an algorithm that sorts strings (sequences of symbols drawn from some finite, ordered alphabet Σ) in lexicographical order. In the case of numbers in the unit interval, the corresponding strings are determined by the b-ary expansions of the numbers for some base $b \geq 2$.

The variant of radix sort considered in this talk distributes a list of strings into b sublists according to their first symbol in the underlying alphabet $\Sigma = \{\sigma_1, \dots, \sigma_b\}$. The algorithm continues recursively in each sublist that contains more than one element by splitting the sublist according to the next symbol in each string.

In order to get an idea how radix sort typically performs on strings, it takes a reasonable model for the input. Normally, such an input is considered to be random and, more precisely, the input strings are i.i.d. copies of some string $\Xi = (\xi_i)_{i\geq 1}$.

Under a fairly simple model, known as the Bernoulli source model, $(\xi_i)_{i\geq 1}$ is a sequence of i.i.d. Bernoulli distributed random variables. If B_n denotes the number of bucket operations performed by radix sort on n strings in the Bernoulli source model (where a bucket operations includes reading a symbol from a string and placing the string into a sublist), then, even in this simple model, several interesting asymptotic features of B_n occur. In particular, it is known that as $n \to \infty$, depending on $p = \mathbb{P}(\xi_i = 1)$,

$$\mathbb{E}[B_n] = \frac{1}{H} n \log n + n\omega_p(\log n) + o(n), \tag{1}$$

where $H = -p\log p - (1-p)\log(1-p)$ is the source entropy and $\omega_p : \mathbb{R} \to \mathbb{R}$ might be either a continuous, periodic function or constant depending on whether $\log(p)/\log(1-p)$ is rational or not. It is also known that

$$Var(B_n) \sim \begin{cases} n\omega(\log n), & \text{if } p = 1/2, \\ \sigma^2 n \log n, & \text{if } p \neq 1/2 \end{cases} \tag{2}$$

where ω is a continuous, periodic function and $\sigma^2 > 0$ is a constant depending on p. Finally, there is a normal limit after rescaling:

$$\frac{B_n - \mathbb{E}[B_n]}{\sqrt{Var(B_n)}} \xrightarrow{d} \mathcal{N}(0,1). \tag{3}$$

The main topic of the (first part of the) talk is to check if (1) - (3) also holds for a more general input model.

In fact, it turns out that under the assumption that each string is distributed as $\Xi = (\xi_i)_{i\geq 1}$ for some Markov chain $(\xi_i)_{i\geq 1}$ on $\{0,1\}$ (called the Markov source model), (1)-(3) also holds for suitably generalized constants H and σ^2:

If $P = (p_{ij})_{i,j\in\{0,1\}}$ denotes the transition matrix of the underlying Markov chain and (π_0, π_1) denotes the weights of its stationary distribution, the constants in (1) and (2) become for Markov Sources

$$H = \pi_0 H_0 + \pi_1 H_1, \qquad H_i = -p_{i0}\log(p_{i0}) - p_{i1}\log(p_{i1}),$$

$$\sigma^2 = \frac{\pi_0 p_{00} p_{01}}{H^3}\left(\log\left(\frac{p_{00}}{p_{01}}\right) + \frac{H_1 - H_0}{p_{01}+p_{10}}\right)^2 + \frac{\pi_1 p_{10} p_{11}}{H^3}\left(\log\left(\frac{p_{10}}{p_{11}}\right) + \frac{H_1 - H_0}{p_{01}+p_{10}}\right)^2.$$

In particular, the first case in (2) only occurs for $p_{ij} = 1/2$, $i,j \in \{0,1\}$.

The analysis is based on the following distributional recursion: If B_n^μ denotes the number of Bucket operations for n strings with initial distribution μ (the transition matrix is fixed), then B_n^μ satisfies for $n \geq 2$

$$B_n^\mu \stackrel{d}{=} B_{K_n}^{p_{00}\delta_0 + p_{01}\delta_1} + B_{n-K_n}^{p_{10}\delta_0 + p_{11}\delta_1} + n \tag{4}$$

with $(B_k^{p_{00}\delta_0+p_{01}\delta_1})_{k\geq 0}$, $(B_k^{p_{10}\delta_0+p_{11}\delta_1})_{k\geq 0}$, K_n independent and $K_n \sim B(n, \mu(0))$.

It turns out that the expansion (1) is not sufficient to derive a limit law with the method used in this talk (which is the contraction method) since it usually requires an asymptotic expansion of $\mathbb{E}[B_n]$ up to the order of $o(\sqrt{Var(B_n)})$. This problem is solved by showing that

$$\mathbb{E}[B_n] = \frac{1}{H} n\log n + f(n), \qquad f \text{ Lipschitz continuous},$$

which leads to the necessary asymptotic of the toll term after rescaling (4).

The second part of the talk considers the radix select algorithm. In this one sided version of radix sort, the algorithm returns the element of rank ℓ by recursively splitting the sublist that contains the element of rank ℓ (which is determined by the sizes of the sublists).

Some results derived for Markov sources are:

- The first order asymptotic of the expected number of bucket operations when selecting rank $\ell = \lfloor tn \rfloor + 1$, $t \in [0, 1)$.
- The first order asymptotic of all moments and a law of large numbers for the maximal number of bucket operations (maximized over all $\ell \in \{1, \dots, n\}$).
- A limit law for the number of bucket operations when selecting a rank that is uniformly distributed on $\{1, \dots, n\}$ and independent of the strings (grand averages model).

