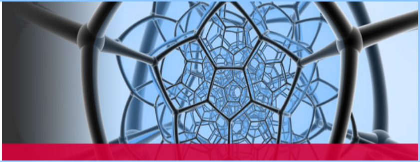
PHASE CHANGES IN RANDOM STRUCTURES AND ALGORITHMS

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**

RANDOM BSTs: THE UNDERLYING RECURRENCE

All moments satisfy the recurrence

$$a_n = b_n + \frac{2}{n} \sum_{0 \le j < n} a_j \qquad (n \ge 1).$$

Exact solution by elementary means

Assume $a_0 = 0$.

$$a_n = b_n + 2(n+1) \sum_{1 \le j < n} \frac{b_j}{(j+1)(j+2)} \qquad (n \ge 1).$$

THE DIFFERENTIAL EQUATION

The generating function $f(z) = \sum_n a_n z^n$ satisfies the DE

$$f'(z) = g'(z) + \frac{2}{1-z} f(z),$$

where $g(z) = \sum_n b_n z^n$**.**

Exact solution when $f(0) = 0$

$$f(z) = (1-z)^{-2} \int_0^z (1-t)^2 g'(t) \mathbf{d}t.$$

Everything is easy!!

A GENERAL FORM

A useful differential operator

Let $\theta := (1-z)\mathbb{D}$. Then the DE

$$f'(z) = g'(z) + \frac{2}{1-z} f(z)$$

can be written as $(\theta - 2)f = (1-z)g'$.

DEs of *Cauchy-Euler* type

Many DEs arising in AofA are of the form

$$\textbf{Polynomial}(\theta) f = g.$$

THE θ-OPERATOR

Properties

Let $\theta := (1-z)\mathbb{D}$**. Then**

$$(1-z)^j\mathbb{D}^j = \theta(\theta+1)\cdots(\theta+j-1) =: \theta^{\bar{j}} \qquad (j \geq 1).$$

So the DE $(1-z)^r f^{(r)} = \sum_{0\leq j<r} c_j(1-z)^j f^{(j)} + g$ **can be expressed as**

$$\underbrace{\left(\theta^{\bar{r}} - \sum_{0\leq j<r} c_j\theta^{\bar{j}}\right)}_{=:\Lambda(\theta)} f = g.$$

***Asymptotics of** f **depends crucially on the zeros of the indicial equation** $\Lambda(\theta) = 0$.*

A MORE GENERAL FORM

DEs with polynomial coefficients

$$\sum_{0\le j\le m} \textbf{Polynomial}_j(\mathbb{D})f = g.$$

Q: Asymptotics of $[z^n]f(z)$?

Holonomic

A function f is holonomic (or D-finite or P-recursive) if it satisfies a linear homogenous differential equation with polynomial coefficients

$$\sum_{0\le j\le m} \textbf{Polynomial}_j(\mathbb{D})f = 0.$$

EXAMPLES OF DEs WITH POLYNOMIAL COEFFICIENTS

$\theta := (1 - z)\mathbb{D}$; $\theta^{\overline{j}} := \theta(\theta + 1)\cdots(\theta + j - 1)$

- **random *m*-ary search trees:** $\theta^{\overline{m-1}} - m!$
- **random fringe-balanced BSTs (median-of-$(2t + 1)$ quicksort):** $\theta^{\overline{2t+1}} - (2(t + 1))!\theta^{\overline{t}}/(t + 1)!$
- **random generalized *m*-ary search trees (Hennequin's quicksort):** $\theta^{\overline{m(t+1)-1}} - (m(t + 1))!\theta^{\overline{t}}/(t + 1)!$
- **random quadtrees:** $\theta(z\theta)^{d-1} - 2^d$
- **random gridtrees (Devroye):** $\theta^{\overline{m-1}}\left(z^{m-1}\theta^{\overline{m-1}}\right)^{d-1} - m!^d$

EXAMPLES OF DEs WITH POLYNOMIAL COEFFICIENTS

The $\Lambda(\theta)$

- **partial-match queries in random quadtrees:**
  $\theta^{d-1}z^{-1}\theta^{s}z(z-1)-2^{d}z$
- **partial-match queries in random k-d trees:**

$$\left(\theta+\frac{q_k}{z}\right)\cdots\left(\theta+\frac{q_k}{z}\right)-2^{k}\qquad (q_j\in\{0,1\}).$$

- **consecutive records in random permutations**

$$(1-z)\mathbb{D}^{r-1}-(r+(1-z)(y-1))\,\mathbb{D}^{r-2}+(y-1)\sum_{0\le j\le r-3}(z+j+1)\mathbb{D}^{j}$$

- **A huge number of others arise in combinatorics, Calabi-Yau equations, statistical physics, etc.**

