PHASE CHANGES IN RANDOM STRUCTURES AND ALGORITHMS

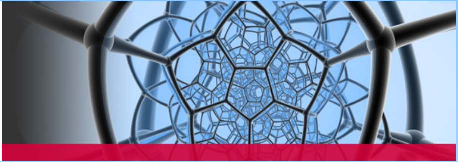
Hsien-Kuei Hwang

Summer School in Applied Probability

May 20, 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**

LECTURE II: METHOD OF MOMENTS FOR RECURSIVELY DEFINED RANDOM VARIABLES AND ITS REFINEMENTS

Part I: Method of moments

The moment of order α of X is $\mathbb{E}(X^\alpha) = \sum_j j^\alpha \mathbb{P}(X = j)$.

MOMENT

Moment was first used in a statistics sense by Karl Pearson in October 1893 in *Nature*: "Now the centre of gravity of the observation curve is found at once, also its area and its first four moments by easy calculation" (OED2).

Karl Pearson (1957–1936)

Accordingly I proceed *not* by the method suggested in Prof. Edgeworth's "Law of Error and the Elimination of Chance" (*Phil. Mag.* p. 318, April 1886), but by a method of higher moments.

Reckoned from O, the distance ON to the vertical through the centre of gravity, G, of the system of rectangles is $c(1+nq)$.

I now calculate the moments of the rectangles round the vertical, OY, and find for the rth moment

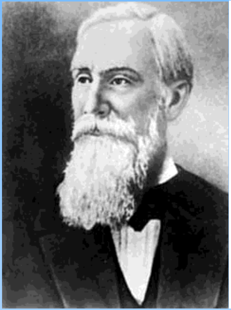
$$M_r = \alpha c^r \frac{d}{dq} q \frac{d}{dq} q \frac{d}{dq} \ldots \text{ to } r \text{ differentiations } \{q(p+q)^n\},$$

where $p+q$ is only to be put unity after differentiation, and c is supposed small. From the first four moments about OY, I find the first four moments about NG with the following results:—

$$\mu_1 = 0,$$
$$\mu_2 = npq\alpha c^2,$$
$$\mu_3 = npq(p-q)\alpha c^3,$$
$$\mu_4 = npq\{1+3(n-2)qp\}\alpha c^4.$$

Now the centre of gravity of the observation curve is found at once, also its area and its first four moments by easy calculation. Thus the position of NG, α, μ_2, μ_3, and μ_4 are given

The method of moments (or the method of mathematical expectation) dates back to work by Pafnuty Lvovich Chebyshev (1821–1894) in his version of the classical central limit theorems.



OEUVRES

DE

P. L. TCHEBYCHEF,

PUBLIÉES PAR LES SOINS

de MM. A. MARKOFF et N. SONIN,

MEMBRES ORDINAIRES DE L'ACADÉMIE IMPÉRIALE DES SCIENCES.

METHOD OF MOMENTS

Frechet-Shohat 1931

If $\mathbb{E}(X_n^m) \to \mu_m < \infty$, as $n \to \infty$ and for $m = 1, 2, \ldots$, and the sequence $\{\mu_m\}$ determines uniquely a distribution, then

$$X_n \xrightarrow{d} X,$$

where $\mathbb{E}(X^m) = \mu_m$.

Carleman's condition

If

$$\sum_k \mu_{2k}^{-1/(2k)} = \infty,$$

$\Longrightarrow$ unique determination of the distribution.

A special case: $\sum_k \mu_k x^k / k!$ is entire.

A TOY EXAMPLE

$X_n \sim \text{Binomial}(n; p)$

$$\mathbb{P}(X_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$\mathbb{E}(e^{X_n s}) = (1 + p(e^s - 1))^n.$$

$\mathbb{E}(X_n) = pn$

$$\mathbb{E}(e^{(X_n - pn)s}) = e^{n \log(1+p(e^s-1)) - pns}$$

$$= \exp\left(n \sum_{j\geq 2} \kappa_j \frac{s^j}{j!}\right).$$

In particular, $\kappa_2 = p(1-p)$ **and** $\kappa_3 = -p(1-p)(1-2p)$**.**

A TOY EXAMPLE

Thus taking coefficients of s^m

$$\mathbb{E}(X_n - pn)^m = m! \sum_{\substack{2j_2+\cdots+mj_m=m \\ j_2,\ldots,j_m\geq 0}} \frac{1}{j_2!}\left(\frac{n\kappa_2}{2!}\right)^{j_2} \cdots \frac{1}{j_m!}\left(\frac{n\kappa_m}{m!}\right)^{j_m}.$$