Moreover, if $Y_n(\ell)$ denotes the number of bucket operations when selecting the string of rank ℓ among n i.i.d. string generated by a *Bernoulli source*, then, in the Skorokhod topology, as $n \to \infty$,

$$\left(\frac{Y_n(\lfloor tn \rfloor + 1) - m(t)n}{\sqrt{n}} \right)_{t \in [0,1]} \xrightarrow{d} G$$

where G is a centered Gaussian process and m is a affine linear function determined by the source. This convergence does not hold for Markov sources due to a lack of continuity in the first order asymptotic of the mean. However, it is still an open problem to check if the marginals of the rescaled process converge for Markov sources.

References

[1] K. Leckey,R. Neininger and H. Sulzbach, *Analysis of radix selection on Markov sources.* Proceedings of the 25th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA'14) (2014), 253–264.

[2] K. Leckey, R. Neininger and W. Szpankowski, *Towards More Realistic Probabilistic Models for Data Structures: The External Path Length in Tries under the Markov Model.* Proceedings ACM-SIAM Symp. Disc. Algo. (SODA) (2013), 877–886.

Dependence and phase changes of random m-ary search trees

Hsien-Kuei Hwang, Ralph Neininger

The m-ary search tree is a class of data structures introduced by Muntz and Uzgalis [9] in 1971 in computer algorithms to support efficient searching and sorting of data. When constructed from a random permutation of n elements, the space requirement of such a *random m-ary search tree* ($m \geq 2$) can be defined in the following self-contained way

$$S_n \stackrel{d}{=} \begin{cases} 0, & \text{if } n = 0 \\ 1, & \text{if } 1 \leq n < m \\ S_{I_1}^{[1]} + \cdots + S_{I_m}^{[m]} + 1, & \text{if } n \geq m \end{cases}$$

where

$$\mathbf{P}(I_1 = i_1, \dots, I_m = i_m) = \frac{1}{\binom{n}{m-1}},$$

when $i_1, \dots, i_m \geq 0$ and $i_1 + \cdots + i_m = n - m + 1$. Except for $m = 2$ (in that case $S_n = n$), the S_n's are for $m \geq 3$ and general n non-degenerate random variables. It is known that S_n exhibits a *phase change phenomenon*: its distribution tends to Gaussian for large n when the branching factor m satisfies $3 \leq m \leq 26$ but does not approach a fixed law when $m \geq 27$; see [8, 7, 2, 1] and the references therein. More precise approximation results are also known; see [5, 3].

On the other hand, we prove that the *total node path length* T_n, defined as the sum of the distances between the root to each *node* of the tree, or, in terms of S_n,

$$T_n \stackrel{d}{=} \begin{cases} 0, & \text{if } n = 0 \leq n < m \\ T^{[1]}_{I_1} + \cdots + T^{[m]}_{I_m} + S^{[1]}_{I_1} + \cdots + S^{[m]}_{I_m}, & \text{if } n \geq m, \end{cases}$$

converges, after properly centered and normalized, to a limit law for all $m \geq 2$, with mean of order $n \log n$ and variance of order n^2.

Our next question was "to which extent does T_n depend on S_n"? To our surprise, the random variables T_n, despite the strong dependence of its definition on S_n, turn out to be asymptotically independent of S_n when $3 \leq m \leq 26$, and dependent when $m \geq 27$. More precise statement of the results are given below.

We prove first that the covariance undergoes another phase change at $m = 13$

$$\mathrm{Cov}(S_n, T_n) \sim \begin{cases} C_R n, & \text{if } 3 \leq m \leq 13; \\ F_c(\beta \log n) n^{\alpha}, & \text{if } m \geq 14; \end{cases}$$

where C_R is a suitable constant and $F_c(z)$ is a 1-periodic function which can be explicitly computed. Here $\alpha - 1 + \beta i$ ($\beta > 0$) denotes the second largest zero (in real part) of the equation

$$z(z+1)\cdots(z+m-2) - m! = 0,$$

This implies the following asymptotic estimates for the correlation coefficient

$$\boxed{\rho(S_n, T_n) \sim \begin{cases} 0, & \text{if } 3 \leq m \leq 26; \\ F_\rho(\beta \log n), & \text{if } m \geq 27, \end{cases}}$$

where F_ρ is another 1-periodic function; see Figure 1 for graphical rendering of the periodic function for a few selected m.

While one might ascribe this counter-intuitive result to the possibly nonlinear correlation between T_n and S_n, we enhance such an asymptotically uncorrelated phenomenon by a stronger joint limit law, which again puts an accent on the asymptotic independence between T_n and S_n when $3 \leq m \leq 26$. For larger m, they are asymptotically dependent and we have a precise characterization of the joint limit law.

Moreover, replacing the total node path length by total key path length (summing over all keys instead of over all nodes) does not change the major phenomena. Indeed, the same types of results hold in a more general setting where S_n satisfies

$$S_n \stackrel{d}{=} S^{[1]}_{I_1} + \cdots + S^{[m]}_{I_m} + \begin{cases} c + o(n^{-\varepsilon}), & \text{if } 2 \leq m \leq 13; \\ o(n^{\alpha - 1}), & \text{if } m \geq 14 \end{cases}$$

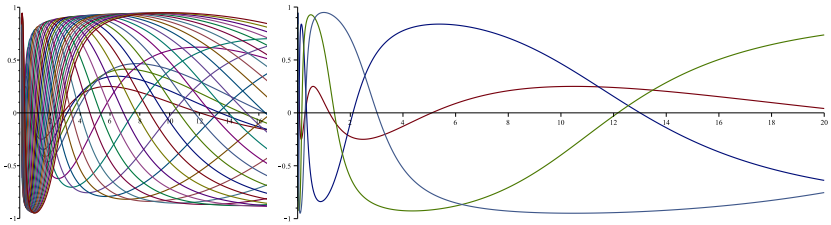

FIGURE 1. *Periodic fluctuations of the limiting correlation coefficient of space requirement and total key path length for n up to $20,000$ and $m = 30, \dots, 175$ (left), and $n \le 200,000$ and $m = 27, 60, 120, 175$ (right).*

and T_n are either of the following two forms

- $T_n \stackrel{d}{=} T_{I_1}^{[1]} + \cdots + T_{I_m}^{[m]} + n + h_n$ with $h_n = o(n)$ and $\left|\sum_n h_n n^{-2}\right| < \infty$,
- $T_n \stackrel{d}{=} T_{I_1}^{[1]} + \cdots + T_{I_m}^{[m]} + S_{I_1}^{[1]} + \cdots + S_{I_m}^{[m]}$.