THE GENERAL ANALYTIC APPROACH

From an ODE-theoretic viewpoint

In general, ODEs of the form

$$\sum_{0\le j\le m} \textbf{Polynomial}_j(\mathbb{D})f = 0,$$

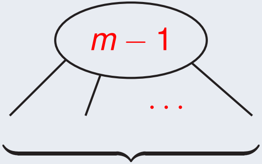
are easy in the sense that dominant singularity, leading order, precise asymptotic behaviors can be readily derived by classical theory (e.g., Frobenius method); in the simplest case, one has

$$f(z) \sim C(\rho - z)^{-\alpha} \qquad (z \sim \rho),$$

then the asymptotics of $[z^n]f(z) = f^{(n)}(0)/n!$ **(the coeff-operator) can be derived by singularity analysis.**

Hard part: an analytic expression for C***?***

RANDOM m-ARY SEARCH TREES



m **branches:** $X_n \stackrel{d}{=} X^{(1)}_{J_1} + \cdots + X^{(m)}_{J_m} + T_n$

Space requirement: $P_n(y) := \mathbb{E}(e^{X_n y})$

$$P_n(y) = \frac{e^y}{\binom{n}{m-1}} \sum_{j_1+\cdots+j_m=n-m+1} P_{j_1}(y)\cdots P_{j_m}(y),$$

so the bivariate GF $P(z,y) := \sum_n P_n(y) z^n$ **satisfies**

$$\frac{\partial^{m-1}}{\partial z^{m-1}} P(z,y) = (m-1)! e^y P^m(z,y).$$

SPACE REQUIREMENT OF RANDOM m-ARY SEARCH TREES

Two explicitly solvable cases

If $m = 2$ (BSTs), then ($X_n \equiv n$)

$$P(z, y) = \frac{1}{1 - e^y z}.$$

If $m = 3$, then

$$z = e^{-y/2} \int_1^{P(z,y)} \frac{\mathrm{d}v}{\sqrt{v^4 + e^y - 1}};$$

(expressible as generalized hypergeometric or Weierstrass's $\wp$ functions).

No closed-form solutions are known for $m \geq 4$.

The phase change

Mahmhoud and Pittel (1989), Lew and Mahmoud (1994), Chern and H. (2001): The space requirement X_n exhibits the phase change: if $3 \leq m \leq 26$, then

$$\frac{X_n - \mu n}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1);$$

if $m \geq 27$, then the sequence of random variables $(X_n - \mathbb{E}(X_n))/\sqrt{\mathbb{V}(X_n)}$ does not converge to a fixed limit law.

For other results, see H. (2003), Janson (2005), Chauvin and Pouyanne (2005), Fill and Kapur (2004, 2005), Dean and Majumdar (2005).

SPACE REQUIREMENT: MEAN

The recurrence $\mu_n := \mathbb{E}(X_n)$

$$\mu_n = 1 + \frac{m}{\binom{n}{m-1}} \sum_{0 \le j < n} \binom{n-1-j}{m-2} \mu_j,$$

The GF $M(z) := \sum_n \mu_n z^n$ satisfies

$$M^{(m-1)}(z) - \frac{m!}{(1-z)^{m-1}} M(z) = \frac{(m-1)!}{(1-z)^m},$$

$$\left(\theta^{\overline{m-1}} - m!\right) M(z) = \frac{(m-1)!}{1-z},$$

with the initial conditions $M(0) = 0$ **and** $M^{(j)}(0) = j!$**,** $1 \le j \le m-2$**.**

Q: asymptotics of μ_n from properties of $M(z)$?

EXACT SOLUTION OF THE DE

Two simple lemmas

- $f(z) = f(0)(1-z)^{-\rho} + (1-z)^{-\rho}\int_0^z (1-x)^{\rho-1} g(x)dx$ **is the solution to the DE**

$$(\vartheta - \rho)f(z) = g(z),$$

with initial value $f(0)$**.**

- $f(z) = c(1-z)^{-\rho} + A\dfrac{(1-z)^{-s}}{s-\rho}$, $s \neq \rho$, **is the general solution to the DE**

