Fourth Colloquium on Mathematics and Computer Science DMTCS proc. **AG**, 2006, 1–30

Concentration Properties of Extremal Parameters in Random Discrete Structures

Michael Drmota[1]

[1]*Institute of Discrete Mathematics and Geometry, TU Wien, Wiedner Hauptstrasse 8-10, A-1040 Vienna, Austria*

The purpose of this survey is to present recent results concerning concentration properties of extremal parameters of random discrete structures. A main emphasis is placed on the height and maximum degree of several kinds of random trees. We also provide exponential tail estimates for the height distribution of scale-free trees.

Keywords: concentration inequalities, random discrete structures, random trees

Contents

1 Introduction

Concentration of distribution has always been a *hot topic* in probability theory. In fact, the law of large numbers, the central limit theorem and the law of the iterated logarithm are very prominent examples of the *early times* of probability theory where concentration occurs. However, even recently several books and survey articles on concentration have been published, we just mention few of them (22; 44; 65; 67).

Usually concentration is *easy* to prove if the random variable of interest can be written as a sum of (almost) independent random variables. However, if one is interested in extremal parameters the situation gets much more involved.

Of course, there is a counterpart to the central limit theorem, there are only 3 kinds of possible limiting distribution of the maximum of n iid random variables $X_1, \ldots, X_n$, the most prominent is the so-called extremal value distribution or Gumbel distribution (with distribution function $F(x) = \exp(-e^{-x})$).[(i)] But there are only very few cases where one can apply this theorem directly. Nevertheless, the Gumbel distribution appears in several applications, even if the random variables involved are neither identically distribution nor independent.

The aim of this article is to survey some recent results and methods that have been developed to prove concentration and distributional result of extremal parameters of random discrete structures. Of course, it is impossible to present a complete picture. The subsequent choice of topics also reflects the author's interest. Nevertheless, we have tried to collect a representative sample of results and methods.

The first two problems, the chromatic number of random graphs and the travelling salesman problem, can be seen as functionals of iid random variables, where one can use either Azuma's inequality or Talagrand's inequality to provide concentration results. The third example, the longest increasing subsequence in random permutations is of special interest. First there is mysterious connection to random matrices and second, the methods used are more analytic than probabilitic. The next two sections focus on two extremal parameters of random graphs resp. of random trees, the diamter resp. the height and the maximum node degree. We present two different classes of random graphs, the $G(n,p)$-model and a scale-free model that follows the ideas of Barabasi and Albert. Finally, we have collected result on several kinds of random trees. The advantage of random trees is that due to their recursive structure they are usually easy to describe from a combinatorial point of view. The recursive structure is usually translated into recurrences (of different types) for corresponding generating functions that can be – usually – handled analytically.

In a final section we also present a proof that the height of scale-free trees is highly concentrated, that is, we provide exponential tail estimates. For example, this implies that the variance of the height stays bounded.

One also observes that – usually – extremal parameters are concentrated. In our collection there are only two parameters that are not concentrated, the height of Galton-Watson trees and the maximum degree of scale-free trees. In both cases the mean value is of (rather large) order n^α for some $0 < \alpha < 1$, where n is the *size* of the discrete structure. Second, one observes that *small* extremal parameters are *more concentrated* than *large* ones. For example, it seems that – usually – an extremal parameter of order $\log n$ is highly concentrated (which means that the variance is bounded, compare with Section 2.3) Finally, there is another *strange* phenomenon on the side of the proof techniques. In several cases it seems to be more difficult to get asymptotics for the mean than to prove concentration. This is, for example, true if one applies Azuma's or Talagrand's inequality. Further the concentration proof for the height of scale-free trees (Section 8) does not provide an asymptotic expansion for the mean, either.

2 Types of Concentration

Concentration is usually not formally defined. For our purposes we say that a sequence of random variables X_n is concentrated if there exists a sequence $a(n)$ with $a(n) \to \infty$ (as $n \to \infty$) and with

$$\lim_{n\to\infty} \mathbb{P}\left\{ \left| \frac{X_n}{a(n)} - 1 \right| \ge \varepsilon \right\} = 0 \tag{1}$$

for all $\varepsilon > 0$. Equivalently one has $X_n/a(n) \xrightarrow{\mathrm{d}} \delta_1$,[(ii)] where δ_1 is the delta distribution concentrated at 1. (Usually one uses the expected value $\mathbb{E}\, X_n$ for the scaling sequence $a(n)$ if it exists.)

In what follows we will observe different *types of concentration*. More precisely we will distinguish between 3 different kinds and, of course, there is also the case of no concentration.

We will now always assume that $\mathbb{E}\, X_n \to \infty$ (as $n \to \infty$).

2.1 No Concentration

A sequence of random variables X_n is not concentrated if $X_n/a(n) \overset{\mathrm{d}}{\not\longrightarrow} \delta_1$ for all sequences $a(n)$ with $a(n) \to \infty$. For example, if $\mathbf{E}\, X_n^2 \sim c \cdot (\mathbf{E}\, X_n)^2$ for some $c > 1$ then usually there is no concentration. Typically one additionally has

$$\frac{X_n}{\mathbf{E}\, X_n} \xrightarrow{\mathrm{d}} Y, \tag{2}$$

where Y is a random variable with distribution different from δ_1, see Figure 1.

(i) The two other types are the Fréchet type distributions and the Weibull type distributions.

(ii) We use the notation $X_n \xrightarrow{\mathrm{d}} X$ for convergence in distribution resp. weak convergence.

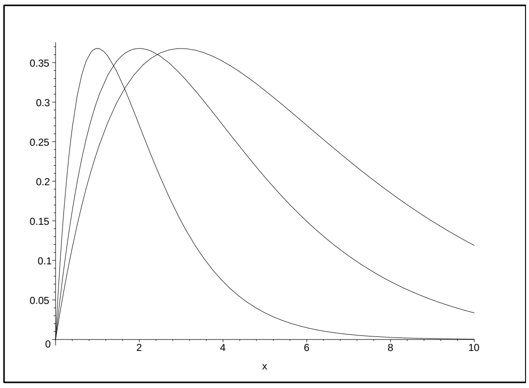

Fig. 1: No concentration

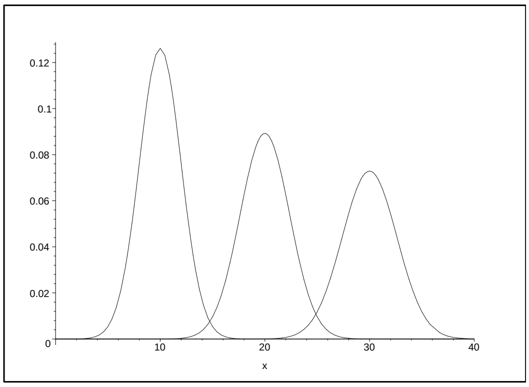

Fig. 2: "Weak" concentration

2.2 "Weak" Concentration

We say that a sequence of random variables X_n is *weakly concentrated* if there are two sequences $a(n), b(n)$ with $a(n) \to \infty$ and $b(n) \to \infty$ but $b(n) = o(a(n))$ (as $n \to \infty$) such that for all $\varepsilon > 0$ there exists $K > 0$ with

$$\limsup_{n\to\infty} \mathbb{P}\left\{ \left| \frac{X_n - a(n)}{b(n)} \right| \geq K \right\} \leq \varepsilon. \tag{3}$$

Note that (3) implies (1). Typically one uses $a(n) = \mathbb{E}\, X_n$ and $b(n) = (\mathbb{V}X_n)^{1/2}$ if second moments exist. For example, if $\mathbb{E}\, X_n^2 \sim (\mathbb{E}\, X_n)^2$ and $\mathbb{V}X_n \to \infty$ (as $n \to \infty$) then we have weak concentration (in this sense) by Chebyshev's inequality. Typically one additionally has

$$\frac{X_n - \mathbb{E}\, X_n}{(\mathbb{V}X_n)^{1/2}} \xrightarrow{\mathrm{d}} Y, \tag{4}$$

where Y is a random variable with distribution different from δ_1, see Figure 2. If Y is normally distributed then we say that X_n satisfies a *central limit theorem*.

2.3 "Strong" Concentration

We call a sequence of random variables X_n *strongly concentrated* (instead of *weakly*) if (3) holds with constant $b(n) = 1$. More precisely we assume that there exists a sequence $a(n)$ with $a(n) \to \infty$ (as $n \to \infty$) such that for all $\varepsilon > 0$ there exists $K > 0$ with

$$\limsup_{n\to\infty} \mathbb{P}\left\{ |X_n - a(n)| \geq K \right\} \leq \varepsilon. \tag{5}$$

Note that (5) is implied if

$$\mathbb{E}\, |X_n - \mathbb{E}\, X_n|^d = O(1) \qquad (n \to \infty) \tag{6}$$

for some $d \geq 1$. If (6) holds for $d = 2$ then the variance stays bounded which is a quite frequent phenomenon for several extremal parameters. Furthermore, one typically has

$$\mathbb{P}\{X_n \leq k\} = F(k - m(n)) + o(1) \tag{7}$$

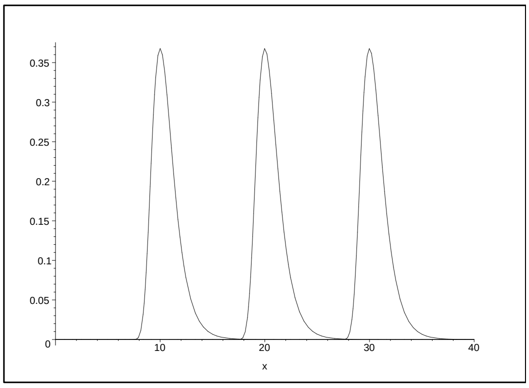

Fig. 3: "Strong" concentration

for some continuous (distribution) function $F(x)$ and some sequence $m(n)$ that is close to the median of X_n. Note that (7) is only valid for integral k. This means that $F(x)$ (appropriately shifted) is a common continuous *envelope* of the discrete distribution functions of X_n. Sometimes such a function $F(x)$ is also called *travelling wave*, compare with Figure 3. Of course, $F(x)$ is not the limiting distribution of X_n (there is usually no limit).

2.4 "Very Strong" Concentration

Condition (5) does not say that $X_n - a(n)$ is bounded (with high probability). Nevertheless, there are cases where X_n is asymptotically concentrated at finitely consecutive values:

$$\mathbb{P}\{m(n) \le X_n \le m(n) + L\} = 1 + o(1). \tag{8}$$

for some sequence $m(n)$ with $m(n) \to \infty$ (as $n \to \infty$) and some integer $L \ge 0$. Here we say that X_n is *very strongly concentrated.*

3 The Chromatic Number of Random Graphs

Let n be a positive integer and p a real number with $0 \le p \le 1$. The *random graph* $G(n,p)$ is a probability space over the set of graphs on the vertex set $\{1, 2, \ldots, n\}$ determined by

$$\mathbb{P}\{(i,j) \in G\} = p$$

for all possible $\binom{n}{2}$ (undirected) edges (i,j) with $1 \le i < j \le n$ with these events mutually independent.

Similarly one also defines random graphs $G(n,m)$, where m is also a given integer with $0 \le m \le \binom{n}{2}$. Here one considers the set of all graphs on the set of vertices $\{1, 2, \ldots, n\}$ with exactly m (undirected) edges and assumes that each of these graphs is equally likely. Due to the law of large numbers $G(n,m)$ will have very similar properties as $G(n,p)$ with $p = m/\binom{n}{2}$.

The *chromatic number* $\chi(G)$ of a graph G is the smallest number k such that there exists a regular k-coloring of the vertices of G, that is, a coloring of k colors of the vertices such that adjacent vertices have different colors.

The chromatic number of random graphs was a intensively studied object in the last decades of the 20th century. The next theorem collects some results of the *expected size* of $\chi(G(n,p))$ if $p = p(n)$ depends on n is a certain way. The results are due to Bollobás (16; 17), Frieze (50), Grimmet and McDiarmid (54), and McDiarmid (68). A more detailed discussion can be found in (56).

In this context we say that a property holds *almost always* if the probability of the exceptional set converges to zero as $n \to \infty$.

Theorem 1

(i) If $C_0/n \le p = p(n) \le (\log n)^{-7}$ (for a proper constant $C_0 > 0$) then almost always

$$\frac{np}{2\log(np) - 2\log\log(np) + 1} \le \chi(G(n,p)) \le \frac{np}{2\log(np) - 40\log\log(np)}.$$

(ii) *If* $(\log n)^{-2} \le p = p(n) \le c$ *(for some arbitrary* $c < 1$*) then almost always*

$$\frac{n}{2\log_b n - \log_b \log_b n} \le \chi(G(n,p)) \le \frac{n}{\log_b n - 6\log_b \log_b n},$$

where $b = 1/(1-p)$.

(iii) *If* $p = p(n) > n^{-\delta}$ *for every* $\delta > 0$ *(and sufficiently large* n*) but* $p = p(n) \le c$ *(for some arbitrary* $c < 1$*) then almost always*

$$\chi(G(n,p)) = \frac{n}{2\log_b n - 2\log_b \log_b n + O(1/p)}.$$

Note that all of these bounds are rather crude although they provide asymptotic equivalence (in most cases). They do not imply any concentration property. Nevertheless there is "very strong" concentration if $p = p(n)$ is sufficiently small. The following result is due to Łuczak (64) and Alon and Krivelevich (6).

Theorem 2 *Fix some* $\varepsilon > 0$*. For every sequence* $p = p(n)$ *there exists a function* $h(n)$ *such that almost always*

(i) *if* $p \ge n^{-\frac{1}{2}-\varepsilon}$ *then* $\chi(G(n,p)) \sim h(n)$*, and*

(ii) *if* $p \le n^{-\frac{1}{2}-\varepsilon}$ *then* $h(n) \le \chi(G(n,p)) \le h(n) + 1$.

The proof of these kinds of concentration properties relies basically on Azuma's inequality. For example, one can show Theorem 2.(ii) for $p \le n^{-6/7}$ with this method (together with some tricky elementary arguments). Alon and Krivelevich used a further ingredience, the Lóvasz Local Lemma, to sharpen this result to $p \le n^{-\frac{1}{2}-\varepsilon}$.

