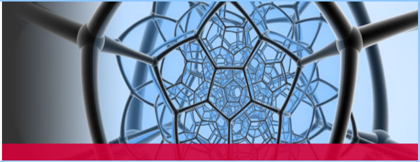
PHASE CHANGES IN RANDOM STRUCTURES AND ALGORITHMS

Hsien-Kuei Hwang

Summer School in Applied Probability

May 21, 2009

Carleton
UNIVERSITY
Canada's Capital University



OUTLINE OF THE LECTURES

1. **Binary search trees, Quicksorts, and phase changes**
2. **Method of moments and its refinements**
3. **Differential equations with polynomial coefficients**
4. **Profiles of random log-trees**

$\sqrt{n}$ vs $\log n$

Typical random trees of height either $\sqrt{n}$ or $\log n$

RANDOM TREES OF HEIGHT $\sqrt{n}$

Examples

- ***Simply generated family of trees*** **(Meir and Moon) or *conditioned Galton-Watson trees*: binary trees, t-ary trees, plane trees, Cayley trees, Motzkin trees, . . .**
- **Non-plane unlabelled trees, non-crossing trees, homeomorphically irreducible trees, free blocky trees, . . .**

RANDOM TREES OF HEIGHT $\log n$

Binomial family

Tries, Patricia tries, digital search trees, bucket digital search trees, etc.

BST (binary search tree) family

BSTs, m-ary STs, median BSTs, quadtrees, simplex trees, gridtrees (of Devroye), etc.

Increasing family

recursive trees, binary increasing trees ($=$ BSTs), plane-oriented recursive trees, etc.

Most of these are Devroye's split trees.

PROFILE OF TREES

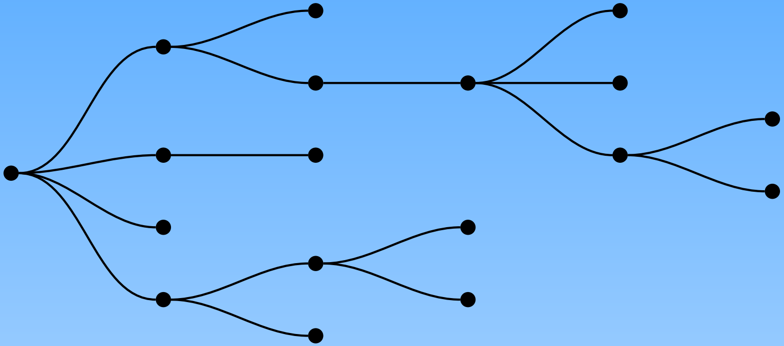


Figure 1: Profile $= \{$**1,4,5,3,3,2**$\}$

MOTIVATIONS

- descendants by generation in branching process
- fine, informative shape characteristic
- related to path length, depth, height, width, etc.
- breadth-first search
- compression algorithms (Jacquet, Szpankowski)
- generation of random trees (Devroye, Robson)
- level-wise analysis of quicksort (Chern, H.)
- parallel quicksort (Evans and Dunbar)

Mathematically interesting

PROFILE OF RANDOM $\sqrt{n}$-TREES

Intensively studied, well understood

Much has been known for such trees: Stepanov, Takács, Aldous, Drmota, Gittenberger, Pitman, Kersting, Janson, Marckert, Bousquet-Mélou, Louchard, . . .

Example: random binary trees

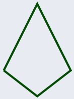
$$\begin{cases} \dfrac{X_{n,k}}{k/4} \xrightarrow{\mathscr{D}} \text{Gamma}(1), & \textbf{if } k \to \infty, k = o(\sqrt{n}), \\ \dfrac{X_{n,k}}{\sqrt{n/8}} \xrightarrow{\mathscr{D}} \text{Stepanov}_\alpha, & \textbf{if } \dfrac{k}{\sqrt{8n}} \to \alpha. \end{cases}$$

A rough picture of the profile of random binary trees

Uniformly for $1 \leq k = o(n^{2/3})$

$$\mathbb{E}(X_{n,k}) \sim \frac{k+2}{2} \exp\left(-\frac{k^2}{4n}\right)$$

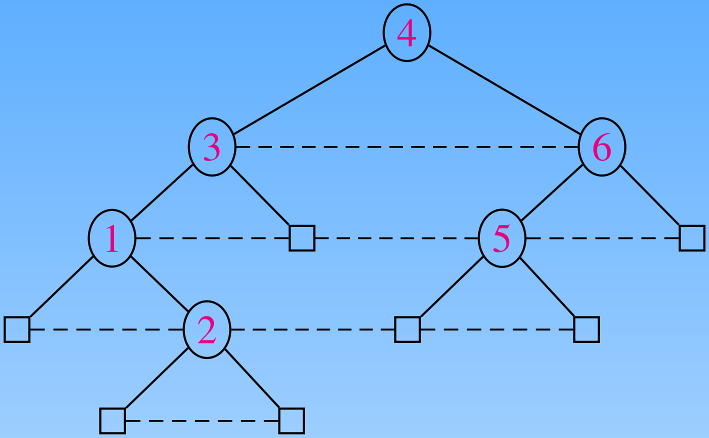
$$\frac{\mathbb{E}(X_{n,k+1})}{\mathbb{E}(X_{n,k})} \sim \frac{k+3}{k+2} \sim 1 \qquad (k = o(\sqrt{n}))$$