$$(\vartheta - \rho)f(z) = A(1-z)^{-s} \qquad (A \in \mathbb{C}).$$

EXACT SOLUTION OF THE DE

A useful consequence

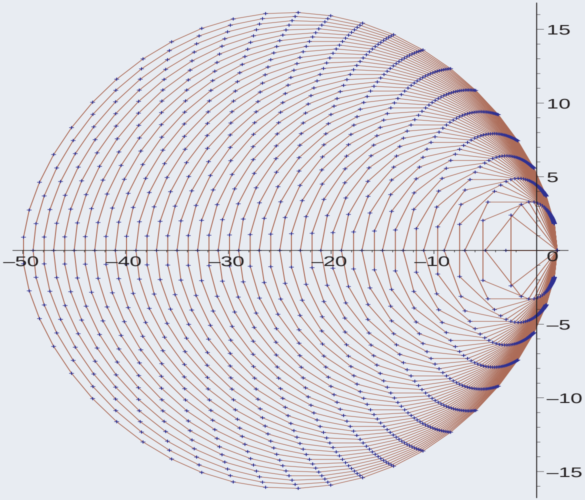
- $f(z) = \dfrac{A(1-z)^{-s}}{(s-\rho_1)\cdots(s-\rho_k)} + \sum_{1\le j\le k} c_j(1-z)^{-\rho_j}$, **where** s **and the** ρ_j**'s are all distinct complex numbers and the** c_j**'s are constants, solves the DE**

$$(\vartheta-\rho_1)\cdots(\vartheta-\rho_k)f(z) = A(1-z)^{-s} \qquad (A\in\mathbb{C}).$$

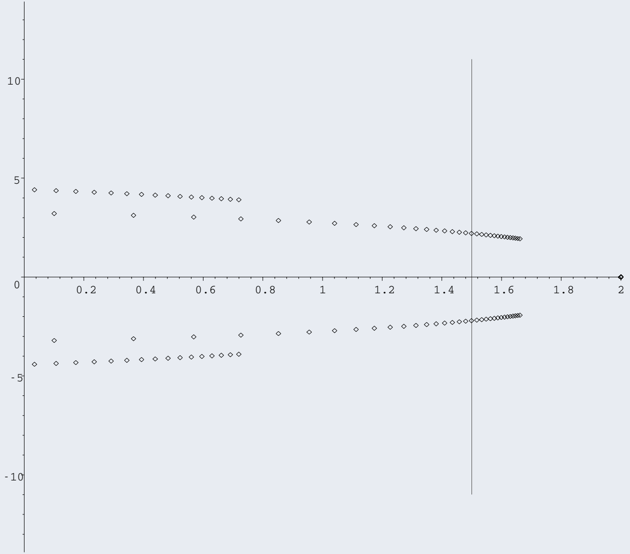
Zeros of the indicial equation $\theta^{\overline{m-1}} - m! = 0$

- $\theta = 2$ **is a zero** ($2^{\overline{m-1}} = m!$)
- **all other zeros are distinct with real part** < 2.

DISTRIBUTION OF THE ZEROS OF $\overline{\theta^{m-1}} - m! = 0$



DISTRIBUTION OF THE ZEROS OF $\overline{\theta^{m-1}} - m! = 0$



EXACT SOLUTION OF THE DE

Let the zeros of $\theta^{\overline{m-1}} - m! = 0$ be λ_j with $\lambda_1 = 2$.

Then $\left(\theta^{\overline{m-1}} - m!\right) M(z) = (m-1)!(1-z)^{-1}$ **becomes**

$$(\theta - 2)(\theta - \lambda_2) \cdots (\theta - \lambda_{m-1}) M(z) = (m-1)!(1-z)^{-1}.$$

Its solution is of the form

$$M(z) = \sum_{1 \le j \le m-1} A_j (1-z)^{-\lambda_j} - \frac{1}{(m-1)(1-z)}.$$

By the initial conditions, we can prove that

$$A_j = \frac{1}{\lambda_j(\lambda_j - 1) \sum_{0 \le \ell \le m-2} \frac{1}{\lambda_j + \ell}} \qquad (1 \le j \le m-2);$$

(see Mahmoud and Pittel, 1989 or Chern and H., 2001)

ASYMPTOTICS OF MEAN

$A_1 = \frac{1}{2(H_m-1)}$, $H_m = \sum_{1\le j\le m} \frac{1}{j}$

$$\mu_n = \sum_{1\le j\le m-1} A_j \binom{\lambda_j + n - 1}{n} - \frac{1}{m-1}$$

$$\sim \frac{n+1}{2(H_m-1)} - \frac{1}{m-1} + \frac{A_2}{\Gamma(\lambda_2)} n^{\lambda_2 - 1} + \frac{A_3}{\Gamma(\lambda_3)} n^{\lambda_3 - 1} + \cdots .$$

where we used the asymptotic relation

$$[z^n](1-z)^{-\alpha} = \binom{n+\alpha-1}{n} = \frac{n^{\alpha-1}}{\Gamma(\alpha)}(1+o(1)),$$

for each finite $\alpha \in \mathbb{C}$ (Flajolet-Odlyzko).

Random m-ary search trees are not space efficient!!

GENERAL DEs?

The success of the above approach relies heavily on the simple form $(1 - z)^{-1}$ of the non-homogeneous part.

