General Edgeworth expansions with applications to profiles of random trees

Alexander Marynych

Taras Shevchenko National University of Kyiv
and
University of Münster

joint work with Z. Kabluchko (Münster, Germany) and
H. Sulzbach (Birmingham, UK)

Assumptions on profiles

A random profile is an arbitrary sequence

$$\mathbb{L}_n = (\mathbb{L}_n(k))_{k\in\mathbb{Z}}, \quad n \in \mathbb{N}$$

of real-valued stochastic processes on $\mathbb{Z}$.

Assumptions on profiles

A random profile is an arbitrary sequence

$$\mathbb{L}_n = (\mathbb{L}_n(k))_{k\in\mathbb{Z}}, \quad n \in \mathbb{N}$$

of real-valued stochastic processes on $\mathbb{Z}$.

Assumption A1: There exists an open interval $(\beta_-, \beta_+) \subset \mathbb{R}$ containing zero such that for every $n \in \mathbb{N}$ and every $\beta \in (\beta_-, \beta_+)$,

$$\sum_{k\in\mathbb{Z}} |\mathbb{L}_n(k)| e^{\beta k} < \infty \quad \text{a.s.}$$

Assumptions on profiles

A random profile is an arbitrary sequence

$$\mathbb{L}_n = (\mathbb{L}_n(k))_{k \in \mathbb{Z}}, \quad n \in \mathbb{N}$$

of real-valued stochastic processes on $\mathbb{Z}$.

Assumption A1: There exists an open interval $(\beta_-, \beta_+) \subset \mathbb{R}$ containing zero such that for every $n \in \mathbb{N}$ and every $\beta \in (\beta_-, \beta_+)$,

$$\sum_{k \in \mathbb{Z}} |\mathbb{L}_n(k)| e^{\beta k} < \infty \quad \text{a.s.}$$

The interval (β_-, β_+) may be infinite. For example, if for every $n \in \mathbb{N}$ the support of a profile $\{k \in \mathbb{Z} \colon \mathbb{L}_n(k) \neq 0\}$ is finite with probability one, then Assumption A1 holds with $(\beta_-, \beta_+) = \mathbb{R}$.

Assumptions on profiles

Assume that there exist

- a sequence $(w_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ such that $\lim_{n\to\infty} w_n = +\infty$;
- an open domain $\mathcal{D} \subset \{\beta \in \mathbb{C} \colon \beta_- < \operatorname{Re}\beta < \beta_+\}$ such that $\mathcal{D} \cap \mathbb{R} = (\beta_-, \beta_+)$;
- a deterministic analytic function $\varphi : \mathcal{D} \to \mathbb{C}$ such that for every $\beta \in (\beta_-, \beta_+)$ we have $\varphi(\beta) \in \mathbb{R}$ and $\varphi''(\beta) > 0$.

From Assumption A1 it follows that with probability one, the normalized Laplace transform

$$W_n(\beta) := e^{-\varphi(\beta) w_n} \sum_{k\in\mathbb{Z}} \mathbb{L}_n(k) e^{\beta k}, \quad \beta \in \mathcal{D},$$

is a random analytic function on $\mathcal{D}$ for every $n \in \mathbb{N}$.

Assumptions on profiles

Assumption A2: With probability one the sequence of random analytic functions $(W_n)_{n\in\mathbb{N}}$ converges locally uniformly on $\mathcal{D}$, as $n\to\infty$, to a random analytic function W_∞ such that $\mathbb{P}[W(\beta)\neq 0$ for all $\beta\in(\beta_-,\beta_+)]=1$.

Assumptions on profiles

Assumption A2: With probability one the sequence of random analytic functions $(W_n)_{n\in\mathbb{N}}$ converges locally uniformly on $\mathcal{D}$, as $n \to \infty$, to a random analytic function W_∞ such that $\mathbb{P}[W(\beta) \neq 0 \text{ for all } \beta \in (\beta_-, \beta_+)] = 1$.

Assume also that the speed of convergence is superpolynomial in w_n:

Assumption A3: For every compact set $K \subset \mathcal{D}$ and every $r \in \mathbb{N}$ there exists an a.s. finite $C_{K,r}$ such that for all $n \in \mathbb{N}$,

$$\sup_{\beta \in K} |W_n(\beta) - W_\infty(\beta)| < C_{K,r} w_n^{-r}.$$

Assumption on profiles

The next assumption is technical. In the classical Edgeworth expansion for sums of i.i.d. variables it corresponds to the assumption that $\mathbb{Z}$ is a minimal lattice containing the support of the step.