While $\begin{cases} \textbf{Expected height} & \sim 2\sqrt{\pi n} \\ \textbf{Expected width} & \sim \sqrt{\pi n/2} \end{cases}$,

$$\mathbb{E}(X_{n,k}) \to \infty \quad \textbf{when} \quad k \leq \sqrt{2n \log n}(1 + o(1)).$$

INTERNAL NODES AND EXTERNAL NODES



PROFILE OF RANDOM BSTs

The probability model

Assume that all permutations of n elements are equally likely. Construct the BST from a random permutation. Call it *random BST*.

$$\begin{cases} X_{n,k} := \text{\# external nodes at distance } k \text{ from the root} \\ I_{n,k} := \text{\# internal nodes at distance } k \text{ from the root} \end{cases}$$

Main questions:

Mean, variance, higher moments, limit distribution of $X_{n,k}$, $I_{n,k}$ for all possible values of k?

Main range: $k \leq K \log n$

Summary of main phenomena

Write throughout $\alpha_{n,k} = \dfrac{k}{\log n}$ **and** $\lim_n \alpha_{n,k} = \alpha$**.**

- $\mathbb{E}(X_{n,k})$ **unimodal, but** $\mathbb{V}(X_{n,k})$ **bimodal**
- **sharp sign-changes of asymptotic correlation coefficient**
- **If** $0.37\cdots < \alpha < 4.31\ldots$**, then** $\dfrac{X_{n,k}}{\mathbb{E}(X_{n,k})} \xrightarrow{\mathscr{D}} X(\alpha)$ **(convergence in distribution).**

Summary of main phenomena

- **If $1 \le \alpha \le 2$, then $\dfrac{X_{n,k}}{\mathbb{E}(X_{n,k})} \xrightarrow{\mathscr{M}} X(\alpha)$ (cv of all moments; best possible range); $X(1) = X(2) \equiv 1$.**

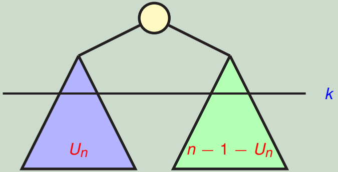
- **If $k \sim j \log n$ and $|k - j \log n| \to \infty$ ($j \in \{1, 2\}$), then $\dfrac{X_{n,k} - \mathbb{E}(X_{n,k})}{\sqrt{\mathbb{V}(X_{n,k})}} \xrightarrow{\mathscr{M}} X'(j)$.**

- **For $k = \begin{matrix} \log n \\ 2 \log n \end{matrix} + O(1)$, $\dfrac{X_{n,k} - \mathbb{E}(X_{n,k})}{\sqrt{\mathbb{V}(X_{n,k})}}$ does not converge to a fixed limit law.**

PROFILE OF RANDOM BSTs

Recurrence

$$\begin{cases} X_{n,0} = \delta_{n,0}, \\ X_{n,k} \stackrel{\mathscr{D}}{=} X_{\text{uniform}[0,n-1],k-1} + X^*_{n-1-\text{uniform}[0,n-1],k-1}; \end{cases}$$



$$\begin{cases} I_{n,0} = 1 - \delta_{n,0}, \\ I_{n,k} \stackrel{\mathscr{D}}{=} I_{\text{uniform}[0,n-1],k-1} + I^*_{n-1-\text{uniform}[0,n-1],k-1}. \end{cases}$$

PROFILE OF RANDOM BSTs

Expected profile: $\mu_{n,k} = \frac{2}{n} \sum_{0 \le j < n} \mu_{j,k-1}$

Mean values known since 1960's: Lynch (1965), Knuth (1998), Brown, Shubert (1984), Mahmoud, Pittel (1984), Pittel (1984), Louchard (1987), Devroye (1988).

$$\begin{aligned} \mathbb{E}(X_{n,k}) &= \frac{2^k}{n!} \text{StirlingFirstKind}(n, k) \\ &= \frac{(2\log n)^k}{\Gamma(\alpha_{n,k}) k! n} \left(1 + O\left(\frac{1}{\log n}\right)\right), \end{aligned}$$

uniformly for $1 \le k = O(\log n)$**, where** $\alpha_{n,k} := k/\log n$**.**

The BST constants

$$\frac{\log \mathbb{E}(X_{n,k})}{\log n} \to \lambda(\alpha) := \alpha - 1 - \alpha \log(\alpha/2).$$