How to deal with general non-homogenenous part? Needed for higher moments.

THE GENERAL DEs: AN ANALYTIC APPROACH

A simple analytic scheme by ODE theory

Consider the DE

$$\sum_{0\le j\le k} c_j(1-z)^j f^{(j)}(z) = h(z) \qquad (k\ge 1;\, c_k = 1).$$

Here $h \not\equiv 0$ **is *FO-admissible* with** $h(z)\sim A(1-z)^{-s}$**, as** $z\to 1$**, where** $A, s\in\mathbb{C}$**.**

Let $\rho\in\mathbb{C}$ **be the largest zero (in real part) of the indicial equation** $\Lambda(\theta) := \theta^{\overline{k}} + \sum_{0\le j<k} c_j\theta^{\overline{j}} = 0$**.**

If ρ **is a simple zero and the real parts of all other zeros are** $<\Re(\rho)$**, then for** $z\sim 1$

$$f(z)\sim\begin{cases} \frac{A}{\Lambda(s)}(1-z)^{-s}, & \text{if } \Re(s)>\Re(\rho);\\ C(1-z)^{-\rho}, & \text{if } \Re(s)<\Re(\rho);\\ C(1-z)^{-\rho}+C'(1-z)^{-s}, & \text{if } s\ne\rho \text{ and } \Re(s)=\Re(\rho);\\ -\frac{A}{\Lambda'(\rho)}(1-z)^{-\rho}\log(1-z), & \text{if } s=\rho.\end{cases}$$

ELEMENTARY VS ANALYTIC

In number-theoretic sense

$$\textbf{Elementary} \longrightarrow \textbf{real analysis}$$
$$\textbf{Analytic} \longrightarrow \textbf{complex analysis}.$$

Analytic approach

Operationally easier, resulting expressions simpler, but requires stronger analytic conditions

Elementary approach

Operationally less elegant, but gives stronger result

A hybrid general approach

An appropriate development of the elementary approach, assisted by procedures of analytic approach

ASYMPTOTIC TRANSFER RESULTS

The underlying recurrence

Consider

$$a_n = \frac{m}{\binom{n}{m-1}} \sum_{0\le j<n} \binom{n-1-j}{m-2} a_j + b_n \qquad (n \ge m-1),$$

with the initial conditions $a_n = b_n$ **for** $0 \le n \le m-2$.

Then ($C := \frac{1}{H_m - 1} \sum_{j\ge 0} \frac{b_j}{(j+1)(j+2)}$**)**

$$a_n \sim Cn \textbf{ iff } b_n = o(n) \textbf{ and } \left|\sum_j b_j j^{-2}\right| < \infty,$$

$$a_n \sim cn^\alpha \textbf{ iff } b_n \sim \frac{cn^\alpha}{1 - \frac{m!}{(\alpha+1)\cdots(\alpha+m-1)}} \quad (\alpha > 1).$$

The DE: $\theta^{\overline{m-1}} - m! = (\theta - \lambda_1)\cdots(\theta - \lambda_{m-1})$

$$(\theta - \lambda_1)\cdots(\theta - \lambda_{m-1})f(z) = (1-z)^{m-1}g^{(m-1)}(z),$$

where $f(z) = \sum_n a_n z^n$ **and** $g(z) = \sum_n b_n z^n$.

Idea: successive applications of linear operators

$$(\theta - \lambda_1)\underbrace{(\theta - \lambda_2)\cdots(\theta - \lambda_{m-1})f(z)}_{f_1(z)}$$

$$= (\theta - \lambda_1)f_1(z)$$

So we focus first on DE of the form $(\theta - \rho)f = g$.

THE LINEAR-OPERATOR APPROACH

$(\theta-\rho)f=g$

$f(z)=f(0)(1-z)^{-\rho}+(1-z)^{-\rho}\int_0^z(1-x)^{\rho-1}g(x)dx$ **is the solution to the DE** $(\vartheta-\rho)f(z)=g(z)$ **with initial value** $f(0)$.