In what follows we give some hints how one can use Azuma's inequality to problems like the chromatic number of random graphs. We recall that a martingale is a sequence of random variables $Y_0, Y_1 \ldots, Y_n$ on a probabilty space $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{E}\,(Y_{k+1}|\mathcal{F}_k) = Y_k$, where $\mathcal{F}_0 = \{\emptyset, \Omega\} \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n = \mathcal{F}$ is an increasing sequence of σ-fields.

Theorem 3 (Azuma's Inequality) *Suppose that* $Y_0, Y_1 \ldots, Y_n$ *is a martingale with constant* Y_0 *and that*

$$|Y_{k+1} - Y_k| \le c_k \tag{9}$$

for some some constants c_k $(0 \le k < n)$*. Then, for every* $t > 0$,

$$\mathbb{P}\{|Y_n - \mathbb{E}\,Y_n| \ge t\} \le 2\,\exp\left(-\frac{t^2}{2\sum_{k=1}^n c_k^2}\right). \tag{10}$$

A very useful application of Azuma's inequality is the following property that is also called *independent bounded difference inequality* (that is due to McDiarmid (67)).

Theorem 4 *Let* $X_1, \ldots, X_n$ *be independent random variables, with* X_k *taking values in a set* Ω_k*. Suppose that a function* $f : \Omega_1 \times \cdots \times \Omega_n \to \mathbb{R}$ *satisfies the property that*

$$|f(x_1, \ldots, x_n) - f(y_1, \ldots, y_n)| \le c_k \tag{11}$$

if $(x_1, \ldots, x_n)$ *and* $(y_1, \ldots, y_n)$ *differ only at the* k*-th coordinate, that is* $x_j = y_j$ *for* $j \ne k$.

Then, the random variable $Y = f(X_1, \ldots, X_n)$ *satisfies, for any* $t \ge 0$,

$$\mathbb{P}\{|Y - \mathbb{E}\,Y| \ge t\} \le 2\,\exp\left(-\frac{t^2}{2\sum_{k=1}^n c_k^2}\right). \tag{12}$$

Proof: Let $\mathcal{F}_k$ denote the σ-field generated y $X_1, \ldots, X_k$ and consider the corresponding martingale defined by $Y_k = \mathbb{E}\,(f(X_1, \ldots, X_n)|\mathcal{F}_k)$, $k = 0, 1, \ldots, n$. Then it is easy to show that (11) implies (9). Thus, Theorem 3 applies. □

In the context or random graphs one frequently uses the *vertex exposure martingale* that is constructed in the following way. Let $A_k = \{(j,k) : 1 \le j < k\}$ the set of (undirected) edges of the complete graph on vertex set $\{1, 2, \ldots, n\}$ that connect k with a vertex $j < k$ and let $X_k = (\mathbb{I}_{[e \in G(n,p)]} : e \in A_k)$ denote the random vector of indicators of edges in A_k. Now let f be any *graph theoretical function* (for example, the chromatic number). Then $Y_k = \mathbb{E}\left(f(G(n,p))|\mathcal{F}_k\right)$, where $\mathcal{F}_k$ is the σ-field generated by $X_1,\ldots,X_k$ is a martingale, the so-called vertex exposure martingale. It can be interpreted as the conditional expectation of f with partial information on the first k vertices and their internal edges. Here (11) says that $|f(H_1)-f(H_2)|\le c_k$ if H_1,H_2 are subgraphs of the complete graph on the vertices $\{1,2,\ldots,n\}$ such that the symmetric difference of the edge sets of H_1 and H_2 is contained in A_k.

Let us illustrate this property with the chromatic number. Let H_1,H_2 be subgraphs of the complete graph on the vertices $\{1,2,\ldots,n\}$ with the property that the symmetric difference of the edge sets of H_1 and H_2 is contained in A_k. Let H denote the graph that corresponds to the intersetion of the edge sets of H_1 and H_2. Then H and H_1 resp. H and H_2 differ (at most) by the vertex k and some edges that connect k with vertices $j<k$. Since we can color the vertex k always with one additional color we surely have $\chi(H)\le\chi(H_1)\le\chi(H)+1$ and similarly $\chi(H)\le\chi(H_2)\le\chi(H)+1$. Thus, $|\chi(H_1)-\chi(H_2)|\le 1$ and we can use (11) with $c_k=1$. In particular, we immediately obtain the following concentration property that is due to Shamir and Spencer (87).

Theorem 5 *Let n,p arbitrary. Set $c=\mathbb{E}\left(\chi(G(n,p))\right)$ Then*

$$\mathbb{P}\{|\chi(G(n,p))-c|>\lambda\sqrt{n-1}\}<2\,e^{-\lambda^2/2}.$$

Of course, this does not prove the concentration properties of Theorem 2.(ii). However, it provides a first bound that is probably not too far from optimality if p is *large*.

4 Talagrand's Inequality and the Travelling Salesman Problem

Let $\mathbf{X}=(X_1,X_2,\ldots,X_n)$ be an n-tuple of random point selected uniformly and independently in the unit square $[0,1]^2$ and let $\mathrm{TSP}(\mathbf{X})$ the length of the minumum (travelling salesman) tour through these points:

$$\mathrm{TSP}(\mathbf{X})=\min_{\pi\in S_n}\sum_{j=1}^{n}\left|X_{\pi(j)}-X_{\pi(j+1)}\right|.$$

The exact behaviour of $\mathbb{E}\left(\mathrm{TSP}(\mathbf{X})\right)$ or the median $\mathrm{M}\left(\mathrm{TSP}(\mathbf{X})\right)$ is not known. However, by a classical theorem of Beardwood, Halton and Hammersley (9) we have

$$\frac{\mathrm{TSP}(\mathbf{X})}{\sqrt{n}}\to\beta_2$$

in probability, where $\beta_2>0$ is a constant for which no analytic expression is known, compare also with Karp (58) and Steele (84).

However, by an application of Talagrand's inequality one can show "strong" concentration.

Theorem 6 *Let $\mathrm{TSP}(\mathbf{X})$ the length of the minumum tour through n independent random points in the unit square. Then there exists a constant $c>0$ with*

$$\mathbb{P}\left\{|\mathrm{TSP}(\mathbf{X})-\mathrm{M}(\mathrm{TSP}(\mathbf{X}))|\ge t\right\}<4e^{-t^2/c}. \tag{13}$$

This is in fact a very strong result and was first proved by Rhee and Talagrand (88) by very much involved arguments that are based on a martingale approach.

We shortly describe Talagrand's inequality (89) and how it can be applied to the Travelling Salesman Problem without much effort (compare also with (44; 67)).

Let $\Omega_1,\Omega_2,\ldots,\Omega_n$ probability spaces and Ω the product space. Further, let $\mathbf{X}=(X_1,X_2,\ldots,X_n)$ be an n-tuple of independent random variables X_k taking values in Ω_k. For $\mathbf{x}\in\Omega$ and $A\subset\Omega$ let

$$d_T(\mathbf{x},A)=\sup_{\alpha\ge 0,\|\alpha\|=1}\inf_{\mathbf{y}\in A}d_\alpha(\mathbf{x},\mathbf{y})$$

be Talagrand's *convex distance*, where

$$d_\alpha(\mathbf{x},\mathbf{y})=\sum_{x_i\ne y_i}\alpha_i$$

is a weighted Hamming distance related to $\alpha=(\alpha_1,\ldots,\alpha_n)$ with $\alpha_k\ge 0$ $(1\le k\le n)$. Talagrand's inequality can be now stated as

$$\mathbb{P}\{\mathbf{X}\in A\}\cdot\mathbb{P}\{d_T(\mathbf{X},A)\ge t\}<e^{-t^2/4}. \tag{14}$$

For a proof see (67; 89).

The next variant of Talagrand's inequality is quite useful for applications.

Theorem 7 *Let f be a real valued function on $\Omega = \Omega_1 \times \cdots \times \Omega_n$ and suppose that for every $\mathbf{x} \in \Omega$ there exists a non-negative unit n-vector α and a constant $c > 0$ such that for all $\mathbf{y} \in \Omega$*

$$f(\mathbf{x}) \le f(\mathbf{y}) + c\, d_\alpha(\mathbf{x}, \mathbf{y}). \tag{15}$$

Then, for every random n-tuple $\mathbf{X} = (X_1, \ldots, X_n)$ of independent random variables X_k taking values in Ω_k we have

$$\mathbb{P}\{|f(\mathbf{X}) - \mathrm{M}(f(\mathbf{X}))| \ge t\} \le 4\, e^{t^2/(4c^2)}. \tag{16}$$

Proof: For a real number a set $A_a = \{\mathbf{y} \in \Omega : f(\mathbf{y}) \le a\}$. By assumption for every $\mathbf{x} \in \Omega$ there exists a non-negative unit n-vector α such that for all $\mathbf{y} \in A_a$

$$f(\mathbf{x}) \le f(\mathbf{y}) + c\, d_\alpha(\mathbf{x}, \mathbf{y}) \le a + c\, d_\alpha(\mathbf{x}, \mathbf{y}).$$

By taking the miminum over all $\mathbf{y} \in A_a$ we, thus, get

$$f(\mathbf{x}) \le a + c\, d_\alpha(\mathbf{x}, A_a) \le a + c\, d_T(\mathbf{x}, A_a).$$

Hence, if $f(\mathbf{x}) \ge a + t$ then $d_T(\mathbf{x}, A_a) \ge t/c$. Consequently, by Talagrand's inequality

$$\begin{aligned}\mathbb{P}\{f(\mathbf{X}) \le a\} \cdot \mathbb{P}\{f(\mathbf{X}) \ge a + t\} &\le \mathbb{P}\{\mathbf{X} \in A_a\} \cdot \mathbb{P}\{d_T(\mathbf{x}, A_a) \ge t/c\} \\ &\le e^{-t^2/(4c^2)}.\end{aligned}$$

Finally, if we use $a = \mathrm{M}(f(\mathbf{X}))$ then $\mathbb{P}\{f(\mathbf{X}) \le a\} = \frac{1}{2}$ and we get

$$\mathbb{P}\{f(\mathbf{X}) \ge \mathrm{M}(f(\mathbf{X})) + t\} \le 2\, e^{-t^2/(4c^2)}.$$

Similarly, if we use $a = \mathrm{M}(f(\mathbf{X})) - t$ we obtain $\mathbb{P}\{f(\mathbf{X}) \le \mathrm{M}(f(\mathbf{X})) - t\} \le 2\, e^{-t^2/(4c^2)}$. Of course, this proves (16). □

Finally, we show that $f = \mathrm{TSP}$ satisfies (15). Of course, this proves Theorem 6.

Let $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \Omega = ([0,1]^2)^n$. It is an easy exercise that there exists a permutation $\pi \in S_n$ such that

$$\sum_{j=1}^{n} \|x_{\pi(j+1)} - x_{\pi(j)}\|^2 < c$$

for some absolute constant c. For example one just has to divide the unit square $[0,1]^2$ into $\approx n$ subrectangles of diameter $\approx 1/\sqrt{n}$. We denote by $T(\mathbf{x}) = \big(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)}\big)$ the tour corresponding to this permutation. Further, set $\beta_k = \|x_{\pi(k-1)} - x_{\pi(k)}\| + \|x_{\pi(k)} - x_{\pi(k+1)}\|$ and $\beta = (\beta_1, \ldots, \beta_n)$. Then $\sum_{k=1}^n \beta_k^2 \le 4c$. We will show that

$$\mathrm{TSP}(\mathbf{x}) \le \mathrm{TSP}(\mathbf{y}) + d_\beta(\mathbf{x}, \mathbf{y}) \tag{17}$$

holds for all $\mathbf{y} \in \Omega$. Obviously, if we set $\alpha = \beta/\|\beta\|$ then (17) implies

$$\mathrm{TSP}(\mathbf{x}) \le \mathrm{TSP}(\mathbf{y}) + 2\sqrt{c}\, d_\alpha(\mathbf{x}, \mathbf{y}),$$

that is, we have proved (15) for $f = \mathrm{TSP}$ and we are done.

For $\mathbf{x} = (x_1, \ldots, x_n), \mathbf{y} = (y_1, \ldots, y_n) \in \Omega = ([0,1]^2)^n$ we set $\tilde{\mathbf{x}} = \{x_1, \ldots, x_n\}$ and $\tilde{\mathbf{y}} = \{y_1, \ldots, y_n\} \subseteq [0,1]^2$. If $\tilde{\mathbf{x}} \cap \tilde{\mathbf{y}} = \emptyset$ then $d_\beta(\mathbf{x}, \mathbf{y}) = 2\,\ell(T(\mathbf{x}))$, where $\ell(T(\mathbf{x}))$ denotes the length of the tour $T(\mathbf{x})$). Of course, in this case (17) holds trivially. Similarly, if $|\tilde{\mathbf{x}} \cap \tilde{\mathbf{y}}| = 1$ then $d_\beta(\mathbf{x}, \mathbf{y}) \ge \ell(T(\mathbf{x})$ and, thus, (17) holds.

Now suppose that $|\tilde{\mathbf{x}} \cap \tilde{\mathbf{y}}| \ge 2$. The tour $T(\mathbf{x})$ is now divided in *segments* $S = (a, v_1, \ldots, v_j, b)$ with $a, b \in \tilde{\mathbf{x}} \cap \tilde{\mathbf{y}}$ and $v_1, \ldots, v_j \in \tilde{\mathbf{x}} \setminus \tilde{\mathbf{y}}$. We build up a multiset F of edges. The edges (v_i, v_{i+1}) are put into F twice for $1 \le i \le j-1$ and further the shorter of the edges (a, v_1), (v_j, b) is put into F twice, too. This is done for all segments S of $T(\mathbf{x})$ of this form. Note that by this construction the sum of all lenghts of F (according to their multiplicity) is bounded from above by $d_\beta(\mathbf{x}, \mathbf{y})$. Next consider an optimal tour $T^*(\mathbf{y})$ of $\mathbf{y}$ and consider the (multi)graph G with vertex set $\tilde{\mathbf{x}} \cup \tilde{\mathbf{y}}$ and (multi)edge set that consists of the edges of $T^*(\mathbf{y})$ and of F. Then G is connected and Eulerian since every vertex has even degree. The total weight of an Eulerian tour is bounded above by $\mathrm{TSP}(\mathbf{y}) + d_\beta(\mathbf{x}, \mathbf{y})$. Obviously the length of an optimal tour $T^*(\mathbf{x})$ of $\mathbf{x}$ is bounded above by this value. This completes the proof of (17).