Finally, the consideration of random m-ary search trees can be extended to other random trees of logarithmic height such as quadtrees, fringe-balanced binary search trees, etc., and the same phenomena hold. Details of all these results will be contained in a forthcoming paper jointly written with Hua-Huai Chern, Michael Fuchs and the two authors of this abstract, see arXiv:1501.05135.

REFERENCES

[1] Chauvin, B. and Pouyanne, N. (2004) m-ary search trees when $m \ge 27$: a strong asymptotics for the space requirements. *Random Struct. Algor.* **24**, 133–154.

[2] Chern, H.-H. and Hwang, H.-K. (2001) Phase changes in random m-ary search trees and generalized quicksort. *Random Struct. Algor.* **19**, 316–358.

[3] Fill, J. A. and Kapur, N. (2004) The space requirement of m-ary search trees: distributional asymptotics for $m \ge 27$. Invited paper, Proceedings of the 7th Iranian Statistical Conference, 2004.

[4] Fill, J.A. and Kapur, N. (2005) Transfer theorems and asymptotic distributional results for m-ary search trees. *Random Struct. Algor.* **26**, 359–391.

[5] Hwang, H.-K. (2003) Second phase changes in random m-ary search trees and generalized quicksort: convergence rates. *Ann. Probab.* **31**, 609–629.

[6] Lew, W. and Mahmoud, H. M. (1994). The joint distribution in multiway search trees. *SIAM J. Comput.* **23**, 1050–1074.

[7] Mahmoud, H.M. (1992) *Evolution of Random Search Trees*, John Wiley & Sons, New York.

[8] Mahmoud, H. M. and Pittel, B. (1989). Analysis of the space of search trees under the random insertion algorithm. *J. Algorithms* **10**, 52–75.

[9] R. Muntz and R. Uzgalis, Dynamic storage allocation for binary search trees in a two-level memory, in *Proceedings of the Princeton Conference on Information Sciences and Systems*, **4** (1971) 345–349.

[10] Neininger, R. (2001) On a multivariate contraction method for random recursive structures with applications to Quicksort. *Random Struct. Algorit.* **19**, 498–524.

[11] Neininger, R. and Rüschendorf, L. (2004) A general limit theorem for recursive algorithms and combinatorial structures. *Ann. Appl. Probab.* **14**, 378–418.

Random permutation structures

Alexander Gnedin

Symmetric groups S_n $(n = 1, 2, \dots)$ can be endowed in many ways by a system of n-to-1 projections $\pi_n : S_n \to S_{n-1}$, which enable one to consider the uniform distributions on S_n's as marginal distributions of a random permutation growth process, representable by a path in an infinite 'permutation tree'. *Permutation structures* are more general growth processes with marginal distributions conditionally uniform given the value of a permutation statistic.

For $\mathtt{s}$ a general permutation statistic (numerical, multivariate or 'infinite-dimensional' like integer partition), we say that $\mathtt{s}$ is compatible with given projections if $\#\{\sigma \in S_n : \pi_n(\sigma) = \sigma',\ \mathtt{s}(\sigma) = s\}$ depends on $\sigma' \in S_{n-1}$ through $\mathtt{s}(\sigma')$. Probability law of a permutation structure is defined by a sequence of probability distributions P_n on S_n which are consistent (i.e. $\pi_n(P_n) = P_{n-1}$) and satisfy the sufficiency condition: $P_n(\sigma)$ depends only on $\mathtt{s}(\sigma)$ for every n and $\sigma \in S_n$. The conditions define a class of infinite transient Markov chains with common backward transition probabilities.

Many classic statistics are compatible with one of the following plausible projections: deleting n from the cycle notation of σ (type I), deleting n from the one-row notation of σ (type II), and deleting $\sigma(n)$ from the one-row notation of σ then ranking $\sigma(1), \dots, \sigma(n-1)$ by increasing bijection with $[n-1]$ (type III).

For specified projections and statistic $\mathtt{s}$, the *boundary problem* asks one to identify the set of extreme elements (the boundary) within the convex set of permutation structures. From a viewpoint, the problem belongs to the circle of questions around de Finetti's theorem, random symmetries and sufficiency [8], and is intrinsically related to path-counting in the graded graph constructed from the permutation tree by gluing vertices of level n with the same value of $\mathtt{s}$. We shall sketch a few instances to illustrate what kind of solution may appear.

For $\mathtt{s}$ the number of cycles (and π_n's of type I) the extremes are given by the familiar Ewens' distributions

$$P_n(\sigma) = c_n(\theta)\theta^{\mathtt{s}(\sigma)},\ \sigma \in S_n, \tag{1}$$

where $\theta \in [0, \infty]$ and c_n is a normalisation constant [9]. A permutation structure of this kind can be encoded in a sequence of independent random variables $r_1, r_2, \dots$ with r_j taking values in $\{1, \dots, k\}$, and in this coordinate system the number of cycles is representable as a sum $\sum_{j=1}^{n} \varphi_j(r_j)$. The additive representation also holds for $\mathtt{s}$ the number of records or the number of inversions, with the consequence that the associated extreme permutation structures are still of the form (1). In contrast to that, the number of descents is not an additive statistic – the boundary in this case is only countable, with the extreme permutation structures related to random permutations knows as 'a-shuffles' (or inverse riffle-shuffles) [4].