An elementary version: $f'=\frac{\rho}{1-z}f+\frac{g}{1-z}$

Let $g(z)=\sum_{n\geq 0}b_nz^n$ **and** $f(z)=\sum_{n\geq 0}a_nz^n$ **be two formal power series. Then the solution to the recurrence**

$$a_n=\frac{\rho}{n}\sum_{0\leq j<n}a_j+\frac{1}{n}\sum_{0\leq j<n}b_j \qquad (n\geq 1),$$

with a_0 **given, satisfies**

$$a_n=a_0\binom{n+\rho-1}{n}+\sum_{0\leq k<n}\frac{b_k}{k+1}\prod_{k+2\leq j\leq n}\left(1+\frac{\rho-1}{j}\right) \qquad (n\geq 0).$$

Formal-power series

Taking coefficient of z^n on both sides

$$f(z) = f(0)(1-z)^{-\rho} + J_\rho[g](z),$$

where $J_\rho[g](z) := (1-z)^{-\rho} \int_0^z (1-x)^{\rho-1} g(x)dx$**, we also obtain**

$$a_n = a_0 \binom{n+\rho-1}{n} + \sum_{0 \le k < n} \frac{b_k}{k+1} \prod_{k+2 \le j \le n} \left(1 + \frac{\rho-1}{j}\right) \qquad (n \ge 0).$$

We take such a formal-power series point of view for all DEs. Advantages: ***expressions neater, manipulation simpler and without having to worry about analytic properties*****.**

THE LINEAR-OPERATOR APPROACH

Simple, fundamental tools

Let $\mathbf{Q}_\tau := \{f(z) : [z^n]f(z) = o(n^\tau)\}$, where $\tau \in \mathbb{R}$.

- **If $f \in \mathbf{Q}_\tau$, then $J_\rho[Q] \in \mathbf{Q}_\tau$ for $\tau > \Re(\rho) - 1$.**
- **If $[z^n]f(z) \sim cn^v \log^\beta n$, where $c \in \mathbb{C}$ and $\Re(v) > \Re(\rho) - 1$, then**

$$[z^n]J_\rho[f](z) \sim \frac{c}{v + 1 - \rho} n^v \log^\beta n.$$

- **If $[z^n]f(z) \sim cn^{\rho-1} \log^\beta n$, where $\beta > -1$, then**

$$[z^n]J_\rho[f](z) \sim \frac{c}{\beta + 1} n^{\rho-1} \log^{\beta+1} n.$$

Idea of proof

Use

$$[z^n]I_\rho[Q](z) = \sum_{0\le k<n} \frac{[z^k]Q(z)}{k+1} \prod_{k+2\le j\le n}\left(1+\frac{\rho-1}{j}\right),$$

and

$$\begin{aligned}\prod_{k+2\le j\le n}\left(1+\frac{\rho-1}{j}\right) &= \exp\left((\rho-1)\sum_{k+2\le j\le n}\frac{1}{j}\right)\prod_{k+2\le j\le n}\left(1+\frac{\rho-1}{j}\right)e^{-(\rho-1)/j} \\ &= O\left(n^{\Re(\rho)-1}(k+1)^{-\Re(\rho)+1}\right).\end{aligned}$$

THE LINEAR-OPERATOR APPROACH

Now consider $(\theta - \lambda_1)f_1 = (1-z)^{m-1}g^{(m-1)}$

The solution is

$$f_1(z) = f_1(0)(1-z)^{-\lambda_2} + J_{\lambda_1}[(1-z)^{m-1}g^{(m-1)}](z),$$

where, by initial conditions, $\lambda_1 = 2$, and some identities $f_1(0) = m! \sum_{0 \le j \le m-2} \frac{b_j}{(j+1)(j+2)}$.

Let $f_2(z) = (\theta - \lambda_3) \cdots (\theta - \lambda_{m-1})f(z)$.

Applying the same procedure

$$f_2(z) = \frac{f_1(0)}{\lambda_1 - \lambda_2}(1-z)^{-\lambda_1} + I_{\lambda_2}[I_{\lambda_1}[(1-z)^{m-1}g^{(m-1)}]](z) + Q_1(z),$$

where $Q_1(z) \in \mathbf{Q}_1$.

Continuing iterating the same procedure

$$f(z) = \frac{f_1(0)(1-z)^{-\lambda_1}}{(\lambda_1-\lambda_2)\cdots(\lambda_1-\lambda_{m-1})} + J_{\lambda_{m-1}}\left[\cdots J_{\lambda_1}\left[(1-z)^{m-1}g^{(m-1)}\right]\cdots\right](z) + Q_2(z),$$

where $Q_2(z) \in \mathbf{Q}_1$**.**

By induction

$$J_{\lambda_{m-1}}\left[\cdots J_{\lambda_1}\left[(1-z)^{m-1}g^{(m-1)}\right]\cdots\right](z) = \frac{m!J_2[g](z)}{(2-\lambda_2)\cdots(2-\lambda_{m-1})} - \sum_{0\le j\le m-2}\frac{m!b_j}{(j+1)(j+2)}(1-z)^{-\lambda_1} + Q_3(z),$$

where $Q_3(z) \in \mathbf{Q}_1$**.**

The final expression

By using $b_n = o(n)$ **and collecting all estimates,**

$$f(z) = \frac{J_2[g](z)}{H_m - 1} + Q_4(z),$$

where $Q_3(z) \in \mathbf{Q}_1$**. Now**

$$[z^n]J_2[g](z) = (n+1) \sum_{0 \le j < n} \frac{b_j}{(j+1)(j+2)}.$$

This proves that $b_n = o(n)$ **and** $|\sum_n b_n n^{-2}| < \infty$ **imply** $a_n \sim Cn$**; the necessity part is easy and as in BST case.**