Thus $\mathbb{E}(X_{n,k}) \to \infty$ **when** $\alpha_- < \alpha < \alpha_+$**, where** $0 < \alpha_- < 1 < \alpha_+$ **are the two real zeros of the equation** $z - 1 - z\log(z/2)$ **(**$\alpha_- \approx 0.37$, $\alpha_+ \approx 4.31$**).**

Two implications

The estimate for $\mu_{n,k}$ implies an *LLT for depth* and that *the expected height is bounded above by*

$$\mathbb{E}(H_n) \le \alpha_+ \log n - \frac{\alpha_+}{2(\alpha_+ - 1)} \log\log n + O(1).$$

PROFILE OF RANDOM BSTs

Height and expected profile

Drmota (2003), Reed (2003), improving Devroye's

$$\textbf{Expected height} \sim \alpha_+ \log n - \frac{3\alpha_+}{2(\alpha_+ - 1)} \log\log n,$$

$$\mathbb{E}(X_{n,k}) \to \infty \quad \textbf{when} \quad k \le \alpha_+ \log n - \frac{\alpha_+ + \varepsilon}{2(\alpha_+ - 1)} \log\log n.$$

Random binary trees

Flajolet, Odlyzko (1982)

$$\textbf{Expected height} \sim 2\sqrt{\pi n},$$

$$\mathbb{E}(X_{n,k}) \to \infty \quad \textbf{when} \quad k \le \sqrt{2n \log n}(1 - \varepsilon).$$

Expected number of internal nodes at level k

For simplicity, write $L_n := \log n$.

$$\mathbb{E}(Z_{n,k}) = \frac{2^k}{n!} \sum_{j>k} \mathrm{StirlingFirstKind}(n, j)$$

$$\sim \begin{cases} 2^k - \dfrac{(2L_n)^k}{(1-\alpha_{n,k})\Gamma(\alpha_{n,k})nk!}, & \textbf{if } 1 \le k \le L_n - K\sqrt{L_n}; \\ 2^k \Phi(-t_{n,k}), & \textbf{if } t_{n,k} := \dfrac{k - L_n}{\sqrt{L_n}} = o(L_n^{\frac{1}{6}}); \\ \dfrac{(2L_n)^k}{(\alpha_{n,k}-1)\Gamma(\alpha_{n,k})nk!}, & \textbf{if } L_n + K\sqrt{L_n} \le k \le KL_n, \end{cases}$$

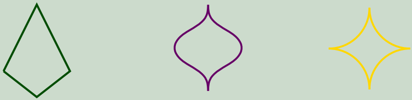
PROFILE OF RANDOM BSTs

A comparison with random binary trees

For external nodes $\frac{\mathbb{E}(X_{n,k+1})}{\mathbb{E}(X_{n,k})} \sim \frac{2\log n}{k+1} \sim \frac{2}{\alpha}$.

For internal nodes $\frac{\mathbb{E}(I_{n,k+1})}{\mathbb{E}(I_{n,k})} \sim \begin{cases} 2, & \textbf{if } \alpha \le 1; \\ \frac{2}{\alpha}, & \textbf{if } \alpha \ge 1. \end{cases}$

For random binary trees, the ratio is asymptotic to 1**.**



Almost all nodes lie at the levels $2\log n + O(\sqrt{\log n})$

(each level having $n/\sqrt{\log n}$ **nodes)**

The general underlying recurrence

Consider

$$a_{n,k} = \frac{2}{n} \sum_{0 \le j < n} a_{j,k-1} + b_{n,k}.$$

Let $a_n(t) := \sum_k a_{n,k} t^k$ **and** $b_n(t) := \sum_k b_{n,k} t^k$**. Then**

$$a_n(t) = \frac{2t}{n} \sum_{0 \le j < n} a_j(t) + b_n(t).$$

Solving it as before and taking coefficients of t^k**, we get**

$$a_{n,k} = b_{n,k} + \frac{2}{n} \sum_{0 \le j < n} \sum_{0 \le r < k} b_{j,k-1-r} [t^r] \prod_{j < \ell < n} \left(1 + \frac{2t}{\ell}\right),$$

where $b_{0,k} := a_{0,k}$**.**

SPECIAL CASES: EXPECTED EXTERNAL AND INTERNAL PROFILES

Expected external profiles: $b_{n,0} = \delta_{n,0}$

$$\mathbb{E}(X_{n,k}) = \frac{2}{n}[t^{k-1}] \prod_{1\le \ell<n} \left(1 + \frac{2t}{\ell}\right)$$
$$= \frac{2^k}{n!} \,\text{StirlingFirstKind}(n, k).$$