This method can be applied to several other problem, for example to the minimal Steiner tree problem etc., see (67; 89).

5 The Longest Increasing Subsequence in Random Permutations

Let S_n denote the set of permutations of the numbers $\{1, 2, \ldots, n\}$, where we assume that every permutation in S_n is equally likely. If $\pi \in S_n$ we say that $\pi(i_1), \pi(i_2), \ldots, \pi(i_k)$ is an increasing subsequence in π if $i_1 < i_2 < \cdots < i_k$ and $\pi(i_1) < \pi(i_2) < \cdots < \pi(i_k)$.

We denote by $L_n = L_n(\pi)$ the length of the longest increasing subsequence and the question is to determine the limiting behaviour of L_n. This question goes back to Ulam (90) in the early 60's who conjectured that $\mathbb{E}\, L_n \sim c\sqrt{N}$ for some constant $c > 0$. Erdős and Szekeres (45) could show $c \geq \frac{1}{2}$, later Logan and Shepp (62) proved $c \geq 2$ and almost simultanously Vershik and Kerov (91) settled Ulam's problem and proved $c = 2$. Alternate proofs are due to Aldous and Diaconis (4), Seppäläinen (86) and Johansson (57). Frieze (51) was then the first who proved that L_n is concentrated with help of martingale methods. His result was improved by Bollobás and Brightwell (18) and by Talagrand (89) who showed that the variance is bounded by $\mathbb{V}L_n = O(\sqrt{N})$. However, it turned out that this order of magnitude is not optimal. Calculations of Odlyzko and Rains (73) indicated that the exact order should be $N^{1/3}$ which is in fact true.

Eventually Baik, Deift, and Johansson (11) settled the problem completely. They proved the following weak concentration property.

Theorem 8 *Let S_n be the group of permutations of n numbers with uniform distribution and L_n the longest increasing subsequence. Then there exists a random variable Y such that*

$$\frac{L_n - 2\sqrt{n}}{n^{1/6}} \xrightarrow{\mathrm{d}} Y. \tag{18}$$

Furthermore, we have convergence of all moments.

It is remarkable that the limiting distribution Y is exactly the same at the limiting distribution of the largest eigenvalue in random Hermitian matrices (see (69)). However, it seems that there is no direct connection between these two problems.

The distribution function $F(t) = \mathbb{P}\{Y \leq t\}$ of Y that is also called *Tracy-Widom distribution* can be characterized by

$$F(t) = \exp\left(\int_t^\infty (x-t)^2 u(x)^2\, dx\right),$$

where $u(x)$ is the solution of the Painlevé II equation

$$u'' = 2u^3 + xu, \quad u(x) \sim Ai(x) \quad (\text{as } x \to \infty);$$

$Ai(x)$ denotes the Airy function.

The proof is an *analytic tour de force*. Basically one determines the asymptotic behaviour of the Poisson transform

$$\phi_k(\lambda) = \sum_{n=0}^\infty \frac{e^{-\lambda}\lambda^n}{n!}\mathbb{P}\{L_n \leq k\}$$

that can be represented as

$$\phi_k(\lambda) = \frac{e^{-\lambda}}{(2\pi)^k k!}\int_{[-\pi,\pi]^k} \exp\left(2\sqrt{\lambda}\sum_{j=1}^k \theta_j\right) \prod_{1\leq j<\ell\leq k} \left|e^{i\theta_j} - e^{i\theta_\ell}\right| d\theta_1 \cdots d\theta_k.$$

One has to use the theory of orthogonal polynomials on the unit circle, sophisticated Riemann-Hilbert problem techniques and certain properties on eigenvalues of random matrices.

There are also studies of the longest increasing subsequence of random permuations under certain restrictions, for example see (30) for permutations that avoid certain patterns of size 3. However, the resuls in these cases are completely different and also depend on the specific pattern that is avoided.

6 Diameter and Maximum Degree in Random Graphs

We will discuss two different kinds of random graphs, first, the usual $G(n, p)$ model and, second, so-called *scale-free random graphs* that model *real world graphs* in the sense of Barabási and Albert (8).

The *diameter* $\mathrm{diam}(G)$ of a graph connected G is the largest distance between two nodes in G. If G is not connected then $\mathrm{diam}(G) = \infty$.

The *maximum degree* of an (undirected) graph G will be denoted by $\Delta(G)$.

6.1 $G(n,p)$-Random Graphs

There are serveral results concerning the diameter of $G(n,p)$. The following theorem collects some of Burtin (23; 24) and Bollobás (13).

Theorem 9

(i) If $(pn)/\log n \to \infty$ *and* $\log n/\log(pn) \to \infty$ *then almost always*

$$\mathrm{diam}(G(n,p)) \sim \frac{\log n}{\log(pn)}.$$

(ii) Let c *be a positive constant and* $p = p(n)$ *and* $d = d(n)$*, an integer* ≥ 2*, be related by* $p^d n^{d-1} = \log(n^2/c)$*. Further suppose that* $(pn)/(\log n)^3 \to \infty$*. Then*

$$\lim_{n\to\infty} \mathbb{P}\{\mathrm{diam}(G(n,p)) = d\} = e^{-c/2}$$

and

$$\lim_{n\to\infty} \mathbb{P}\{\mathrm{diam}(G(n,p)) = d+1\} = 1 - e^{-c/2}.$$

The second part of the result says that the distribution of the diameter is highly concentrated. (Note that $G(n,p)$ is almost always connected if $np = \log n + \omega_n$ for any sequence $\omega_n \to \infty$.)

For example, if $np \sim (\log n)^\ell$ for some $\ell \geq 3$ then the diameter is of order $(\log n)/(\ell \log\log n)$. However, if $p \sim n^{-\delta}$ for some $\delta < 1$ then the diameter is bounded. In particular, if $p^2 n - 2\log n \to \infty$ and $n^2(1-p) \to \infty$ then $\mathrm{diam}(G(n,p)) = 2$ (almost always).

The distribution of the maximum degree will depend on the behaviour of

$$\lambda_k = \lambda_k(n) := n\binom{n-1}{k} p(n)^k (1-p(n))^{n-k-1}.$$

We first state a result on "very strong" concentration properties of the maximal degree Δ that is due to Bollobás (15).

Theorem 10 *Suppose that* $p = p(n) = o(\log n/n)$ *and let* $k = k(n) \geq np$ *be such that the quantity* $\max\{\lambda_k, 1/\lambda_k\}$ *is minimal.*

(i) If $0 < \liminf \lambda_k \leq \limsup \lambda_k < \infty$ *then, as* $n \to \infty$*,*

$$\mathbb{P}\{\Delta = k(n)\} = 1 - e^{-\lambda_k} + o(1)$$

and

$$\mathbb{P}\{\Delta = k(n)\} = e^{-\lambda_k} + o(1).$$

(ii) If $\lim \lambda_k = \infty$ *then*

$$\mathbb{P}\{\Delta = k(n)\} = 1 + o(1).$$

(iii) If $\lim \lambda_k = 0$ *then*

$$\mathbb{P}\{\Delta = k(n) - 1\} = 1 + o(1).$$

(iv) If there is a function $D(n)$ *with* $\mathbb{P}\{\Delta = D(n)\} = 1 + o(1)$ *as* $n \to \infty$ *then* $p = p(n) = o(\log n/n)$*.*

The situation is completely different if p is constant. There is no strong concentration (see (12; 13; 82)).

Theorem 11 *Suppose that* $0 < p < 1$ *is fixed and* $q = 1 - p$*.*

(i) For every real number y *we have*

$$\mathbb{P}\left\{\Delta \leq pn + \sqrt{2pqn\log n}\left(1 - \frac{\log\log n}{4\log n} + \frac{y - 2\sqrt{\pi}}{2\log n}\right)\right\}$$
$$= \exp\left(-e^{-y}\right) + o(1).$$

(ii) Almost always we have

$$\left|\Delta - pn - \sqrt{2pqn\log n} + \log\log n\sqrt{\frac{pqn}{8\log n}}\right| \le \log\log n\sqrt{\frac{n}{\log n}}.$$

(iii) For every real number b there exists $c(b)$ such that

$$\mathbb{P}\{\Delta < pn + b\sqrt{npq}\} = (c(b)+o(1))^n.$$

There are several extension of this result concerning the m-th largest degree, the difference between the largest and the smallest degree etc., see also the paper of Ivchenko (55).

If $np \to 0$ then $G(n,p)$ is definitely disconnected. However, one can study the largest diameter of all components. This has been worked out by Łuczak (63).

6.2 Scale-Free Random Graphs

There has been substantial interest in random graph models where vertices are added to the graph succesively and are connected to several already existing nodes according to some given law. The so-called Barabási-Albert model (8) joins a new node to an existing one with probability proportionally to the degree. The idea behind is to model various *real-word graphs* like the internet or social networks.

It turns out that the Barabási-Albert model is not unambigously defined. Therefore Bollobás and Riordan (20) introduced a precise version for a random (multi)graph G_m^n. In order to make the presentation simpler we start with the description of G_1^n. One starts with an initial node 1 with a loop. (This means that 1 has degree 2). Then at step k we add one node that is connected to $j \le k$ with propability

$$\begin{array}{ll} \dfrac{\deg_{G_1^{k-1}}(j)}{2k-1} & \text{if } j < k, \\ \dfrac{1}{2k-1} & \text{if } j = k. \end{array}$$

Of course, after n steps we have produced a random (multi)graph with vertex set $\{1,2,\ldots,n\}$ and n edges. This graph is closely related to plane oriented recursive trees that will discuss these (and related kinds of trees) in Section 7.

Now the random graph G_m^n is constructed from G_1^{mn} by identifying the nodes $\{(\ell-1)m+1, (\ell-1)m+2,\ldots,\ell m\}$ $(1 \le \ell \le n)$ of G_1^{mn} to a new node ℓ (and all edged within the nodes $\{(\ell-1)m+1, (\ell-1)m+2,\ldots,\ell m\}$ are now loops of the new node ℓ). Of course, this procedure results in a random (multi)graph with vertex set $\{1,2,\ldots,n\}$ and mn edges. (It is also possible to construct G_m^n recursively from G_m^{n-1} by adding a new node and exactly m edges according to the degree distribution of G_m^{n-1} but this is a little bit more difficult to state.)

It turns out that the degree distribution of G_m^n satisfies a *power law*. The probability that a randomly chosen node of G_m^n has degree d is asymptotically $2/d^3$ (see (21)). Graphs with this property are called *scale-free*.

Bollobás and Riordan (20) proved the following result for the diameter of G_m^n.

Theorem 12 *Suppose that $m \ge 2$. Then for every $\varepsilon > 0$*

$$\lim_{n\to\infty} \mathbb{P}\left\{(1-\varepsilon)\frac{\log n}{\log\log n} \le \operatorname{diam}(G_m^n) \le (1+\varepsilon)\frac{\log n}{\log\log n}\right\} = 1.$$

For $m = 1$ the result is different, the diameter is of order $\log n$ (see Pittel (79)).

As already noted there are very precise results on the degree distribution of G_m^n but there are no results on the maximum degree. (Only in the case $m = 1$ we can adopt the results on trees that we will discuss in the next section.)

7 Height and Maximum Degree in Random Trees

Properties of trees are important for many algorithms. Further they are used as data structures in computer science in various ways so that there are several *natural probabilistic models*. Therefore we put a special emphasis on trees. Here we focus on two extremal statistics, on the height of trees and on the the maximum degree.

7.1 Galton-Watson Trees

Let ξ be a non-negative integer valued random variable with $\mathbb{E}\,\xi = 1$, $0 < \mathbb{V}\,\xi = \sigma^2 < \infty$.(iii) The Galton-Watson branching process $(Z_k)_{k\ge 0}$ is now given by $Z_0 = 1$, and for $k \ge 1$ by

$$Z_k = \sum_{j=1}^{Z_{k-1}} \xi_j^{(k)},$$

where the $(\xi_j^{(k)})_{k,j}$ are iid random variables distributed as ξ.

It is well-known that Galton-Watson branching processes can be represented by ordered (finite or infinite) rooted trees T such that the sequence $(Z_k)_{k\ge 0}$ is just the profile sequence and $\sum_{k\ge 0} Z_k$, which is called the *total progeny*, is just the number of nodes $|T|$ of T.(iv) We will denote $\nu(T)$ the probability that T occurs. The generating function $y(x) = \sum_{n\ge 1} y_n x^n$ of the numbers

$$y_n = \mathbf{Pr}[|T| = n] = \sum_{|T|=n} \nu(T)$$

satisfies the functional equation

$$y(x) = x\varphi(y(x)),$$

where $\varphi(t) = \mathbb{E}\,t^\xi = \sum \varphi_i t^i$ with $\varphi_i = \mathbf{Pr}\{\xi = i\}$. If $\mathcal{T}_n$ denotes the set of rooted trees T of size $|T| = n$ then

$$\nu_n(T) := \frac{\nu(T)}{y_n}$$

is a probability distribution on $\mathcal{T}_n$ which we will use in the sequel. The *random tree* in $\mathcal{T}_n$ will be denoted by T_n. Note that

$$y_n \sim \frac{d}{\sqrt{2\pi}\sigma} n^{-3/2} \qquad (n \equiv 1 \bmod d), \tag{19}$$

where $d = \gcd\{i > 0 : \varphi_i > 0\}$.(v) Meir and Moon (70) have considered the same kind of trees under the notion of *simply generated trees*. In fact they introduced general non-negative weights φ_i. However, by a proper scaling simply generated trees can be reduced to (critical) Galton-Watson trees.

For example, for $\varphi(t) = \mathbb{E}\,t^\xi = (1+t)^2/4 = \frac{1}{4} + \frac{t}{2} + \frac{t^2}{4}$ we just recover the class of binary trees with n (internal) nodes, where each binary tree (of size n) has equal probability. Another well-know example is that of planted plane trees (again with uniform distribution on trees of size n). They are induced by $\varphi(t) = \mathbb{E}\,t^\xi = 1/(2-t) = \frac{1}{2} + \frac{t}{4} + \frac{t^2}{8} + \cdots$. Further, $\varphi(t) = e^{t-1}$ models rooted labeled trees, that is, all n^{n-2} rooted labeled trees are equally likely.