If $m = 2r$, then

$$\mathbb{E}(X_n - pn)^{2r} \sim \frac{(2r)!}{r!2^r}(\kappa_2 n)^r$$

If $m = 2r+1$, then

$$\mathbb{E}(X_n - pn)^{2r+1} \sim \frac{(2r+1)!}{3(r-1)!2^r}\kappa_2^{r-1}\kappa_3 n^r$$

A TOY EXAMPLE

Thus ($\kappa_2 = p(1-p)$) for $r = 1, 2, \ldots$

$$\mathbb{E}\left(\frac{X_n - pn}{\sqrt{p(1-p)n}}\right)^{2r} \sim \frac{(2r)!}{r!2^r}$$

$$\mathbb{E}\left(\frac{X_n - pn}{\sqrt{p(1-p)n}}\right)^{2r-1} = o(1).$$

Since $\sum_{k\geq 0} \frac{(2k)!}{k!2^k} \cdot \frac{z^{2k}}{(2k)!} = e^{z^2/2}$ is entire

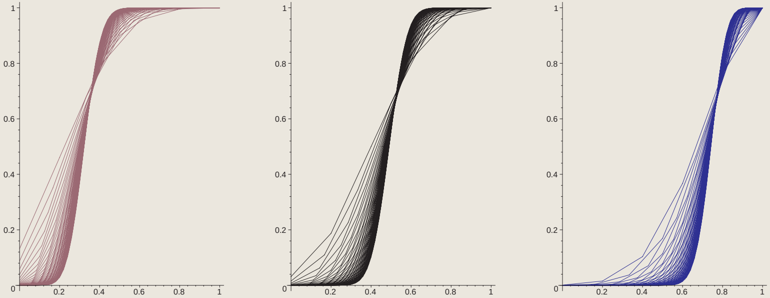
we conclude, by the moment convergence theorem, that

$$\frac{X_n - pn}{\sqrt{p(1-p)n}} \xrightarrow{d} \mathscr{N}(0,1),$$

with convergence of all moments.

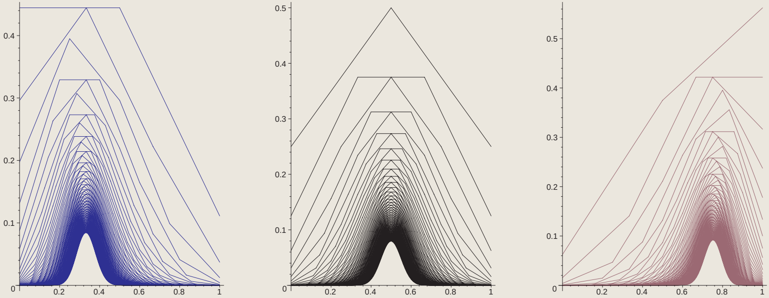
BINOMIAL $\Longrightarrow$ NORMAL (CLT)

$$p = \frac{1}{3}, \frac{1}{2}, \frac{3}{4}$$



BINOMIAL $\Longrightarrow$ NORMAL (LLT)

$$p = \frac{1}{3}, \frac{1}{2}, \frac{3}{4}$$



METHOD OF MOMENTS

Features

- **Brute force (?)**
- **Less probability (less modern, more classical)**
- **More transparent**
- **Stronger than weak convergence**

Use it as the last weapon

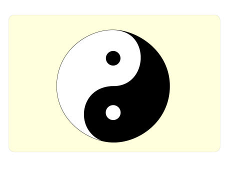
New features

When applied to random recursive structures:

- **All moments satisfy the same recurrence**
- **Asymptotic transfers**
- **Refinements $\Rightarrow$ convergence rate and LLT**

RECURSIVE RANDOM VARIABLES

Recursion is ubiquitous in Computer Algorithms and in Combinatorial Structures.



Famous numbers

- **Binomial (successes in n trials)**

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$$

- **Stirling first (cycles in permutations)**

$$s(n,k) = (n-1)s(n-1,k) + s(n-1,k-1)$$

- **Stirling second (blocks in set partitions)**

$$S(n,k) = kS(n-1,k) + S(n-1,k-1)$$

- **Eulerian (runs in permutations)**

$$A(n,k) = (k+1)A(n-1,k) + (n-k)A(n-1,k-1)$$

- **Eulerian second (leaves in plane-ordered recursive trees)**

$$a(n,k) = ka(n-1,k) + (2n-k)a(n-1,k-1)$$

RECURSIVE RANDOM VARIABLES

Examples

- **Quicksort (binary search trees)**

$$P_n(y) = \frac{Q_n(y)}{n} \sum_{0 \le k < n} P_k(y) P_{n-1-k}(y)$$

$$X_n \stackrel{d}{=} X_{\text{Uniform}[0,n-1]} + X_{n-1-\text{Uniform}[0,n-1]} + Y_n$$