THE LINEAR-OPERATOR APPROACH

The large toll function case

If $b_n \sim cn^\alpha$, $\alpha > 1$, then the asymptotics of a_n can be easily "guessed" as follows. Assume $a_n \sim Cn^\alpha$. Then

$$\begin{aligned}
a_n &= \frac{m}{\binom{n}{m-1}} \sum_{0 \le j < n} \binom{n-1-j}{m-2} a_j + b_n \\
&\sim \frac{Cm}{\binom{n}{m-1}} \sum_{j<n} \binom{n-1-j}{m-2} j^\alpha + cn^\alpha \\
&\sim Cm! n^{1-m} \sum_{j<n} \frac{(n-1-j)^{m-2}}{(m-2)!} j^\alpha + cn^\alpha \\
&\sim C \frac{\Gamma(\alpha+1)m!}{\Gamma(\alpha+m)} n^\alpha + cn^\alpha \sim Cn^\alpha,
\end{aligned}$$

so that C is solved to be $\frac{c}{1-\frac{m!}{(\alpha+1)\cdots(\alpha+m-1)}}$.

The proof follows the same linear-operator procedure and is much simpler.

SPACE REQUIREMENT: RECURRENCE OF CENTRAL MOMENTS

$$P_n(y) = \frac{e^y}{\binom{n}{m-1}} \sum_{j_1+\cdots+j_m=n-m+1} P_{j_1}(y)\cdots P_{j_m}(y)$$

Let $\overline{P}_n(y) := \mathbb{E}(e^{(X_n-\mu_n)y})$ **and** $P_{n,k} := \mathbb{E}(X_n-\mu_n)^k = \overline{P}_n^{(m)}(0)$. **Then**

$$\overline{P}_n(y) = \sum_{j_1+\cdots+j_m=n-m+1} \overline{P}_{j_1}(y)\cdots\overline{P}_{j_m}(y)e^{\Delta_n(\mathbf{j})y},$$

where $\Delta_n(\mathbf{j}) := 1+\mu_{j_1}+\cdots+\mu_{j_m}-\mu_n$. **It follows that**

$$P_{n,k} = \frac{m}{\binom{n}{m-1}} \sum_{j_1+\cdots+j_m=n-m+1} \binom{m-1-j}{m-2} P_{j,k} + Q_{n,k},$$

where

$$Q_{n,k} := \frac{1}{\binom{n}{m-1}} \sum_{\substack{i_0+\cdots+i_m=k\\ i_1,\ldots,i_m<k}} \binom{k}{i_0,\ldots,i_m} \sum_{j_1+\cdots+j_m=n-m+1} P_{j_1,i_1}\cdots P_{j_m,i_m}\Delta_n(\mathbf{j})^{i_0}.$$

VARIANCE OF THE SPACE REQUIREMENT

Let $\alpha = \Re(\lambda_2)$.

Since $\mu_n - \frac{n}{H_m - 1} \asymp \max\{1, n^{\alpha-1}\}$**,**

$$\begin{aligned}\Delta_n(\mathbf{j}) &= 1 + \mu_{j_1} + \cdots + \mu_{j_m} - \mu_n \\ &\asymp \max\{1, n^{\alpha-1}\}\end{aligned}$$

Numeric values of α

m	3	4	5	6	7	8	9	10
α	-3	-2.5	-1.5	-0.768	-0.266	0.101	0.366	0.568

m	11	12	13	14	15	16	17	18
α	0.726	0.852	0.955	1.041	1.112	1.173	1.226	1.272

m	19	20	21	22	23	24	25	26
α	1.313	1.349	1.381	1.409	1.435	1.459	1.479	1.499

$\alpha - 1 > 1/2$ **for** $m \geq 27$

VARIANCE OF THE SPACE REQUIREMENT

$V_n := \mathbb{V}(X_n) = P_{n,2}$ satisfies

$$V_n = \frac{m}{\binom{n}{m-1}} \sum_{j_1+\cdots+j_m=n-m+1} \binom{m-1-j}{m-2} V_j + Q_{n,2},$$

where

$$Q_{n,2} := \frac{1}{\binom{n}{m-1}} \sum_{j_1+\cdots+j_m=n-m+1} \Delta_n(\mathbf{j})^2$$

$$\asymp \max\{n^{2\alpha-2}, 1\}.$$

Thus

- **for** $3 \le m \le 26$: $Q_{n,2} = o(n)$ **and** $|\sum_n Q_{n,2} n^{-2}| < \infty$, **so** $V_n \sim \sigma^2 n$;
- **for** $m \ge 27$: $n^{2\alpha-2} \gg n^{1+\varepsilon}$, **so** $V_n \asymp n^{2\alpha-2}$.