The first contribution to the height of special Galton-Watson trees is due to de Bruijn, Knuth and Rice (29) who considered the case of planted plane trees ($\varphi(t) = 1/(2-t)$). Later Flajolet and Odlyzko (48) and Flajolet, Gao, Odlyzko and Richmond (47) determined the distribution of the height for a very general class of simply generated trees, that is, in terms of Galton-Watson trees, where $\mathbb{E}\,t^\xi$ exists for some $t > 1$. Later, Aldous (1; 2; 3) introduced the notion of a *continuum random tree* that is (in a proper sense) the weak limit of scaled Galton-Watson trees. Since the height is a continuous functional (in this context) one directly gets a weak limit theorem for the height under the *weak assumption* that the second moment of the offspring distribution exists.

Theorem 13 *Suppose that the second moment $\mathbb{E}\,\xi^2$ is finite. Then*

$$\frac{1}{\sqrt{n}} H_n \xrightarrow{\mathrm{d}} \frac{2}{\sigma} \max_{0\le t\le 1} e(t),$$

where $(e(t), 0 \le t \le 1)$ denotes Brownian excursion of duration 1. Furthermore, if $\varphi(t) = \mathbb{E}\,t^\xi$ exists for some $t > 1$ then we also have convergence of all moments. For every $r \ge 0$ we have, as $n \to \infty$,

$$\mathbb{E}\,(H_n^r) = 2^{r/2}\sigma^{-r} r(r-1)\Gamma(r/2)\zeta(r) \cdot n^{r/2} \left(1 + O(n^{-\frac{1}{4}+\eta})\right)$$

where $\zeta(s)$ denotes the Riemann Zeta-function and $(r-1)\zeta(r) = 1$ for $r = 1$ and η is any positive number.

(iii) Usually it is not assumed that $\mathbb{E}\,\xi = 1$ (which characterize so-called *critical branching processes*). However, for our purposes it is no loss of generality to make this assumption (see (59)).

(iv) For critical branching processes the probability that the total progeny is finite equals 1.

(v) In what follows we will always assume that $d = 1$. The case $d > 1$ is completely analogous.

In particular, the height is not concentrated. This is quite unusual for extremal parameters. However, this phenomenon is easy to explain with help of the continuum random tree approximation.

Note that the distribution function of $M := \max_{0\le t\le 1} e(t)$ is explicitly given by

$$\mathbb{P}\{M \le x\} = 1 - 2\sum_{k=1}^{\infty}(4x^2k^2-1)e^{-2x^2k^2}$$

and the moments by

$$\mathbb{E}\,(M^r) = 2^{-r/2}r(r-1)\Gamma(r/2)\zeta(r).$$

The approach of Aldous is quite general, but it does not give an error term. The only known method that provides an error term is due to Flajolet and Odlyzko (48) and is based on generating function. We quickly sketch their approach.

Let $y_k(x)$ denote the generating functions

$$y_k(x) = \sum_{n\ge 1}\mathbb{P}\{|T| = n,\, h_T \le k\}\, x^n,$$

where h_T denotes the height of T, then $y_0(x) = \varphi_0 x$, and recursively

$$y_{k+1}(x) = x\varphi(y_k(x)), \qquad (k \ge 0). \tag{20}$$

With help of these function one gets the generating function of the expected height:

$$H(x) := \sum_{n\ge 1}\mathbb{E}\; H_n \cdot y_n \cdot x^n = \sum_{k\ge 0}(y(x) - y_k(x)).$$

After a subtle analysis of the above recurrence (20) it is possible to derive a local representation of the form

$$H(x) = \frac{1}{\sigma^2}\log\frac{1}{1-x} + K + O\left(|1-x|^{\frac14-\eta}\right)$$

for some constant K and every (fixed) $\eta > 0$ (see (48)). Thus, with help of the Transfer Lemma (see (49)) and (19) we directly get, as $n\to\infty$,

$$\mathbb{E}\; H_n = \frac{\sqrt{2\pi}}{\sigma}\cdot\sqrt{n} + O\left(n^{\frac14+\eta}\right).$$

In a similar way one gets corresponding asymptotic equivalents for higher moments which characterize again the distribution of $n^{-1/2}H_n$ of Theorem 13.

In rooted trees one usually considers the out-degree of a node, which is the same as the degree at the root and differs by 1 from the degree for all nodes that are different from the root. Further, we only have to discuss cases where $\varphi(t)$ is not a polynomial. Namely, if $\varphi(t)$ is a polynomial of degree d then the maximum out-degree equals d for almost all trees in $\mathcal{T}_n$.

The maximum out-degree will be again denoted by $\Delta(T_n)$. The following theorem collects some results of Meir and Moon (71) and of Carr, Goh and Schmutz (25).

Theorem 14

(i) Suppose that $\varphi_i = \Pr\{\xi = i\} > 0$ *for sufficiently large* $i \ge i_0$ *and that* $\varphi_{i+1}/\varphi_i \to 0$ *as* $i\to\infty$. *Then*

$$\mathbb{P}\{|\Delta(T_n) - \delta(n)| \le 1\} = 1 + o(1),$$

where $\delta(n) = \max\{k\ge 0 : \mathbb{P}\{\xi \ge k\} \ge 1/n\}$.

(ii) If $\varphi(t) = e^{t-1}$ *then there exists a sequence* $\delta'(n)$ *that is asymptotically equivalent to* $\delta'(n) \sim \frac{\log n}{\log\log n}$ *such that*

$$\mathbb{P}\{\delta'(n) \le \Delta(T_n) \le \delta'(n) + 1\} = 1 + o(1).$$

(iii) If $\varphi(t) = 1/(2-t)$ *then we have uniformly for all* $k \ge 0$

$$\mathbb{P}\{\Delta(T_n) \le k\} = \exp\left(-2^{-(k-\log_2 n+1)}\right) + o(1)$$

The proofs of these results rely (again) on generating functions. Set

$$y_d(x) = \sum_{n\ge 1} \mathbb{P}\{|T| = n,\, \Delta(T_n) \le d\}\, x^n.$$

Then this generating function satisfies the functional equation

$$y_d(x) = x\,\varphi_d(y_d(x)) \quad \text{with} \quad \varphi_d(t) = \sum_{i\le d} \varphi_i t^i.$$

Thus, we are in a similar situation as in the original problem. For every fixed d it follows that

$$\mathbf{Pr}\{|T| = n,\, \Delta(T_n) \le d\} \sim C_d(\varphi_d'(\tau_d))^n n^{-3/2},$$

where $\tau_d > 0$ is determined by the equation $\tau_d\varphi_d'(\tau_d) = \varphi_d(\tau_d)$ and C_d is a certain constant. This implies that

$$\mathbb{P}\{\Delta(T_n) \le d\} = \frac{\mathbf{Pr}\{|T| = n,\, \Delta(T_n) \le d\}}{\mathbf{Pr}\{|T| = n\}} \sim \sqrt{2\pi}\sigma C_d\, (\varphi_d'(\tau_d))^n.$$

The only problem is to make this asymptotic relation uniform for d in the interesting range (which is non-trivial) and to determine the (asymptotic) behaviour of $\tau_d - 1$ (which is usually easy).

Another interesting (extremal) parameter of trees is the *width*, that is, the maximal number of nodes in the same level. In (40) it was shown that (up to scaling) the width has the same limiting distribution as the height. Furthermore one also has convergence of moments, see (41). Again there is no concentration.

7.2 Pólya Trees

We now consider rooted unlabeled (non-planar) trees and let t_n denote the number of trees of this type of size n. It has been first observed by Pólya (80) — this explains the notion *Pólya Trees* – that the generating function

$$t(x) = \sum_{n\ge 1} t_n x^n.$$

satisfies the functional equation

$$t(x) = x \exp\left(t(x) + \frac{1}{2}t(x^2) + \frac{1}{3}t(x^3) + \cdots \right). \tag{21}$$

Later, Otter (74) determined the asymptotic structure and also that the generating function $\tilde{t}(x)$ of unrooted unlabeld trees is given by $\tilde{t}(x) = t(x) - \frac{1}{2}\left(t(x)^2 - t(x^2)\right)$.

The radius of convergence of $t(x)$ (and of $\tilde{t}(x)$) is approximately $\rho \approx 0.338219$ and is given by $t(\rho) = 1$, that is, $t(x)$ is convergent at $x = \rho$. Further, $t(x)$ has a local expansion of the form

$$t(x) = 1 - b(\rho - x)^{1/2} + c(\rho - x) + O(|\rho - x)^{3/2}), \tag{22}$$

where $b \approx 2.6811266$ and $c = b^2/3 \approx 2.3961466$, which implies via asymptotic transfer that

$$t_n = \frac{b\sqrt{\rho}}{2\sqrt{\pi}} n^{-3/2}\rho^{-n}\left(1 + O(n^{-1})\right). \tag{23}$$

The essential observation is that $\rho < 1$. Hence the part $\frac{1}{2}t(x^2) + \frac{1}{3}t(x^3) + \cdots$ in the functional equation (21) behaves *nicely* if x is close to the singulartity. Hence, (21) has a similiar structure as the functional equation $y(x) = x\,\varphi(y(x))$ for Galton-Watson trees. It is therefore not unexpected that the height and maximum degree behave similar to (proper) Galton-Watson trees if we assume that Pólya trees of size n are considered to be equally likely.

First the height H_n satisfies the same properties as the height of Galton-Watson trees, Theorem 13 applies for Pólya trees, too. This is shown in a forthcoming paper (42).

The distribution of the maximum degree was determined by Goh and Schmutz (52).

Theorem 15 *Let $\Delta(T_n)$ denote the maximum out-degree of Pólya trees of size n. Then*

$$\mathbb{P}\{\Delta(T_n) \le k\} = \exp\left(-c_0\eta^{k-\mu_n}\right) + o(1)$$

with $c_0 = 3.262\ldots$, $\eta = 0.3383\ldots$, and $\mu_n = 0.9227\ldots \cdot \log n$.

This means that the maximum degree of Pólya trees behaves quite similiar to the maximum degree of planted plane trees, however, with a different scaling. Nevertheless, in both cases we have "strong" concentration and the limiting behaviour is described with help of the extreme value (or Gumbel) distribution.

The proof runs (again) along the lines that have been indicated for Galton-Watson trees. However, it is much more involved. For example, one has to make use of asymptotic properties of the cycle index of the symmetric group.

7.3 m-Ary Search Trees

Fringe balanced m-ary search trees are characterized by two integer parameters $m \ge 2$ and $t \ge 0$. The search tree is built from a set of n distinct keys taken from some totally ordered set such as the real numbers or integers; for our purposes we can assume that the keys are the integers $1, \ldots, n$. The search tree will be an m-ary tree where each node has at most m children; moreover, each node will store one or several of the keys, up to at most $m-1$ keys in each node. The parameter t affects the probability distribution of the trees; higher values of t tend to make the tree more balanced. We remark that the simplest, and most often studied, case is the random *binary search tree* obtained by taking $m = 2$ and $t = 0$.

To describe the construction of the search tree, we begin with the simplest case $t = 0$. If $n = 0$, the tree is empty. If $1 \le n \le m-1$, the tree consists of a root only, with all keys stored in the root. If $n \ge m$, we randomly select $m-1$ keys that are called *pivots* (with the uniform distribution over all sets of $m-1$ keys). The pivots are stored in the root. The $m-1$ pivots split the set of the remaining $n-m+1$ keys into m subsets $I_1, \ldots, I_m$: if the pivots are $x_1 < x_2 < \ldots x_{m-1}$, then $I_1 := \{x_i : x_i < x_1\}$, $I_2 := \{x_i : x_1 < x_i < x_2\}$, …, $I_m := \{x_i : x_{m-1} < x_i\}$. We then construct recursively a search tree for each of the sets I_i of keys (ignoring it if I_i is empty), and attach the roots of these trees as children of the root in the search tree. As already mentioned, in the case $m = 2$, $t = 0$, we thus have the well-studied *binary search tree*, see (66).

In the case $t \ge 1$, the only difference is that the pivots are selected in a different way, which affects the probability distribution of the set of pivots and thus of the trees. We now select $mt + m - 1$ keys at random, order them as $y_1 < \cdots < y_{mt+m-1}$, and let the pivots be $y_{t+1}, y_{2(t+1)}, \ldots, y_{(m-1)(t+1)}$. In the case $m \le n < mt+m-1$, when this procedure is impossible, we select the pivots by some supplementary rule (possibly random, but depending only on the order properties of the keys). This splitting procedure was first introduced by Hennequin for the study of variants of the Quicksort algorithm and is referred to as the generalized Hennequin Quicksort, see (28).

We describe the splitting of the keys by the random vector $\mathbf{V}_n = (V_{n,1}, V_{n,2}, \ldots, V_{n,m})$, where $V_{n,k} := |I_k|$ is the number of keys in the kth subset, and thus the number of nodes in the kth subtree of the root (including empty subtrees). We thus always have, provided $n \ge m$,

$$V_{n,1} + V_{n,2} + \cdots + V_{n,m} = n - (m-1) = n + 1 - m$$

and elementary combinatorics, counting the number of possible choices of the $mt + m - 1$ selected keys, shows that the probability distribution is, for $n \ge mt + m - 1$ and $n_1 + n_2 + \cdots + n_m = n - m + 1$,

$$\mathbb{P}\{\mathbf{V}_n = (n_1, \ldots, n_m)\} = \frac{\binom{n_1}{t} \cdots \binom{n_m}{t}}{\binom{n}{mt+m-1}}. \tag{24}$$

For this random model of trees let $H_n^{(m,t)}$ denote the height and $\overline{H}_n^{(m,t)}$ the saturation level, that is, the maximal level up to which the tree is a complete m-ary tree. The above description of m-ary search trees directly leads to an explicit recurrence of the form

$$\begin{aligned} &\mathbb{P}\{H_n^{(m,t)} \le k+1\} \\ &= \sum_{n_1+n_2+\cdots+n_m = n-m+1} \frac{\binom{n_1}{t}\binom{n_2}{t} \cdots \binom{n_m}{t}}{\binom{n}{m(t+1)-1}} \mathbb{P}\{H_{n_1}^{(m,t)} \le k\} \cdots \mathbb{P}\{H_{n_m}^{(m,t)} \le k\}. \end{aligned} \tag{25}$$

The reason is that a tree of size n has height $\le k+1$ if and only if all subtrees of the root (of sizes $n_1, \ldots, n_m$) have heights $\le k$.