Assumption on profiles

The next assumption is technical. In the classical Edgeworth expansion for sums of i.i.d. variables it corresponds to the assumption that $\mathbb{Z}$ is a minimal lattice containing the support of the step.

Assumption A4: For every compact set $K \subset (\beta_-, \beta_+)$, every $a > 0$ and $r \in \mathbb{N}_0$, we have

$$\sup_{\beta \in K} \left[e^{-\varphi(\beta) w_n} \int_a^{\pi} \left| \sum_{k \in \mathbb{Z}} \mathbb{L}_n(k) e^{k(\beta + iu)} \right| \mathrm{d}u \right] = o(w_n^{-r}) \quad \text{a.s.}$$

as $n \to \infty$.

Bell polynomials

The complete Bell polynomials $B_j(z_1, \ldots, z_j)$ are defined by the formal identity

$$\exp\left\{\sum_{j=1}^{\infty} \frac{x^j}{j!} z_j\right\} = \sum_{j=0}^{\infty} \frac{x^j}{j!} B_j(z_1, \ldots, z_j).$$

Therefore, $B_0 = 1$ and for $j \in \mathbb{N}$,

$$B_j(z_1, \ldots, z_j) = \sum{}' \frac{j!}{i_1! \ldots i_j!} \left(\frac{z_1}{1!}\right)^{i_1} \ldots \left(\frac{z_j}{j!}\right)^{i_j},$$

where the sum $\sum{}'$ is taken over all $i_1, \ldots, i_j \in \mathbb{N}_0$ such that $1i_1 + 2i_2 + \ldots + ji_j = j$.

Bell polynomials

The complete Bell polynomials $B_j(z_1, \ldots, z_j)$ are defined by the formal identity

$$\exp\left\{\sum_{j=1}^{\infty} \frac{x^j}{j!} z_j\right\} = \sum_{j=0}^{\infty} \frac{x^j}{j!} B_j(z_1, \ldots, z_j).$$

Therefore, $B_0 = 1$ and for $j \in \mathbb{N}$,

$$B_j(z_1, \ldots, z_j) = \sum{}' \frac{j!}{i_1! \ldots i_j!} \left(\frac{z_1}{1!}\right)^{i_1} \ldots \left(\frac{z_j}{j!}\right)^{i_j},$$

where the sum $\sum{}'$ is taken over all $i_1, \ldots, i_j \in \mathbb{N}_0$ such that $1i_1 + 2i_2 + \ldots + ji_j = j$.

The first few Bell polynomials:

$$B_1(z_1) = z_1, \quad B_2(z_1, z_2) = z_1^2 + z_2 \quad B_3(z_1, z_2, z_3) = z_1^3 + 3z_1 z_2 + z_3.$$

Hermite polynomials

The "probabilist" Hermite polynomial $\mathrm{He}_n(x)$ of degree n is defined by

$$\mathrm{He}_n(x) = e^{\frac{1}{2}x^2}\left(-\frac{\mathrm{d}}{\mathrm{d}x}\right)^n e^{-\frac{1}{2}x^2}.$$

The first few Hermite polynomials are:

$$\mathrm{He}_1(x) = x, \quad \mathrm{He}_2(x) = x^2 - 1, \quad \mathrm{He}_3(x) = x^3 - 3x,$$
$$\mathrm{He}_4(x) = x^4 - 6x^2 + 3, \quad \mathrm{He}_6(x) = x^6 - 15x^4 + 45x^2 - 15.$$

The main theorem

Consider a random profile $\mathbb{L}_1, \mathbb{L}_2, \ldots$ which satisfies assumptions A1–A4.

Consider the tilted profile $k \mapsto e^{\beta k - \varphi(\beta) w_n} \mathbb{L}_n(k)$ and define its "mean" $\mu(\beta)$ and the "standard deviation" $\sigma(\beta)$:

$$\mu(\beta) = \varphi'(\beta), \quad \sigma^2(\beta) = \varphi''(\beta),$$

and "the normalized coordinates"

$$x_n(k) = x_n(k; \beta) = \frac{k - \mu(\beta) w_n}{\sigma(\beta)\sqrt{w_n}}, \quad k \in \mathbb{Z}.$$

The main theorem

Consider a random profile $\mathbb{L}_1, \mathbb{L}_2, \ldots$ which satisfies assumptions A1–A4.