When $\mathtt{s}$ is the integer partition derived from factoring permutation in cycles (and if projections are of type I), a permutation structure can be associated with an exchangeable partition of $\mathbb{N}$. By Kingman's 'paintbox' representation the extremes are parameterised by points of the infinite-dimensional simplex

$\Delta = \{(x_1, x_2, \dots) : x_1 \geq x_2 \geq \cdots \geq 0, \sum_j x_j \leq 1\}$. A distinguished permutation structure generalising (1) has two parameters [10] and can be simulated by a simple Markov chain.

When $\mathtt{s}$ is the set of descents and projections are of type II, a permutation structure corresponds to a random order on $\mathbb{N}$ with the property that every infinite suborder has the same distribution (spreadability) [7, 8]. The boundary is parameterised by partitions of $[0, 1]$ in oriented intervals [5].

The statistic 'the set of records in permutation' is compatible with projections of both types II and III. For type III, the boundary is parameterised by points of an infinite-dimensional simplex [1], in line with Pitman's [10] theory of partially exchangeable partitions. For type II, a typical permutation structure has a record in almost every position, with relatively rare exceptions [2]. In the latter setting the topology of the boundary is that of a cone with uncountable totally disconnected base. For some unknown reason the boundary is of the same kind as for another graded graph (Young-Fibonacci lattice [6]) related to permutations only indirectly via a solitaire game introduced in [3].

References

[1] A. Gnedin, *Coherent random permutations with biased record statistics*, Discrete Math. **311** (2011), 80–91.

[2] A. Gnedin, V. Gorin, *Record-dependent measures on the symmetric groups*, Random Structures and Algorithms (2014).

[3] A. Gnedin and S. Kerov, *Fibonacci solitaire*, Random Structures and Algorithms, **20** (2002), 71–88.

[4] A. Gnedin and G. Olshanski, *The boundary of the Eulerian number triangle*, Moscow Math. J. **5**, (2006), 461–475.

[5] A. Gnedin and G. Olshanski, *Coherent random permutations with descent statistic and the boundary problem for the graph of zigzag diagrams*, Intern. Math. Res. Notes, Article ID 51968 (2006), 1–39.

[6] F. Goodman and S. Kerov, *The Martin boundary of the Young-Fibonacci Lattice*, J. Algebr. Combinatorics, **11** (2000), 17–48.

[7] S. Jacka and J. Warren, *Random orderings of the integers and card shuffling*, Stoch. Proc. Appl., **117** (2007), 708–719

[8] O. Kallenberg, *Probabilistic symmetries and invariance principles*, Springer, 2005.

[9] G. Olshanski, *Random permutations and related topics*, In: The Oxford Handbook of Random Matrix Theory, Chapter 25, Oxford University Press, 2011.

[10] J. Pitman, *Exchangeable and partially exchangeable random partitions*, Prob. Theory Related Fields, **102** (1995), 145–158.

And/or trees: A local limit point of view

Cécile Mailler

(joint work with Nicolas Broutin)

Random Boolean trees, and in particular and/or trees have been thoroughly studied in the literature - mainly by Analaytic Combinatorics methods. We present a new approach of such problems based on local limit properties of the considered

random trees. With this approach, we intend to be more universal than the Analytic Combinatorics approach, and thus become able to study random tree models that cannot directly be studied by Analytic Combinatorics.

Literature

Let us first briefly review the literature concerning and/or trees. We focus here on two models: the Catalan tree and the binary search tree (BST). We will end this short overview by two open problems, being our motivations for this work.

An and/or tree (cf. Figure 1) is a plane rooted tree (with no node having exactly one child) whose internal nodes are labelled from the set $\{\wedge, \vee\}$ and the leaves from the set $\{x_1, \bar{x}_1, \ldots, x_k, \bar{x}_k\}$, with k an integer. This labelled tree is thus equivalent to a Boolean expression, where $\wedge$ is the logical connective `and`, $\vee$ is the logical connective `or`, and $\bar{x}_i$ is the negation of x_i. This Boolean tree thus calculates a Boolean function from $\{0,1\}^k$ onto $\{0,1\}$.

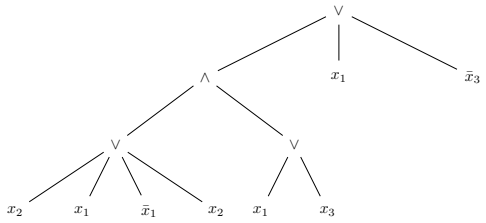

Figure 1. An and/or tree of size 8 calculating the Boolean function $(x_i)_{i\geq 1} \mapsto x_1$.

The following models propose different ways to pick up at random an and/or tree, being interested in the induced distribution on the set of Boolean functions.

The Catalan tree model. The first random and/or model was introduced by Paris et al. [5] and later studied by Lefmann and Savický [4] and Chauvin et al. [1] – the latter article is the seminal Analytic Combinatorics approach of the question. They pick up a tree uniformly on the set of binary and/or trees having n leaves (and labelled with k variables), and denote by $\mathbb{P}_{n,k}$ the law of the random Boolean function it calculates. They then prove that $\mathbb{P}_{n,k}$ converges to an asymptotic probability distribution $\mathbb{P}_k$ when the size of the tree n tends to infinity. With an interesting Analytic Combinatorics approach, Kozik later proved the following result: for all Boolean function f, asymptotically when k tends to infinity,

$$\mathbb{P}_k(f) := \lim_{n\to+\infty} \mathbb{P}_{n,k}(f) = \Theta\left(\frac{1}{k^{L(f)+1}}\right),$$

where $L(f)$, called the complexity of the Boolean function f is defined as the size (number of leaves) of the smallest trees calculating f, except for the two constant functions `true` and `false`, whose complexity is defined to be 0.