The height $H_n = H_n^{(2,0)}$ of binary search trees (and its variants) has a long history (compare with (37)). In 1986 Devroye (32) proved that the expected value $\mathbf{E}\, H_n$ satisfies the asymptotic relation $\mathbf{E}\, H_n \sim c \log n$ (as $n \to \infty$), where $c = 4.31107\ldots$ is the largest real solution of the equation $\left(\frac{2e}{c}\right)^c = e$. (Earlier Pittel (76) had shown that $H_n / \log n \to \gamma$ almost surely as $n \to \infty$, where $\gamma \le c$, compare also with Robson

(83).) Based on numerical data Robson conjectured that the variance $\mathbf{Var}\, H_n$ is bounded. Eventually, Reed (81) and independently Drmota (37) settled Robson's conjecture and proved that

$$\mathbb{V}\, H_n = O(1).$$

In (37) the distribution of H_n was also asymptotically determined.

In what follows we present a result from (26) that generalizes this result to the height (and saturation level) of fringe balanced m-ary search trees. Let $\beta_1 > 0$ and $\beta_2 < 0$ be the two solutions of the equation

$$\sum_{j=0}^{(m-1)(t+1)-1} \log(\beta+t+1+j) - \log\left(\frac{(m(t+1))!}{(t+1)!}\right) = \sum_{j=0}^{(m-1)(t+1)-1} \frac{\beta}{\beta+t+1+j} \tag{26}$$

and set

$$\rho_1 = \exp\left(\sum_{j=0}^{(m-1)(t+1)-1} \frac{1}{\beta_1+t+1+j}\right) \quad \text{and} \quad \rho_2 = \exp\left(\sum_{j=0}^{(m-1)(t+1)-1} \frac{1}{\beta_2+t+1+j}\right).$$

Next consider the functional equation

$$F(x/\rho) = \mathbb{E}\,\left(F(xV_1)\cdots F(xV_m)\right), \tag{27}$$

where the random vector $\mathbf{V} = (V_1, V_2 \ldots, V_m)$ is supported on the simplex $\Delta = \{(s_1,\ldots,s_m) : s_j \geq 0, s_1+\cdots+s_m = 1\}$ with density

$$f(s_1,\ldots,s_m) = \frac{((t+1)m-1)!}{(t!)^m}\,(s_1\cdots s_m)^t.$$

In (26) it is shown that (27) has (up to scaling) a unique solution $F^{(m,t)}(x)$ for $\rho = \rho_1$ and a solution $G^{(m,t)}(x)$ for $\rho = \rho_2$ with the properties

$$1 - F^{(m,t)}(x) \sim d_1 x^{\beta_1} \log x \quad (x \to 0+) \tag{28}$$

and

$$1 - G^{(m,t)}(x) \sim d_2 x^{\beta_2} \log x \quad (x \to \infty) \tag{29}$$

for non-zero real constants d_1, d_2. Furthermore, $F^{(m,t)}(x)$ and $G^{(m,t)}(x)$ are strictly decreasing resp. increasing, continuous, and satisfy

$$\lim_{x\to\infty} F^{(m,t)}(x) = \lim_{x\to 0+} G^{(m,t)}(x) = 0. \tag{30}$$

Theorem 16 *Let $m \geq 2$ and $t \geq 0$ be integers. There exist sequences c_k, d_k with*

$$\lim_{k\to\infty} \frac{c_{k+1}}{c_k} = \rho_1 \quad \textit{and} \quad \lim_{k\to\infty} \frac{d_{k+1}}{d_k} = \rho_2$$

such that

$$\mathbb{P}\{H_n^{(m,t)} \leq k\} = F^{(m,t)}(n/c_k) + o(1) \quad \textit{and} \quad \mathbb{P}\{\overline{H}_n^{(m,t)} > k\} = G^{(m,t)}(n/d_k) + o(1).$$

Furthermore, set $k_1(n) = \max\{k \geq 0 : c_k \leq n\} \sim \log n/\log\rho_1$ and $k_2(n) = \max\{k \geq 0 : d_k \leq n\} \sim \log n/\log\rho_2$. Then

$$\mathbb{E}\, H_n^{(m,t)} = k_1(n) + O(1) \quad \textit{and} \quad \mathbb{E}\,\overline{H}_n^{(m,t)} = k_2(n) + O(1)$$

and there exists $\eta > 0$ with

$$\mathbb{P}\{|H_n^{(m,t)} - \mathbb{E}\, H_n^{(m,t)}| \geq y\} = O(e^{-\eta y}) \quad \textit{and} \quad \mathbb{P}\{|\overline{H}_n^{(m,t)} - \mathbb{E}\,\overline{H}_n^{(m,t)}| \geq y\} = O(e^{-\eta y}).$$

In particular we have, as $n \to \infty$,

$$\mathbb{V} H_n^{(m,t)} = O(1) \quad \textit{and} \quad \mathbb{V}\overline{H}_n^{(m,t)} = O(1).$$

This result shows that the height and saturation level of m-ary search trees are strongly concentrated.

The proof is based on an analysis of generating functions. Set

$$y_k(x) = \sum_{n\ge 0} \mathbb{P}\{H_n^{(m,t)} \le k\} \cdot x^n \quad \text{and} \quad \overline{y}_k(x) = \sum_{n\ge 0} \mathbb{P}\{\overline{H}_n^{(m,t)} \ge k\} \cdot x^n.$$

Then (25) is restated as

$$y_{k+1}^{(m(t+1)-1)}(x) = \frac{(m(t+1)-1)!}{(t!)^m} \left(y_k^{(t)}(x)\right)^m, \tag{31}$$

$$\overline{y}_{k+1}^{(m(t+1)-1)}(x) = \frac{(m(t+1)-1)!}{(t!)^m} \left(\overline{y}_k^{(t)}(x)\right)^m \tag{32}$$

with initial conditions

$$y_0(x) = 1,\ y_k(0) = y_k'(0) = \cdots = y_k^{(m-1)}(0) = 1$$

and

$$\overline{y}_0(x) = \frac{x}{1-x},\ y_k(0) = y_k'(0) = \cdots = y_k^{(m-1)}(0) = 0.$$

In Section 8 we will present a proof of strong concentration for the height of scale free trees that is very similar to the proof of concentration in the present case.

As for Galton-Watson trees one is also interested in the width W_n of m-ary search trees. However, there are only few resulst for binary search trees ($m = 2$, $t = 0$). It is known (27) that

$$\frac{W_n}{n/\sqrt{4\pi \log n}} \to 1 \quad a.s.$$

Further, a similar relation holds for the expected profile: $\mathbb{E}\, W_n \sim n/\sqrt{4\pi \log n}$ (43; 35). Devroye and Hwang (35) also showed that the level K_n of the width is strongly concentrated.

7.4 Recursive Trees

A recursive tree is a (non-planar) rooted tree (with n nodes) where the nodes are labeled with $1, 2, \ldots, n$ such that all successors of each node have a larger label. In particular, the root has label 1, and every path from the root to a leaf has strictly increasing labels. It is also possible to consider a recursive tree as the result of an evolution process. The *process* starts with the root (that gets label 1). Next, another node is attached to the root (that gets label 2) and in every step a new node is attached to an already existing node (and gets the next label). The labels are the *history* of the tree evolution. By definition it is clear the number y_n of increasing trees of size n is given by $(n-1)!$. Note that the generating function

$$y(z) = \sum_{n\ge 1} y_n \frac{z^n}{n!} = \sum_{n\ge 1} \frac{z^n}{n} = \log \frac{1}{1-x}$$

satisfies the differential equation

$$y'(z) = e^{y(z)} = 1 + y(z) + \frac{1}{2} y(z)^2 + \cdots,$$

which has a natural combinatorial explanation.

Let H_n denote the height of random recursive trees of size n. By using this combinatorial interpretation we also get a recurrence relation for the generating functions

$$y_k(z) = \sum_{n\ge 0} \mathbb{P}\{H_n \le k\} \frac{z^n}{n}.$$

They satisfy the recurrence

$$y_{k+1}'(z) = e^{y_k(z)}$$

with $y_0(z) = 0$ and $y_{k+1}(0) = 0$.

Szymanski (85) was the first who studied the height H_n of recursive trees. He proved then $(1-\varepsilon) \le H_n/\log n \le e$ in probability. Later Pittel (79) proved that the upper bound is the correct limit: $H_n/\log n \to e$ in probability. The following result that settles the distribution and proves strong concentration is due to the author (and is contained in a forthcoming paper (38)).

Theorem 17 *Let H_n denote the height of random recursive trees of size n.*

$$\mathbb{E}\, H_n = e \log n + O\big(\sqrt{\log n}\,(\log\log n)\big).$$

Furthermore we have (uniformly for all $k \ge 0$ as $n \to \infty$)

$$\mathbb{P}\{H_n \le k\} = F(n/y_k'(1)) + o(1),$$

where $F(y)$ satisfies the integral equation

$$y\,F(y/e^{1/e}) = \int_0^y F(z/e^{1/e})F(y-z)\,dz. \tag{33}$$

Moreover, as $n \to \infty$,

$$\mathbb{V} H_n = O(1)$$

and there exist $\eta > 0$ and $c > 0$ such that

$$\mathbb{P}\{|H_n - \mathbb{E}\, H_n| \ge y\} \le c\, e^{-\eta y}$$

for all $y \ge 0$.

Again the proof of the concentration property is similar to that of Section 8 with is based on generating functions.

The maximum out-degree $\Delta(T_n)$ of random recursive trees was studied by Szymanski (85), Devroye and Lu (36) and Goh and Schmutz (53).

Theorem 18 *Let $\Delta(T_n)$ denote the maximum out-degree or random recursive trees. Then we have $\mathbb{E}\,\Delta(T_n) \sim \log_2 n$ and the distribution is given by*

$$\mathbf{P}\{\Delta(T_n) \le k\} = \exp\left(-2^{-(k-\log_2 n+1)}\right) + o(1). \tag{34}$$

Unfortunately it is not known whether the variance (and other central moments) stay bounded as $n \to \infty$ although Theorem 18 suggests such a property.

The proof of (34) relies on a careful analysis of the generating functions

$$y_d(z) = \sum_{n\ge 0} \mathbb{P}\{\Delta(T_n) \le d\} \frac{z^n}{n},$$

that satisfy the differential equations

$$y_d'(x) = \sum_{\ell=0}^{d} \frac{1}{\ell!}\, y_d(x)^\ell.$$

7.5 Scale-Free Trees

So-called *scale-free trees* are defined similarly to recursive trees. Fix some $r > 0$[(vi)] and let a random (recursive) tree grow by the following random procedure.

The process starts with the root that is labeled with 1. Then at step j a new node (with label j) is attached to any previous node of outdegree d with probability proportial to $d + r$

For $r = 1$ one exactly gets *plane oriented recursive trees* or *heap ordered trees.* Note, too, that this process is almost the same as the G_1^n-construction of scale-free graphs.

The notion *scale-free* comes again from the asymptotic degree distribution. The probability that a random node has out-degree d is asymptically given by

$$\lambda_d = \frac{(r+1)\Gamma(2r+1)\Gamma(r+d)}{\Gamma(r)\Gamma(2r+d+2)} \sim \frac{(r+1)\Gamma(2r+1)}{\Gamma(r)} \cdot d^{-2-r}.$$

Pittel (79) has already shown that the height H_n of scale-free trees satisfies $H_n/\log n \to c_r$ in probability; c_r is defined in Theorem 19. The next theorem (by the author (38)) provides strong concentrations and a distributional result for rational r.

The formulation of the theorem uses a sequence of (generating) functions defined by $y_0(x) = 0$ and $y_{k+1}'(x) = (1 - y_k(x))^{-r}$ with $y_{k+1}(0) = 0$. (In Section 8 we will discuss these functions in detail.)

(vi) Sometimes the parameter $\beta = r - 1 > -1$ is used to define scale-free trees.

Theorem 19 *Suppose that $r = \frac{A}{B} > 0$ is rational (where A, B are positive coprime integers). Set $c_r = r/((r+1)\gamma)$, where γ is the real solution of $\gamma e^{1+\gamma/r} = 1$ Then*

$$\mathbb{E}\, H_n \sim c_r \log n.$$

Furthermore we have (uniformly for all $k \ge 0$ as $n \to \infty$)

$$\mathbf{P}\{H_n \le k\} = G^{(r)}\big((r+1)n/(y_k'(1/(r+1)))^{1+\frac{1}{r}}\big) + o(1). \tag{35}$$

The function $G^{(r)}(y)$ is given by

$$G^{(r)}(y) = \frac{\Gamma\left(\frac{A}{A+B}\right)}{\Gamma\left(\frac{1}{A+B}\right)^A} \int\limits_{z_1+\cdots+z_A=1, z_j\ge 0} \prod_{j=1}^{A} \left(F(yz_j) z_j^{\frac{1}{A+B}-1}\right) d\mathbf{z} \tag{36}$$

and $F(y)$ satisfies the integral equation

$$y^{\frac{1}{d-1}} F(ye^{-1/c_r'}) = \frac{\Gamma\left(1+\frac{1}{A+B}\right)}{\Gamma\left(\frac{1}{A+B}\right)^{A+B+1}} \int\limits_{y_1+\cdots+y_{A+B+1}=y, y_j\ge 0} \prod_{j=1}^{B+1} \left(F(y_j e^{-1/c_r'}) y_j^{\frac{1}{A+B}-1}\right) \times \prod_{\ell=B+2}^{A+B+1} \left(F(y_\ell) y_\ell^{\frac{1}{A+B}-1}\right) d\mathbf{y}.$$

Moreover, for all $r > 0$ we have, as $n \to \infty$,

$$\mathbb{V} H_n = O(1)$$

and there exist $\eta > 0$ and $c > 0$ such that

$$\mathbb{P}\{|H_n - \mathbb{E}\, H_n| \ge y\} \ll e^{-\eta y}$$

for all $y \ge 0$.