ASYMPTOTICS OF HIGHER CENTRAL MOMENTS

Asymptotic normality when $3 \leq m \leq 26$

By induction and the asymptotic transfer,

$$P_{n,2r} = \mathbb{E}\,(X_n - \mu_n)^{2r} \sim \frac{(2r)!}{2^r r!}\sigma^{2r} n^r,$$
$$P_{n,2r-1} = \mathbb{E}\,(X_n - \mu_n)^{2r-1} = o(n^{r-1/2}),$$

for $r \geq 1$. This implies by the method of moments the CLT of X_n.

$m \geq 27$: fluctuations dominate

$$P_{n,k} \sim F_k(\Im(\lambda_2)\log n)n^{k(\alpha-1)} \qquad (k \geq 2).$$

Alternative approaches

Contraction method, martingales, statistical physics.

SECOND PHASE CHANGE: CONVERGENCE RATE

The method of moments can be further refined

$$\sup_{-\infty<x<\infty}\left|\mathbb{P}\left(\frac{X_n-\mu_n}{\sigma_n}<x\right)-\Phi(x)\right|$$
$$=\begin{cases} O(n^{-1/2}), & \textbf{if } 3\le m\le 19;\\ O(n^{-3(3/2-\alpha)}), & \textbf{if } 20\le m\le 26.\end{cases}$$

m	α	$3(3/2-\alpha)$
20	**1.34892881**	**0.45321354**
21	**1.38079786**	**0.35760639**
22	**1.40936978**	**0.27189065**
23	**1.43512896**	**0.19461309**
24	**1.45847025**	**0.12458925**
25	**1.47971848**	**0.06084455**
26	**1.49914326**	**0.00257020**

SECOND PHASE CHANGE: LOCAL LIMIT THEOREM

Moderate deviations LLT

$$\mathbb{P}(X_n = \lfloor \mu_n + x\sigma_n \rfloor) = \frac{e^{-x^2/2}}{\sqrt{2\pi}\,\sigma_n} \times \left(1 + \begin{cases} O\left(\dfrac{1+|x|^3}{n^{1/2}}\right), & \text{if } 3 \le m \le 19 \\ O\left(\dfrac{1+|x|^3}{n^{3(3/2-\alpha)}}\right), & \text{if } 20 \le m \le 26 \end{cases}\right)$$

where the first O-term holds uniformly for $x = o(n^{1/6})$, the second for $x = o(n^{-\alpha+3/2})$.

THE REFINE METHOD OF MOMENTS: SKETCH

Major steps: Let $\bar{\alpha} = \begin{cases} 1/3, & \text{if } 3 \le m \le 19; \\ \alpha - 1, & \text{if } 20 \le m \le 26 \end{cases}$

Define $\phi_n(y) := \mathbb{E}(e^{(X_n - \mathscr{N}(\mu_n, \sigma_n^2))y}) = P_n(y)e^{-\mu_n y - \sigma_n^2 y^2/2}$.

- **Prove** $|\phi_n^{(k)}(0)| \le A^k k! n^{k\bar{\alpha}}$ **using the asymptotic transfer:**

$$\text{if } |b_n| \le cn^\beta, \beta > 1, \text{ then } |a_n| \le \frac{Kcn^\beta}{1 - \frac{m!}{(\beta+1)\cdots(\beta+m-1)}}.$$

- **Derive a uniform estimate for** $|P_n(it)|$ **for** $|t| \le \pi$
- **Rate and LLT by standard Fourier analysis**

SPACE REQUIREMENTS

Open questions

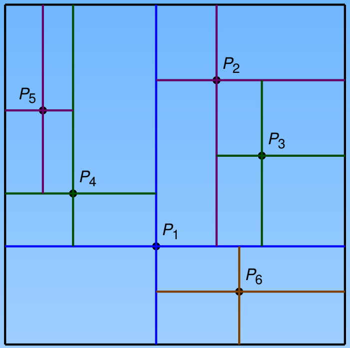
Is there an analytic approach to getting asymptotic expansion for the DE

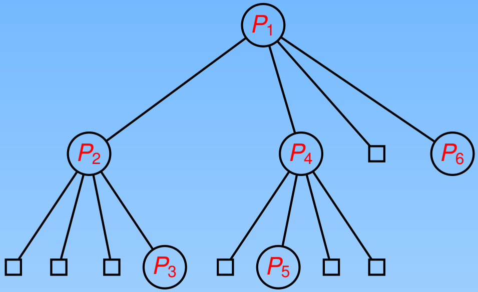
$$\frac{\partial^{m-1}}{\partial z^{m-1}} P(z,y) = (m-1)! y P^m(z,y).$$

for $z, y \sim 1$ **and** $m \geq 4$**? A particular solution is** $(1 - y^{1/(m-1)} z)^{-1}$**.**

This would then have important consequences for, say large deviations.