The BST model. The random binary search tree of size n is the unlabelled binary search tree obtained from n i.i.d. random variables (uniformly distributed on $(0,1)$). Label this random binary tree uniformly at random – meaning that each internal node chooses uniformly at random a label in $\{\wedge, \vee\}$, and each leaf chooses a label uniformly at random in the set $\{x_1, \bar{x}_1, \ldots, x_k, \bar{x}_k\}$, independently from each other. Denote by $p_{n,k}$ the distribution of the random Boolean function it calculates. Chauvin et al. [2] have proven that

$$p_{n,k}(\texttt{true}) = p_{n,k}(\texttt{false}) \to \frac{1}{2},$$

when the size n of the BST tends to infinity. A behaviour completely distinct from the one of the Catalan tree model!

What property of the underlying random trees is the cause of the difference of behaviour in the Catalan and the BST model?

An intuitive explanation is that the Catalan tree has leaves at finite distance from the root whereas the smallest distance between the root and a leaf in the random BST tends to infinity when the size of the BST tends to infinity: this is a local limit property.

The Ford tree model. The Ford tree (or α-model) – studied by Chen et al [3] – is quite a natural random tree, since it interpolates between the Catalan and the BST random trees thanks to a parameter. Our idea is to take the Ford tree of size n, label it at random to get a random and/or tree and study the law of the random Boolean function it calculates. To our knowledge, this and/or tree model cannot be tackled by Analytic Combinatorics methods.

What is the behaviour of the Ford tree model?

Set up and results

For all (deterministic) tree t, we denote by $\hat{t}$ its uniform random labelling: $\hat{t}$ is thus a random and/or tree and calculates a random Boolean function denoted by $f[\hat{t}]$. Our idea is to consider a sequence of random trees $(T_n)_{n\geq 1}$ and study the sequence of random Boolean functions $f[\hat{T}_n]$: does it converge in distribution when n tends to infinity? under which condition is the asymptotic distribution degenerate? can we say more about this asymptotic distribution when it is not degenerate?

Convergence to an asymptotic distribution. We say that a sequence of trees $(t_n)_{n\geq 1}$ converges locally to a tree t_∞ if, and only if, for any integer h, for large enough n, the trees t_n and t_∞ are identical up to height h.

The following result is a generalisation of the BST model described above, and answers our first motivation:

Assume that the sequence of trees $(t_n)_{n\geq 1}$ converges locally to an infinite tree without leaves. Then, asymptotically when n tends to infinity,

$$\mathbb{P}(f[\hat{t}_n] = \texttt{true}) = \mathbb{P}(f[\hat{t}_n] = \texttt{false}) \to \frac{1}{2}.$$

We can also prove convergence to an asymptotic distribution in the following case,

> Assume that the sequence of trees $(t_n)_{n\geq 1}$ converges locally to an infinite tree t_∞ having finitely many ends. Then, for all Boolean function f, asymptotically when n tends to infinity, $\mathbb{P}(f[\hat{t}_n] = f) \to \mathbb{P}(f[\hat{t}_\infty] = f)$.

This result can then be applied to sequences of random trees that converge locally in distribution to infinite tree with finitelay many ends: among them critical Galton-Watson trees conditioned on their size, or the Ford tree. Can we say more about the distribution of $f[\hat{t}_\infty]$ in this last case?

More information in the non-satured case. In the case when T_n converges locally to an infinite tree T_∞ having one unique end, under some additional assumptions, among them self-similarity down the spine and a control of the size of the trees attached to the spine of the limit tree, we are able to obtain that, for all Boolean function f,

$$\mathbb{P}(f[\hat{T}_n] = f) \to \mathbb{P}(f[\hat{T}_\infty] = f) = \Theta\left(\frac{1}{k^{L(f)+1}}\right).$$

This result is very satisfying because of its universality. However, one can regret that its hypothesis are not trivial to check in practise. We are able to apply our theorem to any critical Galton-Watson tree conditioned on its size as well as to the Ford model, giving an answer to our second motivation.

References

[1] B. Chauvin, Ph. Flajolet, D. Gardy and B. Gittenberger, *And/Or trees revisited*, Combinatorics, Probability and Computing, **13** (2004), 475–497.

[2] B. Chauvin, D. Gardy and C. Mailler, *A sprouting tree model for random boolean functions*, Random Structures and Algorithms, doi: 10.1002/rsa.20567 (2014).

[3] B. Chen, D. Ford and M. Winkel *A new family of Markov branching trees: the alpha-gamma model* Electronic Journal of Probability **14** (2009), 400–430.

[4] H. Lefmann and P. Savický, *Some typical properties of large And/Or Boolean formulas*, Random Structures & Algorithms, **10** (1997), 337–351.

[5] J. B. Paris, A. Vencovská and G. M. Wilmers, *A natural prior probability distribution derived from the propositional calculus*, Annals of Pure and Applied Logic, **70** (1994), 243–285.

A functional central limit theorem for branching random walks, and applications to random trees

Zakhar Kabluchko

(joint work with Rudolf Grübel)

Consider a branching random walk in discrete or continuous time. Denote by π_t the point process recording the positions of particles at time t. For $\beta \in \mathbb{C}$ let

$$m(\beta) := \mathbb{E} \sum_{z\in\pi_1} \mathrm{e}^{\beta z}$$

be the Laplace transform of the intensity of π_1. Under suitable conditions, Uchiyama [1] and Biggins [2] proved a martingale convergence of the form

$$W_t(\beta) := \frac{1}{m(\beta)^t} \sum_{z \in \pi_t} \mathrm{e}^{\beta z} \xrightarrow[t\to\infty]{a.s.} W_\infty(\beta).$$

Moreover, for sufficiently small $\varepsilon > 0$, the convergence is a.s. uniform and the limit $W_\infty(\beta)$ is a random analytic function on the disk $\{|\beta| \le \varepsilon\}$. In this talk we will present a functional central limit theorem which states that the random analytic function

$$D_t(u) := m(0)^{\frac{1}{2}t} \left(W_\infty\left(\frac{u}{\sqrt{t}}\right) - W_t\left(\frac{u}{\sqrt{t}}\right)\right)$$

converges to certain Gaussian analytic function, weakly on the space of analytic functions on any disk $\{|u| < R\}$.