Consider the tilted profile $k \mapsto e^{\beta k - \varphi(\beta) w_n} \mathbb{L}_n(k)$ and define its "mean" $\mu(\beta)$ and the "standard deviation" $\sigma(\beta)$:

$$\mu(\beta) = \varphi'(\beta), \quad \sigma^2(\beta) = \varphi''(\beta),$$

and "the normalized coordinates"

$$x_n(k) = x_n(k;\beta) = \frac{k - \mu(\beta) w_n}{\sigma(\beta)\sqrt{w_n}}, \quad k \in \mathbb{Z}.$$

Define the "deterministic cumulants" $\kappa_j(\beta)$ and the "random cumulants" $\chi_j(\beta)$ as follows:

$$\kappa_j(\beta) = \varphi^{(j)}(\beta), \quad \chi_j(\beta) = (\log W_\infty)^{(j)}(\beta).$$

General Edgeworth expansion

Theorem

Let $\mathbb{L}_1, \mathbb{L}_2, \dots$ *be a random profile satisfying assumptions A1–A4. Fix* $r \in \mathbb{N}_0$ *and a compact set* $K \subset (\beta_-, \beta_+)$*. Then*

$$w_n^{\frac{r+1}{2}} \sup_{k \in \mathbb{Z}} \sup_{\beta \in K} \Big| e^{\beta k - \varphi(\beta) w_n} \mathbb{L}_n(k) - \frac{W_\infty(\beta) e^{-\frac{1}{2} x_n^2(k)}}{\sigma(\beta)\sqrt{2\pi w_n}} \sum_{j=0}^{r} \frac{G_j(x_n(k);\beta)}{w_n^{j/2}} \Big| \overset{a.s.}{\to} 0,$$

where $G_j(x) = G_j(x;\beta)$, $j \in \mathbb{N}_0$ *is a polynomial of degree* $3j$, *defined by*

$$G_j(x) = \frac{(-1)^j}{j!} e^{\frac{1}{2}x^2} B_j(D_1, \dots, D_j) e^{-\frac{1}{2}x^2}.$$

Here B_j *is* j*-th Bell polynomial and* $D_1, D_2, \dots$ *are linear differential operators with random coefficients:*

$$D_j = \frac{\varphi^{(j+2)}(\beta)}{(j+1)(j+2)} \left(\frac{1}{\sigma(\beta)} \frac{\mathrm{d}}{\mathrm{d}x}\right)^{j+2} + \chi_j(\beta) \left(\frac{1}{\sigma(\beta)} \frac{\mathrm{d}}{\mathrm{d}x}\right)^{j}.$$

General Edgeworth expansion

The first three terms G_0, G_1, G_2:

$$G_0(x) = 1,$$

$$G_1(x) = \frac{\chi_1(\beta)}{\sigma(\beta)} x + \frac{\kappa_3(\beta)}{6\sigma^3(\beta)} \mathrm{He}_3(x),$$

$$G_2(x) = \frac{\chi_1^2(\beta) + \chi_2(\beta)}{2\sigma^2(\beta)} \mathrm{He}_2(x) + \left(\frac{\kappa_4(\beta)}{24\sigma^4(\beta)} + \frac{\kappa_3(\beta)\chi_1(\beta)}{6\sigma^4(\beta)} \right) \mathrm{He}_4(x) + \frac{\kappa_3^2(\beta)}{72\sigma^6(\beta)} \mathrm{He}_6(x),$$

The classical Edgeworth expansion

Let $Z_1, Z_2, \ldots$ be an i.i.d. sequence of random variables with values in $\mathbb{Z}$, mean $\mu := \mathbb{E}Z_1$, variance $\sigma^2 := \mathbb{D}Z_1 \neq 0$ and cumulant

$$\varphi(\beta) := \log \mathbb{E}e^{\beta Z_1},$$

which is finite on some interval (β_-, β_+) which contains 0.

The classical Edgeworth expansion

Let $Z_1, Z_2, \ldots$ be an i.i.d. sequence of random variables with values in $\mathbb{Z}$, mean $\mu := \mathbb{E}Z_1$, variance $\sigma^2 := \mathbb{D}Z_1 \neq 0$ and cumulant

$$\varphi(\beta) := \log \mathbb{E}e^{\beta Z_1},$$

which is finite on some interval (β_-, β_+) which contains 0.