- **Mergesort**

$$P_n(y) = P_{\alpha(n)}(y) P_{n-\alpha(n)}(y) Q_n(y)$$

$$X_n \stackrel{d}{=} X_{\alpha(n)} + X_{n-\alpha(n)} + Y_n$$

  - **top-down:** $\alpha(n) = \lfloor n/2 \rfloor$
  - **bottom-up:** $\alpha(n) = 2^{\lceil \log_2 n/2 \rceil}$
  - **queue-:** $\alpha(n) = 2^{\lfloor \log_2 2n/3 \rfloor}$

Examples

- **Digital trees**

$$P_{n+b}(y) = Q_n(y) \sum_{0\le k\le n} \binom{n}{k} p^k (1-p)^{n-k} P_k(y) P_{n-k}(y)$$

$$b = 0 \Longrightarrow \text{tries}$$
$$b \ge 1 \Longrightarrow \text{bucket digital search trees}$$

- **Analysis of a trading algorithm (Frieze-Pittel)**

$$P_n(y) = y \sum_{0\le j<n} \frac{n!(n-j)}{n^{n-j+1} j!} P_j(y)$$

RECURSIVE RANDOM VARIABLES

Examples

- **Maxima in right triangle**

$$P_n(y) = y \sum_{j+k+\ell=n-1} \underbrace{\binom{n-1}{j,k,\ell} \frac{(2j+\ell)!(2k+\ell)!}{(2n-1)!} 2^{\ell}}_{\pi_{j,k,\ell}(n)} P_j(y) P_k(y)$$

- **Generalized quicksort (Hennequin)**

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

$$m \geq 2, t = 0 \Longrightarrow \text{m} - \text{ary search tree}$$
$$m = 2, t \geq 0 \Longrightarrow \text{median of } (2\text{t}+1) \text{ quicksort}$$

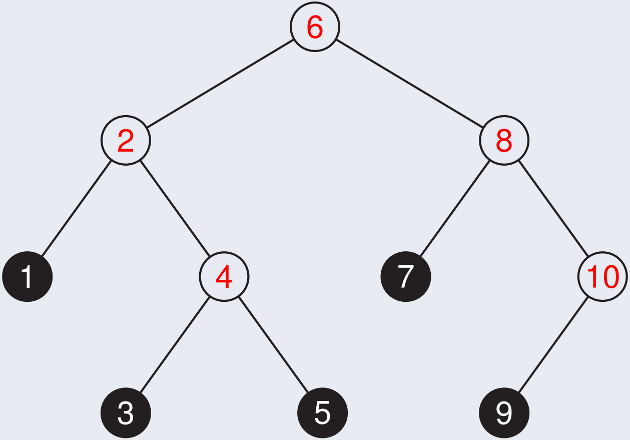
TYPICAL APPROACHES FOR RECURSIVE RVs

Q: Mean, variance, limit law?

- **Complex-analytic method (Bender, Canfield, Richmond, Flajolet, Drmota, . . .)**
- **Contraction method (Rösler, Rüschendorf, Neininger)**
- **Inductive approximation method (Mahmoud, Pittel)**
- **Urn models (Mahmoud, Smythe, Janson)**
- **Branching processes (Devroye, Rösler)**
- **Sum of RVs, Stein's method (Janson, Devroye, . . .)**
- **Method of moments (classical; Flajolet, Fill, H., . . .)**

NUMBER OF LEAVES IN RANDOM BSTs

$$X_n \stackrel{d}{=} X_{\text{uniform}[0,n-1]} + X^*_{n-1-\text{uniform}[0,n-1]} + \delta_{n,1}$$



NUMBER OF LEAVES IN RANDOM BSTs

$X_n \stackrel{d}{=} X_{\text{uniform}[0,n-1]} + X^*_{n-1-\text{uniform}[0,n-1]} + \delta_{n,1}$

Let $P_n(y) := \mathbb{E}(e^{X_n y})$**. Then** $P_0(y) = 1$ **and**

$$P_n(y) = \frac{e^{\delta_{n,1} y}}{n} \sum_{0 \le j < n} P_j(y) P_{n-1-j}(y) \qquad (n \ge 1).$$

In particular, $P_1(y) = P_2(y) = e^y$**,** $P_3(y) = \frac{e^y(2+e^y)}{3}$**.**

The mean $\mu_n := \mathbb{E}(X_n)$ satisfies $\mu_0 = 0$ and

$$\mu_n = \delta_{n,1} + \frac{2}{n} \sum_{0 \le j < n} \mu_j \qquad (n \ge 1).$$