We will give several applications of this result to random trees. Let EPL_n be the external path length of a binary search tree with n vertices. From the work of Régnier [3] on the analysis of the QUICKSORT algorithm it is known that there is a martingale convergence of the form

$$\frac{\mathrm{EPL}_n - 2n\log n}{n} \xrightarrow[n\to\infty]{a.s.} \mathrm{EPL}_\infty.$$

Neininger [4] proved a central limit theorem of the form

$$\sqrt{\frac{n}{2\log n}} \left(\mathrm{EPL}_\infty - \frac{\mathrm{EPL}_n - 2n\log n}{n} \right) \xrightarrow[n\to\infty]{d} \mathrm{N}(0,1).$$

Using an embedding of binary search trees into a branching random walk we recover Neininger's result by essentially taking the derivative at 0 in our functional central limit theorem. Moreover, we prove that the convergence in Neininger's central limit theorem is mixing and establish similar results for uniform random recursive trees and, more generally, Pittel trees.

Finally, we discuss some other examples of tail martingale central limit theorems of the above type including the Pólya urn and the central limit theorem for Galton–Watson processes.

REFERENCES

[1] K. Uchiyama, *Spatial growth of a branching process of particles living in* $\mathbb{R}^d$, Ann. Probab., **10**(1982), 896–918.

[2] J. D. Biggins, *Uniform convergence of martingales in the branching random walk*, Ann. Probab., **20**(1992), 137–151.

[3] M. Régnier, *A limiting distribution for quicksort*, RAIRO Inform. Théor. Appl., **23**(1989), 335–343.

[4] R. Neininger, *Refined quicksort asymptotics*, Rand. Struct. Alg., 2013, To appear.

Limits of randomly growing graphs: Theory and examples

Rudolf Grübel

We write $\mathbb{G}[n]$ for the set of simple graphs with node set $[n] := \{1, 2, \ldots, n\}$, $\mathbb{G} := \bigcup_{n=1}^{\infty} \mathbb{G}[n]$. Our two central questions are: First, what does (or could) convergence of a sequence $(G_n)_{n\in\mathbb{N}} \subset \mathbb{G}$ mean? Secondly, what are the possible limits? Of course, these questions are of interest in the context of other combinatorial structure too that are built on $[n]$, such as the permutations $\mathbb{S}[n]$ of $[n]$.

We first discuss three different notions of convergence. The first of these is standard, the second is heavily used and discussed in detail in [3], and for more on the third we refer to [4].

The set $\mathbb{G}$ may be endowed with the structure of a rooted tree: The single element of $\mathbb{G}[1]$ is the root, and two graphs $G \in \mathbb{G}[n], H \in \mathbb{G}[n+1]$ form an edge (G, H) directed away from the root if we obtain G from H by deleting the node $n+1$ and its incident edges. We then say that $(G_n)_{n\in\mathbb{N}}$ converges *projectively* if, from some n_0 onwards, all G_n's stay on the same ray. This is, of course, the ends compactification of the tree $\mathbb{G}$.

Let $\mathrm{Inj}(k, n)$ be the set of one-to-one functions $f : [k] \to [n]$. Each such f defines a function from $\mathbb{G}[n]$ to $\mathbb{G}[k]$ via

$$G = \big([n], E\big) \quad \mapsto \quad G^f := \big([k], \big\{\{i, j\} : \{f(i), f(j)\} \in E\big\}\big).$$

Let μ_k^G be the push-forward of the uniform distribution on $\mathrm{Inj}(k, n)$ under the mapping $f \to G^f$. Say that $(G_n)_{n\in\mathbb{N}}$ converges in the *subgraph sampling topology* if $(\mu_k^{G_n})_{n\in\mathbb{N}}$ converges for all $k \in \mathbb{N}$.

For our third notion we start with a Markov chain $X = (X_n)_{n\in\mathbb{N}}$ that is adapted to the graded state space $\mathbb{G}$ in the sense that $P(X_n \in \mathbb{G}[n]) = 1$ for all $n \in \mathbb{N}$. We say that $(G_n)_{n\in\mathbb{N}}$ converges *in the Doob-Martin topology* (associated with X) if the conditional distributions $\mathcal{L}\big((X_1, \ldots, X_k)\big|X_n = G_n\big)$ converge for all $k \in \mathbb{N}$ (here we assume for simplicity that $G_n \in \mathbb{G}[n]$ for all $n \in \mathbb{N}$). It is known that, in this topology, $X_n \to X_\infty \in \partial\mathbb{G}$ almost surely, and that the limit X_∞ generates the tail σ-field of X up to null sets. Also, if the chain is of perfect memory type, meaning that X_{n-1} is a function of X_n for all $n \in \mathbb{N}$, then Doob-Martin convergence coincides with projective convergence.

We note in passing that all three topologies can be seen as variants of the Stone-Čech procedure: Given a countable family $\mathcal{F}$ of bounded functions $f : \mathbb{G} \to \mathbb{R}$, we embed $\mathbb{G}$ into the compact space $\mathbb{R}^{\mathcal{F}}$ via $G \mapsto \big(f \mapsto f(G)\big)$.