Consider a deterministic profile $\mathbb{L}_n$:

$$\mathbb{L}_n(k) = \mathbb{P}[Z_1 + \ldots + Z_n = k], \quad k \in \mathbb{Z}.$$

Assumptions A1–A3 hold with φ defined above, $w_n = n$ and $W_\infty(\beta) = W_n(\beta) = 1$. In particular, all cumulants χ_k vanish. Assumption A4 holds if the distribution of Z_1 is 1-arithmetic.

The classical Edgeworth expansion

Let $Z_1, Z_2, \ldots$ be an i.i.d. sequence of random variables with values in $\mathbb{Z}$, mean $\mu := \mathbb{E}Z_1$, variance $\sigma^2 := \mathbb{D}Z_1 \neq 0$ and cumulant

$$\varphi(\beta) := \log \mathbb{E}e^{\beta Z_1},$$

which is finite on some interval (β_-, β_+) which contains 0.

Consider a deterministic profile $\mathbb{L}_n$:

$$\mathbb{L}_n(k) = \mathbb{P}[Z_1 + \ldots + Z_n = k], \quad k \in \mathbb{Z}.$$

Assumptions A1–A3 hold with φ defined above, $w_n = n$ and $W_\infty(\beta) = W_n(\beta) = 1$. In particular, all cumulants χ_k vanish. Assumption A4 holds if the distribution of Z_1 is 1-arithmetic.

From the main theorem with $\beta = 0$ we obtain the classical Chebyshev-Edgeworth-Cramer expansion:

$$\lim_{n\to\infty} n^{\frac{r+1}{2}} \sup_{k\in\mathbb{N}} \left| \mathbb{P}[Z_1 + \ldots + Z_n = k] - \frac{e^{-\frac{1}{2}x_n^2(k)}}{\sigma\sqrt{2\pi n}} \sum_{j=0}^{r} \frac{q_j(x_n(k))}{n^{j/2}} \right| = 0,$$

where q_j is a polynomial of degree $3j$ with coefficients expressible in terms of the cumulants

$\kappa_2, \ldots, \kappa_{j+2}$.

Branching random walks

Consider a system of particles on $\mathbb{Z}$ with the following evolution:

- at time 0 there is a single particle at 0;
- at each step *one* particle is chosen uniformly at random among existing particles;
- the chosen particle is replaced by a cluster of particles whose displacements w.r.t. the mother are described by a point process $\zeta = \sum_{i=1}^{N} \delta_{Z_i}$ (where N, the number of particles, is a.s. finite) on $\mathbb{Z}$;
- all random mechanisms involved in this definition are independent.

Branching random walks

Let S_n be the number of particles after n splitting events, and let their positions be $x_{1,n}, \dots, x_{S_n,n}$.

Let us denote by $\mathbb{L}_n(k)$ the number of particles at site $k \in \mathbb{Z}$ after n splitting events:

$$\mathbb{L}_n(k) = \#\{1 \le j \le S_n : x_{j,n} = k\}. \tag{1}$$

The function $k \mapsto \mathbb{L}_n(k)$ is called the *profile* of the one-split BRW.

Profiles of the one-split BRW

Denote by ν_k the expected number of particles at site $k \in \mathbb{Z}$ in the cluster process ζ:

$$\nu_k = \mathbb{E}\zeta(\{k\}) = \mathbb{E}\left[\sum_{i=1}^{N} \mathbb{1}_{\{Z_i=k\}}\right], \quad k \in \mathbb{Z}.$$

Profiles of the one-split BRW

Denote by ν_k the expected number of particles at site $k \in \mathbb{Z}$ in the cluster process ζ:

$$\nu_k = \mathbb{E}\zeta(\{k\}) = \mathbb{E}\left[\sum_{i=1}^{N} \mathbb{1}_{\{Z_i = k\}}\right], \quad k \in \mathbb{Z}.$$

Assumption B1: We have $\nu_k > 0$ for at least one $k \in \mathbb{Z}\backslash\{0\}$.

Profiles of the one-split BRW

Denote by ν_k the expected number of particles at site $k \in \mathbb{Z}$ in the cluster process ζ:

$$\nu_k = \mathbb{E}\zeta(\{k\}) = \mathbb{E}\left[\sum_{i=1}^{N} \mathbb{1}_{\{Z_i = k\}}\right], \quad k \in \mathbb{Z}.$$

Assumption B1: We have $\nu_k > 0$ for at least one $k \in \mathbb{Z} \backslash \{0\}$.