There is no doubt that (35) is also true for irrational $r > 0$. However, the methods of (38) are not strong enough to prove this more general case. In Section 8 we will present a short proof of the concentration property of H_n.

Recently the maximum out-degree of scale-free trees was discussed by Mori (72)

Theorem 20 *Let $\Delta(T_n)$ denote the maximum degree of scale-free trees with parameter $r > 0$. Then*

$$\frac{\Delta(T_n)}{n^{\frac{1}{1+r}}} \to \mu \quad (a.s.)$$

for some random variable μ (that is related to the degree distribution of T_n). Further

$$\frac{\Delta(T_n) - \mu n^{\frac{1}{1+r}}}{\sqrt{\mu n^{\frac{1}{1+r}}}} \xrightarrow{\mathrm{d}} N(0,1).$$

In particular, we have no concentration if we just scale by average order of magnitude $n^{\frac{1}{1+r}}$. However, if we subtract $n^{\frac{1}{1+r}}\,\mu$ then there is concentration (and a central limit theorem). The proof of this theorem relies on martingale methods.

7.6 Tries

Tries are rooted trees which are used to store data which are labeled with a (possibly) infinite string of symbols from a finite alphabet. For simplicity we first restrict ourselves to the binary alphabet $\{0, 1\}$.

Let $\mathbf{x}_1 = (x_{10}, x_{11}, \ldots), \ldots, \mathbf{x}_n = (x_{n0}, x_{n1}, \ldots) \in \{0,1\}^\omega$ be labels of n data. Each string $\mathbf{x}_j$ defines an infinite path in the (infinite) binary tree; 0 denotes to go to the left subtrees and 1 to go to the right subtree. If the $\mathbf{x}_j$ are different then each infinite path ends with a suffix path that is traversed by that string only.

Let u_j denote the node where the suffix part starts. Then we can *trim* the tree by cutting away everything below node u_j. The node u_j becomes now a leaf representing $\mathbf{x}_j$. If we repeat this procedure for all labels $\mathbf{x}_1, \ldots, \mathbf{x}_n$ then we obtain a finite binary tree with n nodes, called the *trie*.

It is usual to assume that the strings follow the Bernoulli model with probability $p \in (0,1)$, or equivalently they are obtained from a memoryless source on 2 symbols.

In the symmetric case the limiting distribution of the height is well-known, compare with Flajolet (46).

Theorem 21 *Let H_n denote the height of tries in the symmetric case $p = \frac{1}{2}$. Then*

$$\mathbb{P}\{H_n \le k\} = \exp\left(-2^{-(k-2\log_2 n)-1}\right) + o(1).$$

Furthermore, as $n \to \infty$,

$$\mathbb{E}\, H_n = 2\log_2 n + O(1), \quad \mathbb{V} H_n = O(1)$$

and there exist $\eta > 0$ and $c > 0$ such that

$$\mathbb{P}\{|H_n - \mathbb{E}\, H_n| \ge y\} \ll e^{-\eta\, y}$$

for all $y \ge 0$.

This result is, in fact, quite easy to obtain, since the generating function of the height distribution is given by

$$H_k(x) = \sum_{n\ge 0} \mathbb{P}\{H_n \le k\} \frac{x^n}{n!} = \left(1 + \frac{x}{2^k}\right)^{2^k}$$

which directly implies that

$$\mathbb{P}\{H_n \le k\} = \frac{n!}{2^{kn}} \binom{2^k}{n}.$$

However, in (46) this result is proved in a more general context for so-called b-tries.

Note that an infinite 0-1-string can be identified with the binary expansion of a real number in $[0,1]$ and that the symmetric Bernoulli model is equivalent to uniform distribution on $[0,1]$. Thus, there is a natural generalization. One can consider a density $f(x)$ on $[0,1]$ and take $\mathbf{X}_j$ iid according to that law. Devroye (31) analyzed the corresponding trie and obtained a direct generalization of Theorem 21.

Theorem 22 *Let H_n denote the height of tries of generated by iid labels with $f(x)$ as a density on $[0,1]$. If $C := \int_0^1 f(x)^2\, dx < \infty$ then*

$$\mathbb{P}\{H_n \le k\} = \exp\left(-C\, 2^{-(k-2\log_2 n)-1}\right) + o(1).$$

Furthermore,

$$\mathbb{E}\, H_n = 2\log_2 n + O(1).$$

If $\int_0^1 f(x)^2\, dx = \infty$ then $\mathbb{E}\, H_n = \infty$ for all $n \ge 2$.

On the other hand Pittel (78) obtained a very general result in the asymmetric case in the case of an m-ary alphabet with probability distribution $p_1, p_2, \ldots, p_m$.

Theorem 23 *Let H_n denote height of tries of generated by iid labels on an m-ary alphabet with probability distribution $p_1, p_2, \ldots, p_m$ and set*

$$b = \left(\sum_{i=1}^m p_i^2\right)^{-\frac{1}{2}}.$$

Then

$$\mathbb{P}\{H_n \le k\} = \exp\left(-\frac{1}{2}\, b^{-2(k-\log_b n)}\right) + o(1).$$

However, it is unknown whether the variance is bounded (as it is suggested by this result).

We finally want to mention that Devroye (33) applied Talagrand type inequalities to very general trie (and PATRICIA trie) versions and obtained (weak) concentration just by the assumption that $\mathbb{E}\, H_n \to \infty$.

7.7 Digital Search Trees

Digital search trees are again rooted trees which are used to store data which are labeled with a (possibly) infinite string of symbols from a finite alphabet. Digital search trees are closely related to the Lempel-Ziv scheme.

We will first restrict ourselves on the binary alphabet $\{0,1\}$. Suppose that we have n data of that kind. Then the digital search tree algorithm runs as follows. The empty string is stored in the root, while the first item occupies the right or left child of the root depending whether its first symbol is "1" or "0". The remaining items are always stored in the next available node according to the rule that we move to the right if the next symbol is "1" and we move to the left if the next symbol is "0". In the same way we can also search for a specific item. (For more details see (61) and (66).)

As above we assume that the 0-1-strings follow the Bernoulli model with probability $p \in (0,1)$ With respect to this probabilistic model many parameters of digital search trees have been very well studied, compare (again) with (61) and (66). However, it turns out that the symmetric case is much more easy to handle than the asymmetric case.

The following theorem is due to the author (39) (compare also with (60) for a formal asymptotic but non-rigorous approach).

Theorem 24 *Let H_n denote the height of digital search trees in the symmetric case $p = \frac{1}{2}$. Then there exists a sequence k_n that is asymptotically given by*

$$k_n = \log_2 n + \sqrt{2\log_2 n} - \log_2\left(\sqrt{2\log_2 n}\right) + O(1)$$

such that

$$\mathbb{P}\{k_n \le H_n \le k_n + 1\} = 1 + o(1).$$

Furthermore, if $0 \le y \le c_1\sqrt{\log n}$ (for some constant $c_1 > 0$) then there exist constants $c_2, c_3 > 0$ with

$$\mathbb{P}\{|H_n - k_n| \ge y\} \le c_2\, e^{-yc_3\sqrt{\log_2 n}}.$$

This answers a question that is contributed to Kesten (see (5)). The height is concentrated at at most two consecutive levels (and we have extremely small tail estimates).

It is expected that the height is also strongly concentrated in the asymmetric case. But it seems that this is very difficult to prove. The only known result on the height is due to Pittel (77) who has proved that the height of digital search trees with labels on an m-ary alphabet with probability distribution $p_1, p_2, \ldots, p_m$ satisfies

$$\frac{H_n}{\log n} \to \frac{1}{\min\limits_{1\le j\le m} \log 1/p_j} \quad a.s.$$

We also want to mention a universal approach to random trees by Devroye (34) that covers several kinds of trees: m-ary search trees, tries, digital search trees, quadtrees etc. For example, the height H_n of these so-called *split trees* always satisfies the relation

$$\frac{H_n}{\log n} \to c$$

in probability (for some computable constant $c > 0$). It is believed that all split trees have the property that the height is strongly concentrated.

8 The Height of Scale Free Trees

We finally present a proof of the strong concentration property of the height scale-free trees. It is (again) remarkable that it is *easier* to prove concentration than to provide precise asymptotics for the expected height.

We first recall that scale-free trees are a special case of *increasing trees* that have been introduced by Bergeron, Flajolet, and Salvy (10). There one considers the class of all planar recursive trees and associates a *weight* to these trees in the following way. Let ψ_j, $j \ge 0$, be a given sequence of non-negative numbers with $\psi_0 > 0$. Then the weight $\omega(t)$ of a recursive tree t is defined by

$$\omega(t) = \prod_{j\ge 0} \psi_j^{D_j(t)},$$

where $D_j(t)$ denotes the number of nodes in t with j successors. Let J_n denote the set of recursive trees of size n then we set

$$y_n = \sum_{t \in J_n} \omega(t)$$

and

$$y(z) = \sum_{n \geq 0} y_n \frac{z^n}{n!}.$$

By definition it is clear that the generating function $y(z)$ satisfies the differential equation

$$y'(z) = \Psi(y(z)), \quad y(0) = 0, \tag{37}$$

where

$$\Psi(w) = \sum_{j \geq 0} \psi_j w^j.$$

The weights $\omega(t)$ induce a natural probability model on J_n. The probability of a tree $t \in J_n$ is given by $P_n(t) = \omega(t)/y_n$.

For $\Psi(w) = (1+w)^2$ we recover binary search trees (as binary increasing trees) and with $\Psi(w) = e^w$ we can model random recursive trees. Further, if we use $\Psi(w) = (1-w)^{-r}$ then we exactly get scale-free trees.

This can be seen in the following way. In a recent paper Panholzer and Prodinger (75) proved that there are exactly three families where the sequence P_n of probability measures on J_n is induced by a (natural) tree evolution process (described below) if and only if $\Psi(t)$ has one of the three forms:

- $\Psi(w) = \psi_0 \left(1 + (\psi_1/(d\psi_0))w\right)^d$ for some $d \in \{2, 3, \ldots\}$ and $\psi_0 > 0$, $\psi_1 > 0$.
- $\Psi(w) = \psi_0 e^{\frac{\psi_1}{\psi_0} w}$ with $\psi_0 > 0$, $\psi_1 > 0$.
- $\Psi(w) = \psi_0 \left(1 - \dfrac{\psi_1}{r\phi_0} w\right)^{-r}$ for some $r > 0$ and $\psi_0 > 0$, $\psi_1 > 0$.

The corresponding tree evolution process runs as follows The starting point is (again) one node (the root) with label 1. Now assume that a tree t is size n is present. We attach to every node v of t a *local weight* $\rho(v) = (k+1)\psi_{k+1}\psi_0/\psi_k$ when v has k successors and set $\rho(t) = \sum_{v \in t} \rho(v)$. Observe that in a planar tree there are $k+1$ different ways to attack a new (labeled) node to an (already existing) node with k successors. Now choose a node v in t according to the probability distribution $\rho(v)/\rho(t)$ and then independently and uniformly one of the $k+1$ possibilities to attach a new node there (when v has k successors). This construction ensures that in these three particular cases a tree t of size n that occurs with probability proportional to $\omega(t)$ generates a tree t' of size $n+1$ with probability that is proportional to $\omega(t)\psi_{k+1}\psi_0/\psi_k$ which equals $\omega(t')$. Thus, this procedure induces the same probability distribution on J_n as the above mentionen one where a tree $t \in J_n$ has probablility $\omega(t)/y_n$.

Note that if we are only interested in the distributions P_n then we can work (without loss of generality) with some special values for ψ_0 and ψ_1. It is sufficient to consider the generating functions

- $\Psi(w) = (1+w)^d$ for some $d \in \{2, 3, \ldots\}$ (d-ary increasing trees).
- $\Psi(w) = e^w$ (recursive trees).
- $\Psi(w) = (1-w)^{-r}$ for some $r > 0$ (scale-free trees).

In the third class, the probabilty of choosing a node with out-degree j is proportional to $j + r$ (as a short calculation shows).

In particular, this shows that scale-free trees have two completely different ways of description, the tree evolution process and the recursive description (splitting at the root) that corresponds to (37).

In our case ($\Psi(w) = (1-w)^{-r}$) the generating function $y(z) = \sum_{n \geq 1} y_n z^n/n!$ satisfies $y'(z) = (1 - y(z))^{-r}$ and is explicitly given by

$$y(z) = 1 - \left(1 - (r+1)z\right)^{1/(r+1)}.$$

The coefficients are also explicit:

$$y_n = n!(-1)^{n-1}(r+1)^n \binom{1/(r+1)}{n}.$$

The advantage of the recursive description is that one can also get a recurrence of the generating functions of the height distribution. If we set

$$y_k(z) = \sum_{n\ge 0} y_n \mathbb{P}\{H_n \le k\} \frac{z^n}{n!}$$

then we have $y_0(z) = 0$ and recursively

$$y'_{k+1}(z) = \frac{1}{(1-y_k(z))^r} \qquad (y_{k+1}(0) = 0).$$

By taking derivatives it follows that

$$y''_{k+1}(z) = r\big(y'_{k+1}(z)\big)^{1+\frac{1}{r}} y'_k(z).$$

If we set

$$Y_k(z) = y'_k(z)$$

then we have $Y_1(z) = 1$ and the recurrence relation

$$Y'_{k+1}(z) = r\, Y_{k+1}(z)^{1+\frac{1}{r}} Y_k(z) \qquad (Y_{k+1}(0) = 1). \tag{38}$$

Note further that the functions $Y_k(z)$ encodes the height distribution, too:

$$Y_k(z) = y'_k(z) = \sum_{n\ge 0} y_{n+1} \mathbb{P}\{H_{n+1} \le k\} \frac{z^n}{n!}.$$

The first lemma is one of the key properties of the proof. It will provide us upper and lower bounds.