We now compare these topologies in the context of one of the most famous graph models, going back to Erdős and Rényi. For this, we consider the Markov chain X starting at the single element of $\mathbb{G}[1]$, where in the move from n to $n+1$ the edges $\{i, n+1\}$, $i \in [n]$, are added with probability θ, independently of each other. Clearly, X is of perfect memory type, so that the Doob-Martin topology leads to the projective limit.

Now let Π_n be uniformly distributed on $\mathbb{S}[n]$, with the Π_n's being independent of each other and of X. We can then, with some abuse of notation, define the *randomly relabelled* chain $Y = (Y_n)_{n\in\mathbb{N}}$ by $Y_n := \Pi_n \circ X_n$.

Theorem 1. [1] *For the randomly relabelled Erdős-Rényi chain, the Doob-Martin topology is the same as the subgraph sampling topology.*

As another, much more demanding example, we consider a tree growth model, the *Rémy chain*. (The material below is based on joint work with Steve Evans and Anton Wakolbinger).

We need some more notation. By a *full binary tree* we mean a subset x of the set $\{0,1\}^\star$ of all finite 0-1 sequences with the properties that $(u_1,\dots,u_l) \in x$ with $l \neq 0$ implies that $(u_1,\dots,u_{l-1})$ and $(u_1,\dots,u_{l-1},1-u_l)$ are also elements of x (i.e. nodes of the tree); leaves are nodes without descendants. We write $\mathbb{B}_n$ for the set of binary trees with n leaves and $\mathbb{B}[n]$ for the set of labelled such trees, with the leaves labelled by the elements of $[n]$. We may now describe the Rémy algorithm as generating a Markov chain adapted to $\mathbb{B}[n]$, with the transition from n to $n+1$ proceeding in three steps: First, we choose one of the $2n-1$ nodes v of X_n uniformly at random. Then we insert a 'cherry' (the single element of $\mathbb{B}_2$) at v. Finally, we append the subtree of v in X_n to a randomly chosen twin; the other one, a leaf, receives the label $n+1$. We note that $\mathbb{B}[n] = \mathbb{B}_n \times \mathbb{S}[n]$, and that X_n is uniformly distributed on $\mathbb{B}[n]$. Again, X is of perfect memory type. Let Y be the chain that results if we remove the labels; clearly, Y_n is uniformly distributed on $\mathbb{B}_n$.

Trees are graphs, but it is easy to see that a direct application of subgraph sampling does not give interesting limits. This naturally leads to the question of what the proper substructures are if we want to adapt the sampling idea. Here it turns out that *sampling from the leaves* is the right notion: With a subset of the leaves of Y_n we associate the minimal binary tree that contains these leaves. As in the general graph case, this leads to functions from $\mathbb{B}_n$ to $\mathbb{B}_k$ and associated distributions μ_k^y. We obtain the following analogue of Theorem 1.

Theorem 2 [2] *For the unlabelled Rémy chain, the Doob-Martin topology is the same as the leaves sampling topology.*

These results provide some answers for the first of the introductory questions. For the description of the limits in the context of Theorem 1 (by *graphons*) we refer the reader to [3]. In [2] the Doob-Martin boundary of the Rémy chain is described in detail: It turns out that the boundary points can be represented by real trees, together with a diffuse probability measure and a kernel that represents the left-right structure.

References

[1] R. Grübel, *Persisting randomness in randomly growing discrete structures: graphs and search trees*, available at: `arXiv:1407.0808`.

[2] S. N. Evans, R. Grübel and A. Wakolbinger, *Doob–Martin boundary of Rémy's tree growth chain*, available at: `arXiv:1411.2526`.

[3] L. Lovász, *Large networks and graph limits*, American Mathematical Society Colloquium Publications, vol. 60, American Mathematical Society, Providence, RI, 2012.

[4] W. Woess, *Denumerable Markov chains. Generating functions, boundary theory, random walks on trees*, European Mathematical Society (EMS), Zürich, 2009.

Almost sure convergence to the Quicksort Process

Uwe Roesler

The underlying algorithmic challenge is to find the first l-smallest elements in order of an input of n-reals. By a suggestion of Conrado Martinez via Quickselsort (see Wikipedia or Partial Quicksort) we use a variant of Quicksort for sorting.

Input: The input is an iid sequence of random numbers with a continuous distribution.

Algorithm: Use (deterministic) Quicksort with pivot always the first element of the list and continue with the leftmost list needing further treatment.

Output: The l smallest elements in order published at time of identification.

Let $(U_i)_i$ be a sequence of iid rvs with a uniform distribution on the unit interval. Let $X(n,l)$ be the number of comparisons in order to find all l-smallest elements using the input $(U_1, U_2, \ldots, U_n)$ and the above algorithm. Notice, if we interrupt at the time obtaining the l-th smallest, then we made all necessary comparisons for that goal and no one more.

We have a recursion of the $X(n,.)$. Interpret now l as time and consider the normalized rvs

$$Y_n(\frac{l}{n}) = \frac{X(n,l) - EX(n,l)}{n}$$

for $l = 1, 2, \ldots, n$. They satisfy again some recursion. Extend Y_n to a rv with values in the space of cadlag functions. (Actually we use left continuous functions with existing right limits.) Then the main statement of the talk is:

Theorem. Y_n converges almost surely in Skorodhod metric on D to the Quicksort Y process.

The existence of the Quicksort process Y was established in Ragab-Roesler 2014 via an infinite sum. The Quicksort process satisfies a stochastic fixed point equation

$$Y \stackrel{d}{=} (UY^1(1 \wedge \frac{t}{U}) + (1-U)Y^2(\frac{t-U}{1-U} \vee 0) + C(U,t))_t$$

Here C is a know function, Y^1, Y^2, U are independent, U is uniformly distributed and Y^1, Y^2, Y have all the same distribution. ($Y(0) = 0$ by definition.) There is only one distribution on D satisfying this if we require $EY(t) = 0$ for all t and finite second moment of $Y(t)$ for one and then for all t.