Expected internal profiles: $b_{n,0} = 1$ for $n \ge 1$

$$\mathbb{E}(I_{n,k}) = \frac{2}{n}[t^{k-1}] \sum_{1\le j<n} \prod_{j<\ell<n} \left(1 + \frac{2t}{\ell}\right)$$
$$= 2^{k-1}[t^{k-1}] \frac{1}{t-1} \left(\frac{\Gamma(n+t)}{\Gamma(n+1)\Gamma(t+1)} - 1\right).$$

Second moment of $X_{n,k}$

Pittel (1984) derived the expression

$$\mathbb{E}(X_{n,k}^2) = \frac{2^k}{n!} \sum_{1\le t\le n} \frac{1}{(2\pi i)^2} \oint\oint \frac{(\sqrt{8}x/y-1)^{\overline{t-1}}(x^2+t)^{\overline{n-t}}}{yx^{2k-1}\sqrt{1-y^2}}\,\mathrm{d}x\mathrm{d}y,$$

and then showed that for $2-\sqrt{2} \le \alpha \le 2+\sqrt{2}$

$$\mathbb{E}(X_{n,k}^2) = O((\log n)^{3/2} n^{2(\alpha-\alpha\log(\alpha/2)-1)}).$$

Correlation coefficient of $X_{n,k}$ and $X_{n,\ell}$

For $\alpha, \beta \in (2-\sqrt{2}, 2+\sqrt{2})$**, the *correlation coefficient*** $\rho(X_{n,k}, X_{n,\ell})$ **is asymptotic to**

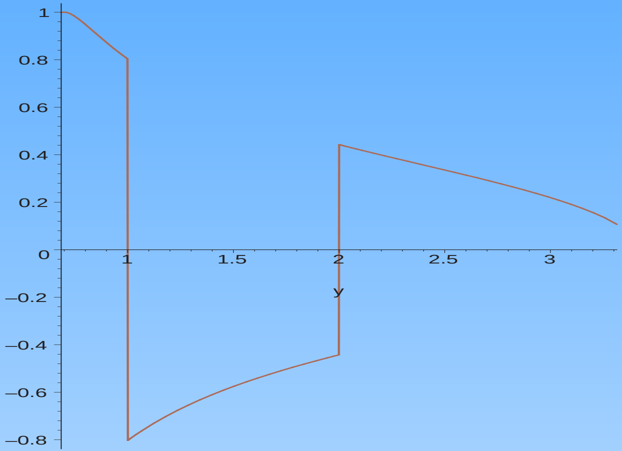
$$\begin{cases} \dfrac{f(\alpha,\beta)}{\sqrt{f(\alpha,\alpha)f(\beta,\beta)}}, & \textbf{if } \alpha,\beta \notin \{1,2\}; \\ \dfrac{f_y(\alpha,\beta)s_{n,\ell} - \frac{1}{2}f_{y^2}(\alpha,\beta)}{\sqrt{f(\alpha,\alpha)p(\beta,\beta;s_{n,\ell},s_{n,\ell})}}, & \textbf{if } \alpha \notin \{1,2\}, \beta \in \{1,2\}; \\ \dfrac{p(\alpha,\beta;t_{n,k},s_{n,\ell})}{\sqrt{p(\alpha,\alpha;t_{n,k},t_{n,k})p(\beta,\beta;s_{n,\ell},s_{n,\ell})}}, & \textbf{if } \alpha,\beta \in \{1,2\}. \end{cases}$$

where $\beta := \lim_n \ell/\log n$, $s_{n,\ell} := \ell - \beta \log n$, $t_{n,k} := k - \alpha \log n$,

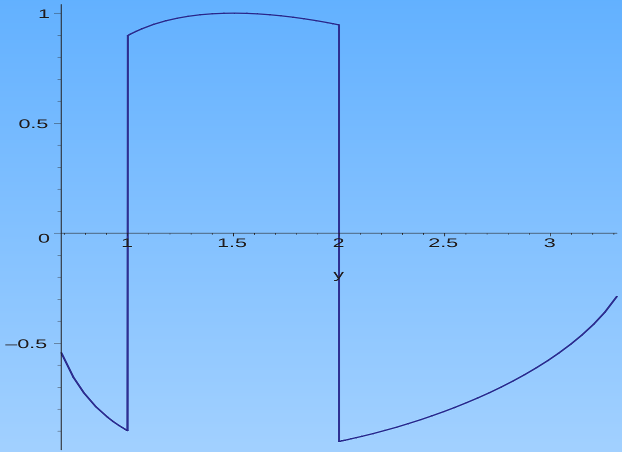
$$f(x,y) := \frac{xy}{(2x+2y-xy-2)\Gamma(x+y-1)} - \frac{1}{\Gamma(x)\Gamma(y)}$$

$$p(j,h;s,t) := f_{xy}(j,h)st - \frac{1}{2}\left(jf_{x^2y}(j,h)t + hf_{xy^2}(j,h)s\right) + \frac{jh}{4}f_{x^2y^2}(j,h).$$