A 2-DIMENSIONAL POINT QUADTREE





RANDOM d-DIMENSIONAL QUADTREES

Random quadtrees

If the n given points are iid from $[0,1]^d$, then the resulting tree is called a *random quadtree*.

$$X_n \stackrel{d}{=} X^{(1)}_{J_1} + \cdots + X^{(2^d)}_{J_{2^d}} + T_n.$$

The underlying recurrence for moments

All moments of any additive cost measures satisfy the recurrence (Flajolet et al., 1995)

$$a_n = b_n + 2^d \sum_{0\le j<n} \pi_{n,j} a_j \qquad (n \ge 1),$$

where $\pi_{n,j} = \binom{n-1}{j} \int_{[0,1]^d} (x_1 \cdots x_d)^j (1 - x_1 \cdots x_d)^{n-1-j} d\mathbf{x}$ **is the probability that one subtree is of size j.**

THE CORRESPONDING DE

$f(z) = \sum_n a_n z^n$

Since

$$\pi_{n,j} = \frac{1}{n} \sum_{j<i_1\leq\ldots\leq i_{d-1}\leq n} \frac{1}{i_1 \cdot \ldots \cdot i_{d-1}},$$

we get the DE

$$\theta(z\theta)^{d-1}(f-g) = 2^d f.$$

An extended Cauchy-Euler DE

Rewrite $\theta(z\theta)^{d-1}(f-g) = 2^d f$ **as**

$$(\theta^d - 2^d)f = \theta(z\theta)^{d-1}g + \sum_{0\leq j<d} (1-z)^j \underbrace{\varpi_j(\theta)}_{\textbf{polynomial}} f$$

The sum on RHS will be asymptotically smaller than LHS since we are interested in the asymptotics of f as $z \sim 1$.

ANALYTIC AND ELEMENTARY APPROACHES

Two different approaches proposed in the literature

- **Analytic approach by Flajolet et al. (1995) based on Euler transform and singularity analysis: analytic properties of g is needed; very precise expansion can be derived.**

$$a_n = b_0+((2^d-1)b_0+b_1)n+\sum_{2\le k\le n}\binom{n}{k}(-1)^k\sum_{2\le j\le k}\frac{b_j^*-b_{j-1}^*}{\prod_{j<\ell\le k}(1-\frac{2^d}{\ell^d})}$$

  where $b_j^* := \sum_{0\le\ell\le j}\binom{j}{\ell}(-1)^\ell b_\ell$.

- **Elementary DE approach (the extended Cauchy-Euler DE): useful when only asymptotics of b_n are known.**

NUMBER OF LEAVES IN RANDOM QUADTREES

A phase change

Chern, Fuchs, H. (2005): ***If*** $1 \leq d \leq 8$***, then the number*** X_n ***of leaves is asymptotically normally distributed; if*** $d \geq 9$***, then the random variables*** $(X_n - \mathbb{E}(X_n))/\sqrt{\mathbb{V}(X_n)}$ ***do not tend to a fixed limit law.***

Asymptotic transfer results used

Let $\beta > 1$**.**

$$a_n \sim Cn \ \textbf{ iff } \ b_n = o(n) \textbf{ and } \left|\sum_n b_n n^{-2}\right| < \infty$$

$$a_n \sim \frac{c(\beta+1)^d}{(\beta+1)^d - 2^d} n^\beta \ \textbf{ iff } \ b_n \sim cn^\beta.$$

THE LINEARITY CONSTANT C

Recall $(\theta^d - 2^d)f = \theta(z\theta)^{d-1}g + \sum_{0\le j<d}(1-z)^j\varpi_j(\theta)f$

$$C := \frac{2}{d}\sum_{k\ge 0} V_k \sum_{j\ge 0}\frac{j!b_j}{(k+2)\cdots(k+j+2)},$$

where $W_k = 0$ if $k < 0$, $W_0 = 1$ and for $k \ge 1$

$$W_k = \sum_{1\le \ell<d}\frac{\varpi_\ell(k+2)}{\varpi_0(k+2)}\,W_{k-\ell}.$$

CONCLUSION

Analysis of algorithms: a rich source of DEs

(phenomenally intriguing, mathematically challenging)