The convergence of finite dimensional distributions of Y_n to Y was shown in Ragab-Roesler. The convergence of $Y_n(1)$ to the Quicksort distribution in Roesler 1991 using the contraction method and earlier $Y_n(1) \to Y(1)$ a.s. by Régnier using a martingale argument. For our purposes we construct a specific process Y satisfying the above recursion also a.e. using a weighted branching process connected to the input $(U_i)_i$.

References

[1] M. Ragab and U. Roesler *The Quicksort process*, Stochastic processes and their Applications **124** (2014), 1036-1054. Online available under ScienceDirect

[2] M. Régnier *A limit distribution for Quicksort*, Informatique théorique et Applications/Theoretical Informatics and Applications **23** (1998), 335-343.

[3] U. Roesler *A limit theorem for "Quicksort"*, Informatique théorique et Applications/Theoretical Informatics and Applications **25** (1991), 85–100.

Reporter: Ralph Neininger

Participants

Prof. Dr. Louigi Addario-Berry
Department of Mathematics & Statistics
McGill University
Burnside Hall
805 Sherbrooke Street West
Montreal QC H3A 2K6
CANADA

Prof. Dr. Gerold Alsmeyer
Institut f. Mathematische Statistik
Universität Münster
Einsteinstr. 62
48149 Münster
GERMANY

Anna Ben-Hamou
LPMA / UMR 7599
Université Denis Diderot (Paris VII)
Boite Courrier 7012
75251 Paris Cedex 05
FRANCE

Prof. Dr. Jean Bertoin
Institut für Mathematik
Universität Zürich
Winterthurerstr. 190
8057 Zürich
SWITZERLAND

Dr. Shankar Bhamidi
Department of Statistics and
Operations Research
University of North Carolina
Chapel Hill, NC 27599-3260
UNITED STATES

Prof. Dr. Sébastien Bubeck
Dept. of Operations Research &
Financial Eng.
Princeton University
Sherred Hall 225
Princeton, NJ 08544-1000
UNITED STATES

Prof. Dr. Luc Devroye
School of Computer Science
McGill University
Montreal Quebec H3A 2K6
CANADA

Prof. Dr. Thomas Duquesne
Laboratoire de Probabilités et
Modèles Aléatoires
Université Paris VI, Boite 188
4, Place Jussieu
75252 Paris Cedex 05
FRANCE

Prof. Dr. James Allen Fill
Dept. of Applied Mathematics &
Statistics
The Johns Hopkins University
Whitehead Hall 306-F
Baltimore, MD 21218-2682
UNITED STATES

Prof. Dr. Bernhard Gittenberger
Institut f. Diskrete Mathematik &
Geometrie
Technische Universität Wien
Wiedner Hauptstr. 8 - 10
1040 Wien
AUSTRIA

Prof. Dr. Alexander Gnedin
School of Mathematical Sciences
Queen Mary
University of London
Mile End Road
London E1 4NS
UNITED KINGDOM

Prof. Dr. Rudolf Grübel
Institut f. Mathematische Stochastik
Universität Hannover
Welfengarten 1
30167 Hannover
GERMANY

Prof. Dr. Hsien-Kuei Hwang
Institute of Statistical Science
Academia Sinica
Taipei 11529
TAIWAN

Prof. Dr. Svante Janson
Matematiska Institutionen
Uppsala Universitet
Box 480
751 06 Uppsala
SWEDEN

Prof. Dr. Zakhar Kabluchko
Universität Ulm
Fakultät für Mathematik
Institut für Stochastik
Helmholtzstr. 18
89069 Ulm
GERMANY

Igor Kortchemski
Institut für Mathematik
Universität Zürich
Winterthurerstr. 190
8057 Zürich
SWITZERLAND

Kevin Leckey
Institut für Mathematik
J. W. Goethe-Universität
Robert-Mayer-Str. 10
60054 Frankfurt am Main
GERMANY

Prof. Dr. Gabor Lugosi
Department of Economics
Pompeu Fabra University
Ramon Trias Fargas 25-27
08005 Barcelona
SPAIN

Dr. Cecile Mailler
Department of Mathematical Sciences
University of Bath
Bath BA2 7AY
UNITED KINGDOM

Prof. Dr. Jean-Francois Marckert
LABRI
Université Bordeaux 1
351, cours de la Liberation
33405 Talence Cedex
FRANCE

Dr. Dieter Mitsche
Laboratoire J.-A. Dieudonne
Université de Nice
Sophia Antipolis
06108 Nice Cedex 2
FRANCE

Noela Müller
Institut für Mathematik
J. W. Goethe-Universität
Robert-Mayer-Str. 10
60054 Frankfurt am Main
GERMANY

Prof. Dr. Tobias Müller
Mathematisch Instituut
Universiteit Utrecht
Budapestlaan 6
P. O. Box 80.010
3508 TA Utrecht
NETHERLANDS

Prof. Dr. Ralph Neininger
Institut für Mathematik
J. W. Goethe-Universität
Robert-Mayer-Str. 10
60325 Frankfurt am Main
GERMANY

Prof. Dr. Ludger Rüschendorf
Institut f. Mathematische Stochastik
Universität Freiburg
Eckerstr. 1
79104 Freiburg i. Br.
GERMANY

Prof. Dr. Uwe Rösler
Mathematisches Seminar
Christian-Albrechts-Universität Kiel
24098 Kiel
GERMANY

Henning Sulzbach
Project RAP
INRIA Paris-Rocquencourt
Domaine de Voluceau
B. P. 105
78153 Le Chesnay Cedex
FRANCE