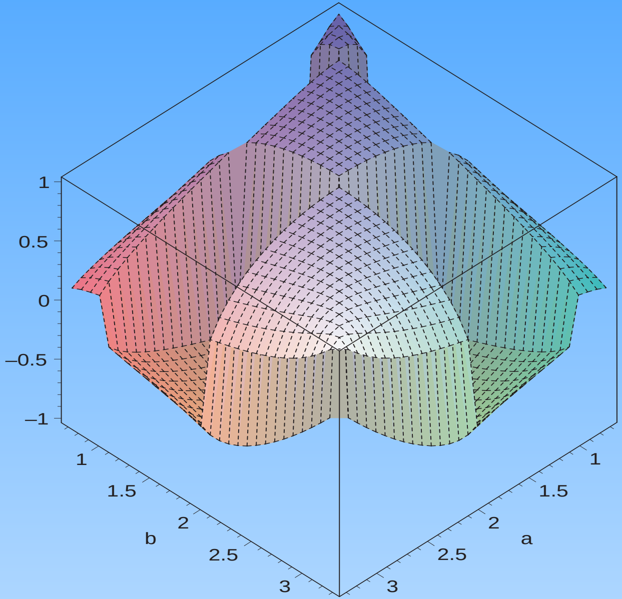
$\rho(X_{n,k}, X_{n,\ell})$ WHEN $\alpha = 0.7$



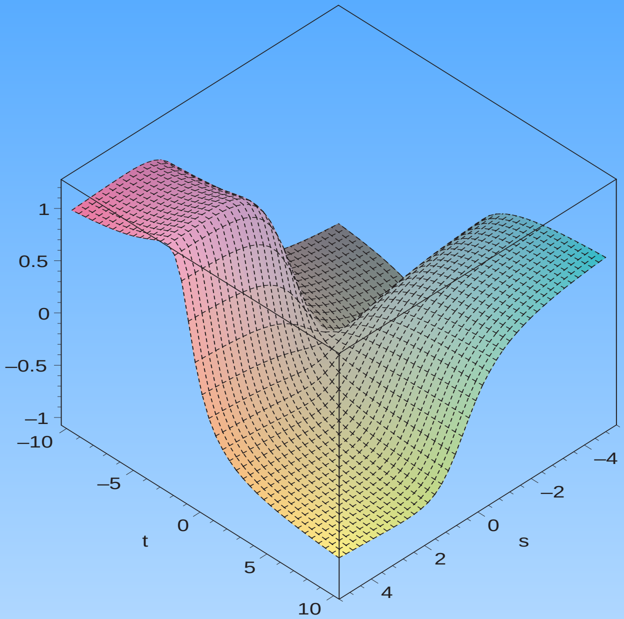
$\rho(X_{n,k}, X_{n,\ell})$ WHEN $\alpha = 1.5$



A 3-D PLOT FOR $\rho(X_{n,k}, X_{n,\ell})$



A 3-D PLOT FOR ρ WHEN $\alpha = 1$ AND $\beta = 2$



SECOND FACTORIAL MOMENT OF $X_{n,k}$

Let $M_k(z) = \sum_n \mathbb{E}(X_{n,k}) z^n$ and
$S_k(z) = \sum_n \mathbb{E}(X_{n,k}(X_{n,k} - 1)) z^n$

$$M_k(z) = \frac{2^k}{n!} \log^k \frac{1}{1-z},$$

and

$$S'_{k+1}(z) = \frac{2}{1-z} S_k(z) + 2M_k(z)^2.$$

Let $F(z, w) := \sum_n S_k(z) w^k$**. Then**

$$F(z,w) = 2w(1-z)^{-2w} \int_0^z (1-t)^{2w} \underbrace{\sum_{k\geq 0} \frac{w^k}{k!k!} \log^{2k} \frac{1}{1-t}}_{\text{modified Bessel function}} \, dt.$$

PROFILE OF RANDOM BSTs

Special cases

If $\alpha = \beta \in \{1, 2\}$ **and** $|k - \alpha \log n|, |\ell - \beta \log n| \to \infty$**, then** $\rho(X_{n,k}, X_{n,\ell}) \sim 1$**.**

Width

Chauvin, Drmota, Jabbour-Hattab (2001):