Assumption B2: The cluster point process ζ is a.s. non-empty, and the probability that it has at least 2 particles is positive. In other words, $N \geq 1$ a.s. and $\mathbb{P}[N = 1] \neq 1$.

Profiles of the one-split BRW

Denote by $m(\beta)$ the moment generating function of the intensity of the cluster point process ζ minus 1:

$$m(\beta) = \sum_{k\in\mathbb{Z}} e^{\beta k}\nu_k - 1 = \mathbb{E}\left[\sum_{i=1}^{N} e^{\beta Z_i}\right] - 1.$$

The expected number of particles at time n is $\mathbb{E}S_n = 1 + m(0)n$, where, by Assumption B2,

$$m(0) = \sum_{k\in\mathbb{Z}} \nu_k - 1 = \mathbb{E}N - 1 > 0.$$

Profiles of the one-split BRW

Denote by $m(\beta)$ the moment generating function of the intensity of the cluster point process ζ minus 1:

$$m(\beta) = \sum_{k \in \mathbb{Z}} e^{\beta k} \nu_k - 1 = \mathbb{E}\left[\sum_{i=1}^{N} e^{\beta Z_i}\right] - 1.$$

The expected number of particles at time n is $\mathbb{E}S_n = 1 + m(0)n$, where, by Assumption B2,

$$m(0) = \sum_{k \in \mathbb{Z}} \nu_k - 1 = \mathbb{E}N - 1 > 0.$$

Assumption B3: The function m is finite on some non-empty open interval $\mathcal{I}$ containing 0.

Profiles of the one-split BRW

Denote by $m(\beta)$ the moment generating function of the intensity of the cluster point process ζ minus 1:

$$m(\beta) = \sum_{k\in\mathbb{Z}} e^{\beta k}\nu_k - 1 = \mathbb{E}\left[\sum_{i=1}^{N} e^{\beta Z_i}\right] - 1.$$

The expected number of particles at time n is $\mathbb{E}S_n = 1 + m(0)n$, where, by Assumption B2,

$$m(0) = \sum_{k\in\mathbb{Z}} \nu_k - 1 = \mathbb{E}N - 1 > 0.$$

Assumption B3: The function m is finite on some non-empty open interval $\mathcal{I}$ containing 0.

Assumption B4: ν is not concentrated on any proper additive subgroup of $\mathbb{Z}$. In other words, $\nu(\mathbb{Z}\backslash a\mathbb{Z}) \neq 0$ for all $a \in \{2, 3, \ldots\}$.

Profiles of the one-split BRW

Define the function

$$\varphi(\beta) = \frac{m(\beta)}{m(0)}, \quad \operatorname{Re}\beta \in \mathcal{I}$$

and denote by $(\beta_-, \beta_+) \subset \mathcal{I}$ the open interval on which $\varphi'(\beta)\beta < \varphi(\beta)$:

$$\beta_- = \inf\{\beta \in \mathcal{I} \colon \varphi'(\beta)\beta < \varphi(\beta)\}, \tag{2}$$

$$\beta_+ = \sup\{\beta \in \mathcal{I} \colon \varphi'(\beta)\beta < \varphi(\beta)\}. \tag{3}$$

The interval (β_-, β_+) is non-empty because it contains 0.

Assumption B5: For any $\beta \in (\beta_-, \beta_+)$ there is $\gamma = \gamma(\beta) > 1$ such that

$$\mathbb{E}\left[\left(\sum_{i=1}^{N} e^{\beta Z_i}\right)^{\gamma}\right] < \infty.$$

Profiles of the one-split BRW

The next theorem states that the sequence of the one-split BRW profiles satisfies Assumptions A2 and A3 with $w_n = \log n$. The proof uses an embedding into a continuous-time BRW in conjunction with well-known Biggins' results.

Profiles of the one-split BRW

The next theorem states that the sequence of the one-split BRW profiles satisfies Assumptions A2 and A3 with $w_n = \log n$. The proof uses an embedding into a continuous-time BRW in conjunction with well-known Biggins' results.