Lemma 1 *Suppose that $Y_1(z), Y_2(z), \overline{Y}_1(z), \overline{Y}_2(z)$ are non-negative continuous functions that are defined for $z \ge 0$ such that $Y_1(0) < \overline{Y}_1(0)$, $Y_2(0) < \overline{Y}_2(0)$, $Y'_2(z) = r\,Y_2(z)^{1+\frac{1}{r}} Y_1(z)$, $\overline{Y}'_2(z) = r\,\overline{Y}_2(z)^{1+\frac{1}{r}} \overline{Y}_1(z)$, and that the difference $\overline{Y}_1(z) - Y_1(z)$ has exactly one positive zero. Then the difference $\overline{Y}_2(z) - Y_2(z)$ has at most one positive zero.*

Proof: For $j = 1, 2$ set

$$y_j(z) = \int_0^z Y_j(t)\,dt \quad \text{and} \quad \overline{y}_j(z) = \int_0^z \overline{Y}_j(t)\,dt.$$

Then we have $y_1(z) < \overline{y}_1(z)$, $y_2(z) < \overline{y}_2(z)$ (at least) for a small interval $0 < z < \zeta$ and also $y'_2(z) = (1-y_1(z))^{-r}$ and $\overline{y}'_2(z) = (1-\overline{y}_1(z))^{-r}$. Furthermore, since $\overline{Y}_1(z) - Y_1(z)$ is positive (for small positive z) and has at most one positive zero, the same follows for

$$\overline{y}_1(z) - y_1(z) = \int_0^z \left(\overline{Y}_1(t) - Y_1(t)\right)\,dt,$$

This can be seen in the following way. Suppose that $\overline{Y}_1(z) \ge Y_1(z)$ for $0 \le z \le z_0$ and $\overline{Y}_1(z) \le Y_1(z)$ for $z \ge z_0$, that is, z_0 is the only (positive) zero of the difference $\overline{Y}_1(z) - Y_1(z)$. Since $\overline{y}'_1(z) - y'_1(z) = \overline{Y}_1(z) - Y_1(z)$ the same is true for the difference $\overline{y}'_1(z) - y'_1(z)$. Hence, the difference $\overline{y}_1(z) - y_1(z)$ is increasing for $0 \le z \le z_0$ and decreasing for $z \ge z_0$. Since $\overline{y}_1(0) > y_1(0)$ it directly follows that the difference $\overline{y}_1(z) - y_1(z)$ has at most one zero.

Now observe that $((1-y)^{-r} - (1-z)^{-r}/(y-z) > 0$ for real $y, z < 1$. Hence,

$$\overline{Y}_2(z) - Y_2(z) = \overline{y}'_2(z) - y'_2(z) = \frac{(1-\overline{y}_1(z))^{-r} - (1-y_1(z))^{-r}}{\overline{y}_1(z) - y_1(z)}\,(\overline{y}_1(z) - y_1(z))$$

has at most one positive zero, too. $\square$

Lemma 2 *The sequence $Y_k(1/(r+1))$ is log-concave, that is,*

$$\frac{Y_{k+2}(1/(r+1))}{Y_{k+1}(1/(r+1))} \le \frac{Y_{k+1}(1/(r+1))}{Y_k(1/(r+1))}.$$

Proof: For $0 \le \gamma < 1$ set

$$V_k(z,\gamma) = \begin{cases} (1-(r+1)z)^{-r/(r+1)} & \text{for } 0 \le z \le \frac{1}{r+1}(1-\gamma), \\ \gamma^{-r/(r+1)} Y_k\left(\frac{z-\frac{1}{r+1}(1-\gamma)}{\gamma}\right) & \text{for } \frac{1}{r+1}(1-\gamma) \le z \le \frac{1}{r+1}. \end{cases}$$

These functions satisfy

$$V_{k+1}'(z,\gamma) = r\,V_{k+1}(z,\gamma)^{1+\frac{1}{r}} V_k(z,\gamma),$$

$V_k(0) = 1$ and $V_k(1/(r+1),\gamma) = \gamma^{-r/(r+1)} Y_k(1/(r+1))$. In particular, for

$$\gamma_k = \left(\frac{Y_k(1/(r+1))}{Y_{k+1}(1/(r+1))}\right)^{1+\frac{1}{r}}$$

we have $V_k(1/(r+1),\gamma_k) = Y_{k+1}(1/(r+1))$. Now inductive application of Lemma 1 shows that $Y_{k+1}(z) - V_k(z,\gamma)$ has (at most) one positive zero. Since $V_k(1/(r+1),\gamma_k) = Y_{k+1}(1/(r+1))$ this implies that

$$Y_{k+1}(z) \le V_k(z,\gamma_k) \quad \text{for} \quad 0 \le z \le \frac{1}{r+1}$$

and after integration

$$Y_{k+2}(1/(r+1)) \le V_{k+1}(1/(r+1),\gamma_k) = \gamma_k^{-r/(r+1)} Y_{k+1}(1/(r+1)) = \frac{Y_{k+1}(1/(r+1))^2}{Y_k(1/(r+1))}.$$

This completes the proof of the lemma. $\square$

Note that Lemma 2 shows that the sequence $Y_{k+1}(1/(r+1))/Y_k(1/(r+1))$ has a limit $\rho \ge 1$. The next step is to show that this limit is actually > 1. This will be done with help of the next lemma.

Lemma 3 *For $0 \le z < \frac{1}{r+1}$ and $k \ge 1$ we have*

$$Y(z) - Y_k(z) \le \left(\frac{2r+1}{r+1}\right)^k \sum_{\ell \ge k} \frac{1}{\ell!}\left(\log\frac{1}{1-(r+1)z)}\right)^\ell. \tag{39}$$

Proof: We proceed by induction. Since $Y_1(z) = 1$ and $(1-(r+1)z)^{-r/(r+1)} \le (1-(r+1)z)^{-1}$ we immediately get (39) for $k = 1$. Now, we can inductively use

$$\begin{aligned} Y(z)' - Y_{k+1}(z)' &= r\left(Y(z)^{2+\frac{1}{r}} - Y_{k+1}(z)^{1+\frac{1}{r}} Y_k(z)\right) \\ &\le r\left(Y(z)^{2+\frac{1}{r}} - Y_k(z)^{2+\frac{1}{r}}\right) \\ &\le r\left(2+\frac{1}{r}\right) Y(z)^{1+\frac{1}{r}} \left(Y(z) - Y_k(z)\right) \\ &= (2r+1)\frac{1}{1-(r+1)z}\left(Y(z) - Y_k(z)\right) \end{aligned}$$

to complete the proof of (39). $\square$

Lemma 4 *For every C with $1 < C < e^{r/(r+1)}$ that satisfies $C > \left(2+\frac{1}{r}\right) e \log C$ we have*

$$\frac{Y_{k+1}(1/(r+1))}{Y_k(1/(r+1))} \ge C \tag{40}$$

for all $k \ge 1$.

Note that if C is close to 1 then the condition $C > \left(2+\frac{1}{r}\right) e \log C$ is surely satisfied.

Proof: Set $c = C^{-r/(r+1)}$ and $z_0 = \frac{1}{r+1}(1-c^k)$. Then $\log\frac{1}{c} < 1$ and we get from (39)

$$\begin{aligned}
Y(z_0) - Y_k(z_0) &\le \left(\frac{2r+1}{r+1}\right)^k \sum_{\ell\ge k} \frac{1}{\ell!}\left(k\log\frac{1}{c}\right)^\ell \\
&\le c_1 \left(\frac{2r+1}{r+1}\right)^k \frac{\left(k\log\frac{1}{c}\right)^k}{k!} \\
&\le c_1 \left(\frac{2r+1}{r+1} e \log\frac{1}{c}\right)^k \\
&= c_1 \left(\left(2+\frac{1}{r}\right) e\log C\right)^k,
\end{aligned}$$

where $c_1 > 0$. On the other hand we have $Y(z_0) = C^k$. Since $C > \left(2+\frac{1}{r}\right) e\log C$ this implies that

$$Y_k(1/(r+1)) \ge Y_k(z_0) \ge C^k(1+o(1)).$$

Thus, the limit has to be bounded by

$$\rho = \lim_{k\to\infty} \frac{Y_{k+1}(1/(r+1))}{Y_k(1/(r+1))} \ge C.$$

Consequently (40) follows from Lemma 2. □

In a next step we will prove bounds for $\mathbb{P}\{H_n \le k\}$.

Lemma 5 *If* $n \ge Y_k(1/(r+1))^{1+\frac{1}{r}}$ *then*

$$\mathbb{P}\{H_n \le k\} = O\left(Y_k(1/(r+1)) \cdot n^{-r/(r+1)}\right). \tag{41}$$

Conversely if $n \le Y_k(1/(r+1))^{1+\frac{1}{r}}$ *then*

$$\mathbb{P}\{H_n > k\} = O\left(Y_k(1/(r+1))^{-1-\frac{1}{r}} \cdot n\right). \tag{42}$$

Proof: Let z_k be defined by $Y(z_k) = Y_k(1/(1+r))$. Set $\eta_k = Y_k(1/(1+r))$. Then

$$z_k = \frac{1}{r+1}\left(1 - \eta_k^{-1-\frac{1}{r}}\right).$$

Further set $\tilde{Y}(z) = (z_k(r+1))^{r/(r+1)} Y(z_k(r+1)z)$. Then $\tilde{Y}(0) < 1 = Y_k(0)$ and

$$\tilde{Y}'(z) = r\tilde{Y}(z)^{2+\frac{1}{r}}.$$

Hence, a successive application of Lemma 1 implies that $\tilde{Y}(z) \le Y_k(z)$ for $0 \le z \le z_k$ and $\tilde{Y}(z) \ge Y_k(z)$ for $z \ge z_k$. Further, by construction we know that $\mathbb{P}\{H_{n+1} \le k\} \le \mathbb{P}\{H_n \le k\}$.

If $z \ge z_k$ then

$$\begin{aligned}
\tilde{Y}(z) &\ge Y_k(z) \\
&\ge \sum_{\ell=0}^{n-1} y_{\ell+1} \mathbb{P}\{H_{\ell+1} \le k\} \frac{z^\ell}{\ell!} \\
&\ge \mathbb{P}\{H_n \le k\} \sum_{\ell=0}^{n-1} y_{\ell+1} \frac{z^\ell}{\ell!}
\end{aligned}$$

Now suppose that $n \ge \eta_k^{1+\frac{1}{r}}$ and use $z = \frac{1}{r+1}$. Then we have (after a short elementary calculation)

$$\tilde{Y}(1/(r+1)) \le c_2 \eta_k$$

and

$$\sum_{\ell=0}^{n-1} y_{\ell+1}\frac{(r+1)^\ell}{\ell!} \ge c_2 n^{r/(r+1)}$$

(for some constants $c_2 > 0$ and $c_3 > 0$). Thus,

$$\mathbb{P}\{H_n \le k\} \le c_4 \eta_k n^{-r/(r+1)}.$$

Similarly we obtain an upper bound for $\mathbb{P}\{H_n > k\}$. If $0 \le z \le z_k$ then

$$\begin{aligned} Y(z) - \tilde{Y}(z) &\ge Y(z) - Y_k(z) \\ &\ge \sum_{\ell=n-1}^{\infty} y_{\ell+1}\mathbb{P}\{H_{\ell+1} > k\}\frac{z^\ell}{\ell!} \\ &\ge \mathbb{P}\{H_n > k\} \sum_{\ell=n-1}^{\infty} y_{\ell+1}\frac{z^\ell}{\ell!}. \end{aligned}$$

Here we use $z' = \frac{1}{r+1}\left(1-\frac{1}{n}\right)$. If $n \le \eta_k^{1+\frac{1}{r}}$ then $z' \le z_k$ and we get (again after a short elementary calculation)

$$Y(z') - \tilde{Y}(z') \le c_5 n^{1+\frac{r}{r+1}} \eta_k^{-1-\frac{1}{r}}$$

and

$$\sum_{\ell=n-1}^{\infty} y_{\ell+1}\frac{(z')^\ell}{\ell!} \ge c_6\, n^{r/(r+1)}$$

(for some constants $c_5 > 0$ and $c_6 > 0$). Of course this proves

$$\mathbb{P}\{H_n > k\} \le c_7 \eta_k^{-1-\frac{1}{r}}\, n$$

and completes the proof of the lemma. □

By combining Lemma 4 and Lemma 5 we finally obtain exponential tail estimates.

Lemma 6 *Let* $k(n) := \max\{\ell \ge 1 : Y_\ell(1/(r+1))^{1+\frac{1}{r}} \le n\}$. *Then*

$$\mathbb{E}\, H_n = k(n) + O(1) \tag{43}$$

and there exist $\eta > 0$ *and* $c > 0$ *such that*

$$\mathbb{P}\{|H_n - \mathbb{E}\, H_n| > y\} \le ce^{-\eta y}$$

for all $y > 0$.

Proof: From Lemma 4 we get for all $\ell \ge 0$

$$\frac{Y_{k(n)+\ell}(1/(r+1))}{Y_{k(n)}(1/(r+1))} \ge C^\ell$$

and consequently (42) gives

$$\begin{aligned} \mathbb{P}\{H_n > k(n)+\ell\} &\le c_8 Y_{k(n)+\ell}(1/(r+1))^{-1-\frac{1}{r}}\, n \\ &\le c_8\, C^{\frac{r+1}{r}\ell}\, Y_{k(n)}(1/(r+1))^{-1-\frac{1}{r}}\, n \\ &\le c_9 C^{-\frac{r+1}{r}\ell}. \end{aligned}$$

Similarly we get (with help of Lemma 4 and (41))

$$\mathbb{P}\{H_n \le k(n) - \ell\} \le c_{10} C^{-\ell}.$$

Thus, there exist $\eta > 0$ and $c > 0$ with

$$\mathbb{P}\{|H_n - k(n)| \ge y\} \le ce^{-\eta y}$$

for all $y > 0$. Of course this implies the lemma. □

The above proof provides stong concentration around the mean but it does not say where the mean value actually is. We only know that it is related to the growth of $Y_k(1/(r+1))$, compare with (43). Thus one has to analyze the recurrence (38) in more detail. One essential tool to do this is the function $G^{(r)}(y)$ that is defined in (36). Namely, if we set

$$\overline{Y}_k(z) = \alpha^{\frac{rk}{1+r}} \Phi\left(\alpha^k(\rho - z)\right)$$

where $\alpha = e^{1/c_r}$ and

$$\Phi(u) = \frac{1}{(r+1)^{\frac{r}{1+r}}\Gamma\left(\frac{r}{1+r}\right)} \int_0^\infty G^{(r)}(y) y^{-\frac{1}{1+r}} e^{-yu}\, dy$$

then we have

$$\overline{Y}'_{k+1}(z) = r\big(\overline{Y}_{k+1}(z)\big)^{1+\frac{1}{r}}\overline{Y}_k(z),$$

$0 < \overline{Y}_k(0) < 1$ and $\overline{Y}_k(1/(r+1)) = \alpha^{rk/(1+r)}$.