$W_n \sim \dfrac{n}{\sqrt{4\pi \log n}}$ **almost surely.**

Devroye and H. (2006)

$$\mathbb{E}(W_n) \sim \frac{n}{\sqrt{4\pi \log n}}$$

$$\mathbb{E}(|W_n - \mathbb{E}(W_n)|^s) = O(n^s (\log n)^{-3s/2}).$$

Variance of $X_{n,k}$: middle range

Uniformly for $\alpha \in (2-\sqrt{2}, 2+\sqrt{2})$

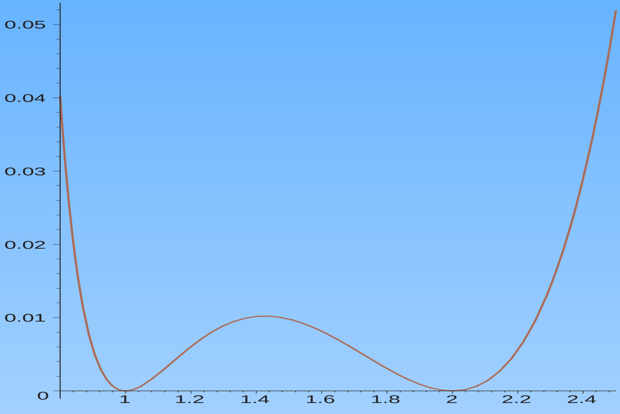
$$\mathbb{V}(X_{n,k}) \sim \phi(\alpha)\left(\mathbb{E}(X_{n,k})\right)^2,$$

where

$$\phi(x) := f(x,x)\Gamma(x)^2 = \frac{\Gamma(x+1)^2}{(4x-x^2-2)\Gamma(2x-1)} - 1.$$

A full asymptotic expansion can be derived.

$\phi(1) = \phi'(1) = \phi(2) = \phi'(2) = 0$



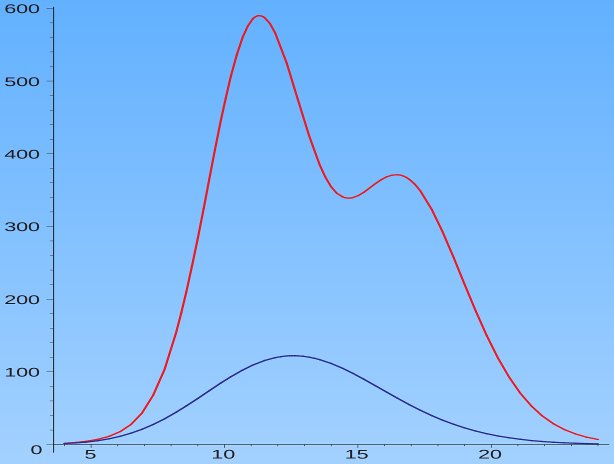
More precise estimates for $\alpha = 1, 2$

$$\mathbb{V}(X_{n,k}) \sim \frac{p(\alpha, \alpha; t_{n,k}, t_{n,k})}{(\log n)^2} \left(\frac{(2 \log n)^k}{k!} \right)^2$$

$$t_{n,k} := k - \alpha \log n$$

p **is a quadratic polynomial in** $t_{n,k}$

$\mathbb{E}(X_{1000,k})$ AND $\mathbb{V}(X_{1000,k})$



Correlation coefficient of $I_{n,k}$ and $I_{n,\ell}$

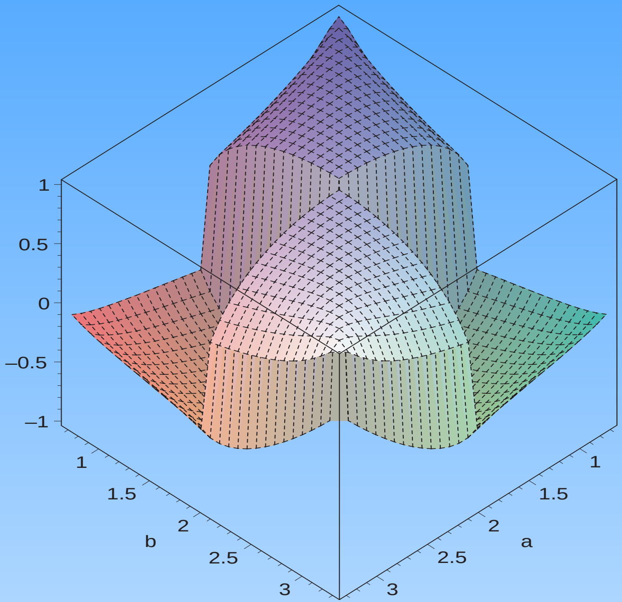
For $\alpha, \beta \in (2-\sqrt{2}, 2+\sqrt{2})$**, the *correlation coefficient*** $\rho(I_{n,k}, I_{n,\ell})$ **is asymptotic to**

$$\begin{cases} \dfrac{\bar{f}(\alpha,\beta)}{\sqrt{\bar{f}(\alpha,\alpha)\bar{f}f(\beta,\beta)}}, & \text{if } \alpha,\beta \notin \{2\}; \\ \dfrac{\bar{f}_\beta(\alpha,2)x - \frac{1}{2}\bar{f}_{\beta^2}(\alpha,2)}{\sqrt{\bar{f}(\alpha,\alpha)p(2,2;s_{n,\ell},s_{n,\ell})}}, & \text{if } \alpha \neq 2, \beta = 2; \\ \dfrac{p(\alpha,\beta;t_{n,k},s_{n,\ell})}{\sqrt{p(\alpha,\alpha;t_{n,k},t_{n,k})p(\beta,\beta;s_{n,\ell},s_{n,\ell})}}, & \text{if } \alpha = \beta = 2. \end{cases}$$