Theorem

Under Assumptions B1–B3 and B5, there is an open neighborhood $\mathcal{D}$ of the interval (β_-, β_+) in the complex plane such that, with probability 1, W_n converges to some random analytic function W_∞ locally uniformly on $\mathcal{D}$. Moreover, for every compact set $K \subset \mathcal{D}$ and $r \in \mathbb{N}$ we can find an a.s. finite random variable $C_{K,r}$ such that for all $n \in \mathbb{N}$,

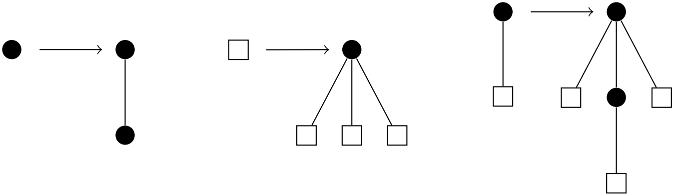
$$\sup_{\beta \in K} |W_n(\beta) - W_\infty(\beta)| < C_{K,r} (\log n)^{-r}.$$

With probability 1, the function W_∞ has no zeros on the interval (β_-, β_+).

Random trees and one-split BRW

There are nodes of two types: the external ones (denoted by □) and the internal ones (denoted by •).

Construction rules for random trees:



Left: RRT. Middle: D-ary recursive with $D = 3$.
Right: PORT.

Random trees and one-split BRW

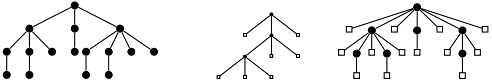


Figure: Sample realizations of random trees. Left: RRT. Middle: D-ary recursive tree with $D = 3$. Right: PORT.

Random trees and the one-split BRW

- particles correspond to (external or internal) nodes;
- positions of particles correspond to the depths of the nodes.

Random trees and the one-split BRW

- particles correspond to (external or internal) nodes;
- positions of particles correspond to the depths of the nodes.

Profile of a tree = the number of nodes at a given level

$\mathbb{L}_n(k) =$ the number of nodes at level k in a tree after n steps

Random trees and the one-split BRW

- particles correspond to (external or internal) nodes;
- positions of particles correspond to the depths of the nodes.

Profile of a tree = the number of nodes at a given level

$\mathbb{L}_n(k)$ = the number of nodes at level k in a tree after n steps

B. Chauvin, L. Devroye, M. Drmota, M. Fuchs, H.-K. Hwang, H. Mahmoud, R. Neininger, A. Panholzer etc.

Random trees and the one-split BRW

(i) Binary search tree correspond to the one-split BRW with the deterministic displacement point process $\zeta = 2\delta_1$ because at any step of the construction an external node at depth k is replaced by two new external nodes at depth $k+1$. We have

$$\varphi(\beta) = 2e^{\beta} - 1, \;\; m(0) = 1, \;\; \mu(0) = \sigma^2(0) = \kappa_j(0) = 2, \;\; j \in \mathbb{N}.$$

The constants $\beta_- \approx -1.678$ and $\beta_+ \approx 0.768$ are the solutions of $2e^{\beta}(1-\beta) = 1$.

Random trees and the one-split BRW

(i) Binary search tree correspond to the one-split BRW with the deterministic displacement point process $\zeta = 2\delta_1$ because at any step of the construction an external node at depth k is replaced by two new external nodes at depth $k+1$. We have

$$\varphi(\beta) = 2e^{\beta} - 1, \ \ m(0) = 1, \ \ \mu(0) = \sigma^2(0) = \kappa_j(0) = 2, \ \ j \in \mathbb{N}.$$

The constants $\beta_- \approx -1.678$ and $\beta_+ \approx 0.768$ are the solutions of $2e^{\beta}(1-\beta) = 1$.

(ii) Random recursive trees (RRTs) correspond to the one-split BRW with the deterministic displacement point process $\zeta = \delta_0 + \delta_1$. In particular,

$$\varphi(\beta) = e^{\beta}, \ \ m(0) = 1, \ \ \mu(0) = \sigma^2(0) = \kappa_j(0) = 1, \ \ j \in \mathbb{N}.$$

We have $\beta_- = -\infty$ and $\beta_+ = 1$.