Thus, the sequence of function $\overline{Y}_k(z)$ satisfies the same recurrence as $Y_k(z)$ and, in fact, it is possible to approximate $Y_k(z)$ with help of $\overline{Y}_k(z)$ in some way. Since we have direct access to $\overline{Y}_k(1/(r+1))$ we also get some information on $Y_k(1/(r+1))$ which determines the expected value. Further, it is also possible to approximate the coefficients of $Y_k(z)$ in terms of the coefficients of $\overline{Y}_k(z)$ which provides the asymptotic distribution of H_n. However, this step is very technical, compare with (38).

Acknowledgements

The author is supported by the Austrian Science Foundation (FWF) FSP-Project S9604.

References

[1] D. J. Aldous, The continuum random tree I, Ann. Prob. 19 (1991), 1–28.

[2] D. J. Aldous, The continuum random tree II: an overview, Stochastic Analysis, M. T. Barlow and N. H. Bingham, Eds., Cambridge University Press 1991, 23–70.

[3] D. J. Aldous, The continuum random tree III, Ann. Prob. 21 (1993), 248–289.

[4] D. Aldous and P. Diaconis, Hammersley interacting particle process and longest increasing subsequences, Prob. Th. and Rel. Fields 103 (1995), 199–213.

[5] D. Aldous and P. Shields, A diffusion limit for a class of random-growing binary trees, Prob. Th. Rel. Fields 79 (1988), 509–542.

[6] N. Alon and M. Krivelevich, The concentration of the chromatic number of random graphs, Combinatorica 17 (1997), 303–313.

[7] N. Alon and J. H. Spencer, The probabilistic method, 2nd ed., Wiley, New York, 2000.

[8] A.-L. Barabási and R. Albert, Emergence of scaling in random networks, Science 286 (1999), 509–512.

[9] J. Beardwood, H. J. Halton and J. M. Hammersley, The shortest path through many points, Proc. Cambridge Phil. Soc. 55 (1959), 299–327.

[10] F. Bergeron, P. Flajolet, and B. Salvy, Varieties of increasing trees, CAAP '92 (Rennes, 1992), 24–48, Lecture Notes in Comput. Sci., 581, Springer, Berlin, 1992.

[11] J. Baik, P. Deift and K. Johansson, On the distribution of the length of the longest increasing subsequence of random permutations, J. Amer. Math. Soc. 12 (1999), 1119–1178.

[12] B. Bollobás, The distribution of the maximum degree of a random graph, Discrete Math. 32 (1980), 201–203.

[13] B. Bollobás, The diameter of random graphs, Trans. Amer. Math. Soc. 267 (1981), 41–52.

[14] B. Bollobás, Degree sequences of random graphs, Discrete Math. 33 (1981), 1–19.

[15] B. Bollobás, Vertices of given degree in a random graph, J. Graph Theory 6 (1982), 147–155.

[16] B. Bollobás, Random graphs, Academic Press, London, 1985.

[17] B. Bollobás, The chromatic number of random graphs, Combinatorica 8 (1988), 49–56.

[18] B. Bollobás and G. Brightwell, The height of a random partial order: concentration of measure, Ann. Appl. Prob. 2 (1992), 1009–1018.

[19] B. Bollobás and F. de la Vega, The diameter of random regular graphs, Combinatorics 2 (1982), 125–134.

[20] B. Bollobás and O. Riordan, The diameter of a scale-free random graph, Combinatorica 24 (2004), 5–34.

[21] B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády, The degree sequence of a scale-free random graph process, Random Struct. Algorithms 18 (2001), 279–290.

[22] S. Boucheron, G. Lugosi, and O. Bousquet, Concentration inequalities, Lecture Notes in Computer Science 3176, Springer, Berlin, 2004.

[23] Ju. D. Burtin, Asymptotic estimates of the diameter and the independence and domination numbers of a random graph, Dokl. Akad. Nauk SSSR 209 (1973), 765–768, translated in Soviet Math. Dokl. 14 (1973), 497—501.

[24] Ju. D. Burtin, Extremal metric characteristics of a random graph. I., Teor. Verojatnost. i Primenen 19 (1974), 740–754.

[25] R. Carr, W. M. Y. Goh and E. Schmutz, The maximum degree in a random tree an related problems, Random Struct. Algorithms 5 (1994), 13–24.

[26] B. Chauvin and M. Drmota, The random multisection problem, travelling waves, and the distribution of the height of m-ary search trees, Algorithmica, to appear (2006).

[27] B. Chauvin, M. Drmota and J. Jabbour-Hattab, The profile of binary search trees, Ann. Appl. Prob. 11 (2001), 1042–1062.

[28] H.-H. Chern, H.-K. Hwang, and T.-H. Tsai, An asymptotic theory for Cauchy–Euler differential equations with applications to the analysis of algorithms. J. Algorithms 44 (2002), 177–225.

[29] N. G. de Bruijn, D. E. Knuth, and S. O. Rice, The average height of planted plane trees, Graph Theory Comput. 15–22, 1972.

[30] E. Deutsch, A. J. Hildebrand and H. S. Wilf, Longest increasing subsequences in pattern-restricted permutations, Electr. J. Combinat. 9 (2003), # R12.

[31] L. Devroye, A probabilistic analysis of the height of tries and of the complexity of triesort, Acta Informatica 21 (1984), 229–237.

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

[33] L. Devroye, Universal asymptotics for random tries and PATRICIA tries, Algorithmica 42 (2005), 11–29.

[34] L. Devroye, Universal limit laws for depths in random trees, SIAM J. Comput. 28 (1999), 409–432.

[35] L. Devroye and H.-K. Hwang, Width and mode of the profile for random trees of logarithmic height, Ann. Appl. Probab. 16 (2006), 886–918.

[36] L. Devroye and J. Lu, The strong convergence of maximal degrees in uniform random recursive trees and DAGS, Random Struct. Algorithms 7 (1995), 1–14.

[37] M. Drmota, An Analytic Approach to the Height of Binary Search Trees II, J. Assoc. Comput. Mach. 50 (2003), 333–374.

[38] M. Drmota, The height of increasing trees, manuscript, 2006.

[39] M. Drmota, Height and saturation level of digital search trees, manuscript, 2006.

[40] M. Drmota and B. Gittenberger, On the profile of random trees, Random Struct. Algorithms 10 (1997), 421–451.

[41] M. Drmota and B. Gittenberger, The width of Galton-Watson trees, Discr. Math. Theoret. Comput. Sci. 6 (2004), 387–400.

[42] M. Drmota and B. Gittenberger, The height and profile of Pólya trees, manuscript, 2006.

[43] M. Drmota and H.-K. Hwang, Profile of random trees: correlation and width of random recursive trees and binary search trees, Adv. Appl. Prob. 37 (2005), 1–21.

[44] D. P. Dubhashi and A. Panconesi, Concentration of measure for the analysis of randomized algorithms, draft available at http://www.dsi.uniroma1.it/∼ale.

[45] P. Erdős and G. Szekeres, A combinatorial theorem in geometry, Composition Math. 2 (1935), 463–470.

[46] P. Flajolet, On the performance evaluation of extendible hashing and trie searching, Acta Informatica 20 (1983), 345–359.

[47] P. Flajolet, Z. Gao, A. M. Odlyzko and B. Richmond The distribution of heights of binary trees and other simple trees, Combin. Probab. Comput. 2 (1993), 145–156.

[48] P. Flajolet and A. M. Odlyzko, The average height of binary trees and other simple trees, J. Comput. Syst. Sci. 25 (1982), 171–213.

[49] P. Flajolet and A. M. Odlyzko, Singularity analysis of generating functions, SIAM J. Discrete Math. 3 (1990), 216–240.

[50] A. M. Frieze, On the independence number of random graphs, Discrete Math. 81 (1990), 171–175.

[51] A. Frieze, On the length of the longest monotone subsequence in a random permutation, Ann. Appl. Prob. 1 (1991), 301–305.

[52] W. M. Y. Goh and E. Schmutz, Unlabeled trees: Distribution of the maximum degree, Random Struct. Algorithms 5 (1994), 411-440.

[53] W. M. Y. Goh and E. Schmutz, Limit distribution for the maximum degree of a random recursive tree, J. Comput. Appl. Math. 142 (2002), 61–82.

[54] G. Grimmet and C. McDiarmid, On colouring random graphs, Math. Proc. Camb. Phil. Soc. 77 (1975), 313–324.

[55] G. I. Ivchenko, On the asymptotic behavior of degrees of vertices in a random graph. Theory Probab. Appl. 18, 188-195 (1973); translation from Teor. Veroyatn. Primen. 18 (1973), 195–203.

[56] S. Janson, T. Łuczak and A. Ruciński, Random graphs, Wiley, New York, 2000.

[57] K. Johansson, The longest increasing subsequence in a random permutation and a unitary random matrix model, Math. Res. Letters 5 (1998), 63–82.

[58] R. M. Karp, Probabilistic analysis of partitioning algorithms for the traveling-salesman problem in the plane, Mathematics in Operations Research 2 (1977), 209–244.

[59] D. P. Kennedy, The Galton-Watson process conditioned on the total progeny, J. Appl. Prob. 12 (1975), 800–806.

[60] C. Knessl and W. Szpankowski, Asymptotic behavior of the height in a digital search tree and the longest phrase of the Lempel-Ziv scheme, SIAM J. Computing 30 (2000), 923–964.

[61] D. E. Knuth, The Art of Computer Programming, Vol. 3, 2nd ed., Addison-Wesley, 1998.

[62] B. F. Logan and L. A. Shepp, A variation problem for random Young tableaux, Advances in Math. 26 (1977), 206–222.

[63] T. Łuczak, Random trees and random graphs, Random Struc. Algoríthms 13 (1998), 485–500.

[64] T. Łuczak, A note on the sharp concentration of the chromatic number of random graphs, Combinatorica 11 (1991), 295–297.

[65] G. Lugosi, Concentration-of-measure inequalities, draft available at http://www.econ.upf.es/∼lugosi.

[66] H. M. Mahmoud, Evolution of Random Search Trees, John Wiley & Sons, New York, 1992.

[67] C. McDiarmid, Concentration, Probabilistic methods for algorithmic discrete mathematics, 195–248, Algorithms Combin., 16, Springer, Berlin, 1998.

[68] C. McDiarmid, On the method of bounded differences. In: Surveys in Combinatorics, Proceedings, Norwich, 1989, London Math. Soc. Lecture Note Ser. 141, Cambridge Univ. Press, Cambridge, 148–188, 1989.

[69] M. L. Mehta, Random matrics, 2nd ed., Academic Press, San Diego, 1991.

[70] A. Meir and J. W. Moon, On the altitude of nodes in random trees, Can. J. Math. 30 (1978), 997–1015.

[71] A. Meir and J. W. Moon,, On nodes of large out-degree in random trees, Proceedings of the Twenty-second Southeastern Conference on Combinatorics, Graph Theory and Computing (Baton Rouge, LA, 1991), Congr. Numer. 82 (1991), 3–13.

[72] T. F. Mori, The maximum degree of the Barabási-Albert random tree, Combinat. Probab. Comput. 14 (2005), 339–348.

[73] A. M. Odlyzko and E. M. Rains, On the longest increasing subsequence in random permutations, in: Analysis, geometry, number theory: the mathematics of Leon Ehrenpreis (Philadelphi, PA, 1998), 439–451, Contemp. Math., 2000.

[74] R. Otter, The number of trees, Ann. Math. 49 (1948), 583–599.

[75] A. Panholzer and H. Prodinger, The level of nodes in increasing trees revisited, Random Struct. Algor., to appear.

[76] B. Pittel, On growing random binary trees, J. Math. Anal. Appl. 103 (1984), 461–480.

[77] B. Pittel, Asymptotic growth of a class of random trees, Ann. Probab. 13 (1985), 414–427.

[78] B. Pittel, Path in a random digital tree: liminting distributions, Adv. Appl. Prob. 18 (1986), 139–155.

[79] B. Pittel, Note on the height of random recursive trees and random m-ary search trees, Random Struct. Algor. 6 (1994), 337–347.

[80] G. Pólya, Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen, Acta Math. 68 (1937), 145–254.

[81] B. Reed, The height of a random binary search tree, J. Assoc. Comput. Mach. 50 (2003), 306–332.

[82] O. Riordan and A. Selby, The maximum degree of a random graph, Combin. Probab. Comput. 9 (2000), 549–572.

[83] J. M. Robson, The height of binary search trees, Austral. Comput. J. 11 (1979), 151–153.

[84] J. M. Steele, Complete convergence of short paths and Karp's algorithm for the TSP, Mathematics in Operations Research 6 (1981), 374–378.

[85] J. Szymanski, On the maximum degree and the height of a random recursive tree, Ranndom Graphs 87, in: M. Karonski, J. Jaworski, A. Rusinski eds., Wiley, New York, 1990, 313–324.

[86] T. Seppäläinen, A microscopic model for the Burgers equation and longest increasing subsequences, Electron. J. Prob. 1 no. 5 (1996).

[87] E. Shamir and J. H. Spencer, Sharp concentration of the chromatic number on random graphs $G_{n,p}$, Combinatorica 7 (1987), 124–129.

[88] W. T. Rhee and M. Talagrand, A sharp deviation for the for the stochastic travelling saleman problem, Ann. Probab. 17 (1989), 1–8.

[89] M. Talagrand, Concentration of measure and isoperimetric inequalities in product spaces, Inst. Hautes Études Sci. Publ. Math. 81 (1995), 73–205.

[90] S. M. Ulam, Monte Carlo calculations in problems of mathematical physics, in: Modern Mathematics for the Engineers, E. F. Beckenbach ed., McGraw-Hill, 261–281, 1961.

[91] A. M. Vershik and S. V. Kerov, Asymptotics of the Plancherel measure of the symmetric group and the limiting form of Young tables, Soviet Math. Dokl. 18 (1977), 527–531.