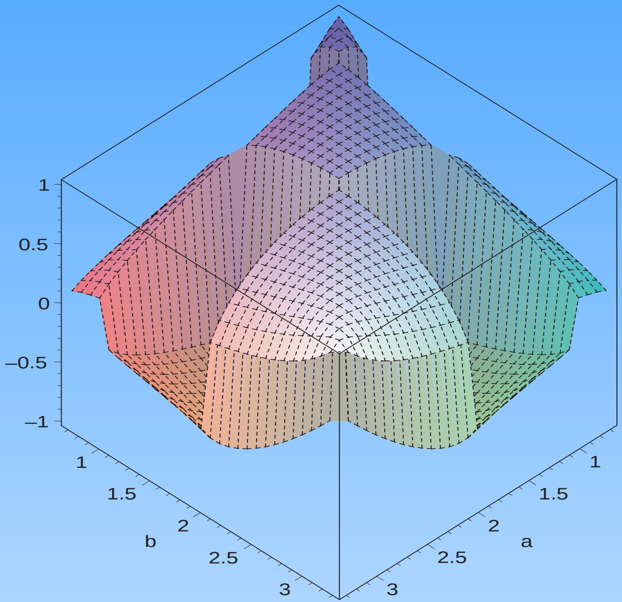
where $\bar{f}(x,y) := f(x,y)/(1-x)/(1-y)$**.**

If $\alpha = \beta = 1$**, then** $\rho(I_{n,k}, I_{n,\ell}) \sim 1$**.**

A 3-D PLOT FOR $\rho(I_{n,k}, I_{n,\ell})$



A 3-D PLOT FOR $\rho(X_{n,k}, X_{n,\ell})$



Limit distributions

Chauvin, Drmota, Jabbour-Hattab (2001): ***almost sure convergence*** **of**

$$\frac{X_{n,k}}{\mathbb{E}(X_{n,k})}, \frac{I_{n,k}}{\mathbb{E}(I_{n,k})} \longrightarrow X(\alpha),$$

for $1.2 \le \alpha \le 2.8$**, where**

$$X(\alpha) \stackrel{\mathscr{D}}{=} \frac{\alpha}{2} U^{\alpha-1} X(\alpha) + \frac{\alpha}{2} (1-U)^{\alpha-1} X(\alpha)^*.$$

Chauvin, Marckert, Klein, Rouault (2005): ***almost sure convergence*** **of** $\frac{X_{n,k}}{\mathbb{E}(X_{n,k})}$ **for** $\alpha_- < \alpha < \alpha_+$**.**

Cv in distribution and cv of all moments

Define $\bar{I}_{n,k} := \begin{cases} 2^k - I_{n,k}, & \text{if } \alpha < 1; \\ I_{n,k}, & \text{if } \alpha \geq 1. \end{cases}$

If $k \sim \alpha \log n$**, where** $\alpha_- < \alpha < \alpha_+$**, then**

$$\frac{X_{n,k}}{\mathbb{E}(X_{n,k})}, \frac{\bar{I}_{n,k}}{\mathbb{E}(\bar{I}_{n,k})} \xrightarrow{\mathscr{D}} X(\alpha),$$

with *convergence of all moments* for $\alpha \in [1,2]$ **but not for** α **outside** $[1,2]$**.**

Moments of the limit law

$\eta_0 = \eta_1 = 1$ **and for** $m \geq 2$

$$\eta_m = \frac{(\alpha/2)^m}{m(\alpha-1)+1-2(\alpha/2)^m} \times \sum_{1\leq j<m} \binom{m}{j} \eta_j \eta_{m-j} \frac{\Gamma(j(\alpha-1)+1)\Gamma((m-j)(\alpha-1)+1)}{\Gamma(m(\alpha-1)+1)}.$$

The polynomial $m(z-1)+1-2(z/2)^m$ **has two positive zeros** z_m^- **and** z_m^+**, where** $z_m^- \uparrow 1$**, and** $z_m^+ \downarrow 2$**.**