Profiles BST

Theorem

Let $(\mathbb{L}_n(k))_{k\in\mathbb{Z}}$ *be the profile of a binary search tree with* $n+1$ *external nodes. For every* $r \in \mathbb{N}_0$, *we have*

$$(\log n)^{\frac{r+1}{2}} \sup_{k\in\mathbb{Z}} \Bigg| \frac{1}{n}\mathbb{L}_n(k) - \frac{e^{-\frac{(k-2\log n)^2}{4\log n}}}{\sqrt{4\pi \log n}} \sum_{j=0}^{r} G_j\left(\frac{k-2\log n}{\sqrt{2\log n}};0\right)\frac{1}{(\log n)^{j/2}}\Bigg| \overset{a.s.}{\to} 0,$$

where $G_j(x;0)$ *is a polynomial in* x *of degree* $3j$ *whose coefficients can be linearly expressed through the derivatives* $W'_\infty(0),\dots,W^{(j)}_\infty(0)$.

Profiles BST

The *mode* u_n and the *width* M_n of a profile are defined by

$$u_n = \arg\max_{k \in \mathbb{Z}} \mathbb{L}_n(k), \quad M_n = \max_{k \in \mathbb{Z}} \mathbb{L}_n(k).$$

Profiles BST

The *mode* u_n and the *width* M_n of a profile are defined by

$$u_n = \arg\max_{k \in \mathbb{Z}} \mathbb{L}_n(k), \quad M_n = \max_{k \in \mathbb{Z}} \mathbb{L}_n(k).$$

Theorem

There is an a.s. finite random variable K such that for $n > K$, the mode u_n of the BST with $n+1$ external nodes is equal to one of the numbers

$$\lfloor 2 \log n + \chi_1(0) - 1/2 \rfloor \text{ or } \lceil 2 \log n + \chi_1(0) - 1/2 \rceil,$$

where $\lfloor \cdot \rfloor$, $\lceil \cdot \rceil$ denote the floor and the ceiling functions, respectively, and $\chi_1(0) = W'_\infty(0)/W_\infty(0)$.

Profiles BST

The *mode* u_n and the *width* M_n of a profile are defined by

$$u_n = \arg\max_{k \in \mathbb{Z}} \mathbb{L}_n(k), \quad M_n = \max_{k \in \mathbb{Z}} \mathbb{L}_n(k).$$

Theorem

There is an a.s. finite random variable K such that for $n > K$, the mode u_n of the BST with $n+1$ external nodes is equal to one of the numbers

$$\lfloor 2\log n + \chi_1(0) - 1/2 \rfloor \text{ or } \lceil 2\log n + \chi_1(0) - 1/2 \rceil,$$

where $\lfloor \cdot \rfloor$, $\lceil \cdot \rceil$ denote the floor and the ceiling functions, respectively, and $\chi_1(0) = W'_\infty(0)/W_\infty(0)$.

Actually, we can say more about the behavior of the mode. The following statements hold with probability 1:

- there are arbitrarily long intervals of consecutive n's for which the mode u_n is unique and $u_n = \lceil 2\log n + \chi_1(0) - 1/2 \rceil$;
- similarly, there are arbitrarily long intervals of consecutive n's for which u_n is unique and $u_n = \lfloor 2\log n + \chi_1(0) - 1/2 \rfloor$;
- the set of $n \in \mathbb{N}$ such that u_n is the integer closest to $2\log n + \chi_1(0) - 1/2$ has asymptotic density one.

Profiles BST

Theorem

Let M_n be the width of a binary search tree with $n+1$ external nodes. With probability 1, the set of subsequential limits of the sequence

$$\tilde{M}_n := 4\log n\left(1 - \frac{\sqrt{4\pi \log n}\, M_n}{n}\right), \quad n \in \mathbb{N},$$

is the interval

$$[\chi_2(0) - 1/12, \chi_2(0) + 1/6].$$

Set $\theta_n = \min_{k\in\mathbb{Z}} |2\log n + \chi_1(0) - 1/2 - k|$. *Then*

$$\tilde{M}_n - \theta_n^2 \overset{a.s.}{\to} \chi_2(0) - \frac{1}{12}.$$

Thank you for attention!

KABLUCHKO, Z., MARYNYCH, A. AND SULZBACH, H. (2017+). General Edgeworth expansions with applications to profiles of random trees. To appear in *Ann. Appl. Probab.* Preprint at http://www.imstat.org/aap/future_papers.html

KABLUCHKO, Z., MARYNYCH, A. AND SULZBACH, H. (2016). Mode and Edgeworth Expansion for the Ewens Distribution and the Stirling Numbers, *Journal of Integer Sequences*, **19**, Article 16.8.8.