Two degenerate cases

$X(1) = X(2) \equiv 1$

The quicksort limit law when $\alpha = 2$

If $k = 2\log n + t_{n,k}$**, where** $t_{n,k} = o(\log n)$ **and** $t_{n,k} \to \infty$**, then**

$$\frac{X_{n,k} - \mathbb{E}(X_{n,k})}{t_n n^{\lambda(\alpha_{n,k})}/\sqrt{4\pi(\log n)^3}}, \frac{I_{n,k} - \mathbb{E}(I_{n,k})}{t_{n,k} n^{\lambda(\alpha_{n,k})}/\sqrt{4\pi(\log n)^3}} \xrightarrow{\mathscr{M}} X'(2).$$

The limit law $X'(2)$ **is essentially the quicksort limit law**

$$X'(2) \stackrel{\mathscr{D}}{=} UX'(2) + (1-U)X'(2)^* + \frac{1}{2} + U\log U + (1-U)\log(1-U).$$

(same law as total path length)

PROFILE OF RANDOM BSTs

No fixed limit law

If $k = 2\log n + O(1)$, then neither of the sequence

$$\left\{ \frac{X_{n,k} - \mathbb{E}(X_{n,k})}{\sqrt{\mathbb{V}(X_{n,k})}}, \frac{I_{n,k} - \mathbb{E}(I_{n,k})}{\sqrt{\mathbb{V}(I_{n,k})}} \right\}$$

converges to a fixed limit law.

Main reason: periodicity ($t_{n,k} = k - \lfloor 2\log n \rfloor - \{2\log n\}$)

$$\mathbb{E}(X_{n,k} - \mathbb{E}(X_{n,k}))^m \sim \underbrace{\text{Polynomial}}_{\text{degree}=m}(t_{n,k}) \left(\frac{(2\log n)^{k-1}}{k!} \right)^m .$$

The range $\alpha = 1$

If $k = \log n + t_{n,k}$, where $t_{n,k} = o(\log n)$ and $t_n \to \infty$, then

$$\frac{X_{n,k} - \mathbb{E}(X_{n,k})}{t_{n,k} n^{\lambda(\alpha_{n,k})} / \sqrt{2\pi (\log n)^3}} \xrightarrow{\mathscr{M}} X'(1).$$

$$X'(1) \stackrel{\mathscr{D}}{=} \frac{1}{2}X'(1) + \frac{1}{2}X'(1)^* + 1 + \frac{1}{2}\log U + \frac{1}{2}\log(1-U).$$

(same as the limit law of $\sum_{j\ge 0} X_{n,j}/2^j$).

If $t_{n,k} = O(1)$, then $\frac{X_{n,k} - \mathbb{E}(X_{n,k})}{\sqrt{\mathbb{V}(X_{n,k})}}$ does not converge to a fixed limit law.

Different behavior for internal nodes

For internal nodes, if $k = \log n + t_{n,k}$**, then, uniformly for** $t_{n,k} = o(\log n)$**,**

$$\frac{I_{n,k} - \mathbb{E}(I_{n,k})}{n^{\lambda(\alpha_{n,k})}/\sqrt{2\pi \log n}} \xrightarrow{\mathscr{M}} X'(1).$$

The normalizing standard variation differs from that of external nodes by a factor of $t_{n,k}/\log n$**.**

No normal limit law for BST-profile

Approaches used

Most proofs rely on handling the double-indexed recurrence

$$a_{n,k} = \frac{2}{n} \sum_{0 \le j < n} a_{j,k-1} + b_{n,k},$$

because all moments (centered or not) satisfy the same recurrence with different $b_{n,k}$.

Develop *asymptotic transfer*; then apply *contraction method* and *the method of moments*.

Functional limit theorems are derived by Drmota et al. (2008).

THE BST-PROFILE PHENOMENA

Universality

For profiles of random trees in the BST family

1. **bimodality of variance near the central range**
2. **sharp sign-changes of correlation coefficient**
3. **cv in distribution in the range when mean** $\to \infty$
4. **cv of all moments in some smaller range, say** $[\alpha_1, \alpha_2]$
5. **convergence of all moments for** $\frac{X_{n,k} - \mathbb{E}(X_{n,k})}{\sqrt{\mathbb{V}(X_{n,k})}}$ **to** $X'(\alpha_1), X'(\alpha_2)$ **when** $\alpha = \alpha_1, \alpha_2$**, resp.**
6. **no fixed limit law when** $k = \alpha_2 \log n + O(1)$

Technicalities more involved

OPEN QUESTIONS

Many questions than answers

- **What happens at the boundary $\alpha = \alpha_-, \alpha_+$?**
- **Are there good process approximations?**
- **How to prove almost-sure convergence in general?**
- **Depth-first search process?**
- **More “humps” in the central range for higher moments?**
- **Asymptotics of central moments outside the middle range?**
- **How to plot or simulate the limit law?**
- **More general theory? and physical connections?**
- $\log^2 n$**-trees?**