THE UNDERLYING RECURRENCE

All moments satisfy the recurrence

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

The exact solution

Since $na_n - (n-1)a_{n-1} = 2a_{n-1} + nb_n - (n-1)b_{n-1}$**, we obtain (assuming** $a_0 = 0$**)**

$$a_n = \frac{n+1}{n} a_{n-1} + \bar{b}_n,$$

where $\bar{b}_n := nb_n - (n-1)b_{n-1}$**.**

SOLUTION TO THE UNDERLYING RECURRENCE

Iterating $a_n = \frac{n+1}{n} a_{n-1} + \bar{b}_n$

$$\begin{aligned} a_n &= \frac{n+1}{n}\left(\frac{n}{n-1} a_{n-2} + \bar{b}_{n-1}\right) + \bar{b}_n \\ &= \frac{n+1}{n-1} a_{n-2} + \frac{n+1}{n} \bar{b}_{n-1} + \bar{b}_n \\ &= \cdots (\textbf{using } a_0 = 0) \\ &= (n+1) \sum_{1\le j\le n} \frac{\bar{b}_j}{j+1}. \end{aligned}$$

Using $\bar{b}_n := nb_n - (n-1)b_{n-1}$, we obtain for $n \ge 1$

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

Ex: solve the case when $a_0 \neq 0$

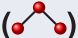
APPLICATIONS OF THE EXACT SOLUTION

Taking $b_n = \delta_{n,1}$

$$\mu_n = \frac{n+1}{3} \qquad (n \geq 2).$$

So about *one third of the nodes are leaves* on average.

Other node types

Ex: **Similarly, there are $(n+1)/3$ nodes on average with only one child (** **or** **), and $(n-2)/3$ nodes with two children (****).**

ASYMPTOTIC TRANSFER RESULTS

From the asymptotics of b_n to those of a_n

By $a_n = b_n + 2(n+1)\sum_{1\le j<n} \frac{b_j}{(j+1)(j+2)}$ **we obtain**

- ***Small "toll functions":*** $C := 2\sum_{j\ge 1} \frac{b_j}{(j+1)(j+2)}$

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

- ***Linear "toll functions"*: If** $b_n \sim cn$**, then**

$$a_n \sim 2cn\log n;$$

- ***Large "toll functions"*: Let** $\alpha > 1$**. Then**

$$a_n \sim cn^{\alpha} \textbf{ iff } b_n \sim c\frac{\alpha+1}{\alpha-1}n^{\alpha}.$$

RECURRENCE OF CENTRAL MOMENTS OF X_n

Consider $\overline{P}_n(y) := \mathbb{E}(e^{(X_n-\mu_n)y})$.

$$\overline{P}_n(y) = \frac{1}{n} \sum_{0\le j<n} \overline{P}_j(y)\overline{P}_{n-1-j}(y)e^{\Delta_{n,j}y},$$

where $\Delta_{n,j} := \delta_{n,1} - \mu_n + \mu_j + \mu_{n-1-j}$ **satisfies**

$$\Delta_{n,j} \in \{-\tfrac{1}{3}, 0, \tfrac{1}{3}, \tfrac{2}{3}\},$$

for $n \ge 1$ **and** $0 \le j < n$**.**

Only boundedness is needed.

RECURRENCE OF CENTRAL MOMENTS OF X_n

By $\overline{P}_n(y) = \frac{1}{n}\sum_{0\le j<n} \overline{P}_j(y)\overline{P}_{n-1-j}(y)e^{\Delta_{n,j}y}$

Let $P_{n,m} := \overline{P}_n^{(m)}(0) = \mathbb{E}(X_n - \mu_n)^m$**. Then** $P_{n,0} = 1, P_{n,1} = 0$ **and for** $m \ge 2$

$$P_{n,m} = \frac{2}{n}\sum_{0\le j<n} P_{j,m} + Q_{n,m},$$

where

$$Q_{n,m} := \sum_{\substack{h+k+\ell=m \\ h,k<m}} \binom{m}{h,k,\ell}\frac{1}{n}\sum_{0\le j<n} P_{j,h}P_{n-1-j,k}\Delta_{n,j}^{\ell}.$$

VARIANCE OF X_n

Recurrence now has the form: $V_n := P_{n,2} = \mathbb{V}(X_n)$

$$V_n = \frac{2}{n} \sum_{0 \le j < n} V_j + Q_{n,2},$$

where $Q_{n,2} := \frac{1}{n} \sum_{0 \le j < n} \Delta_{n,j}^2 = O(1)$, so that $V_n \sim \sigma^2 n$.

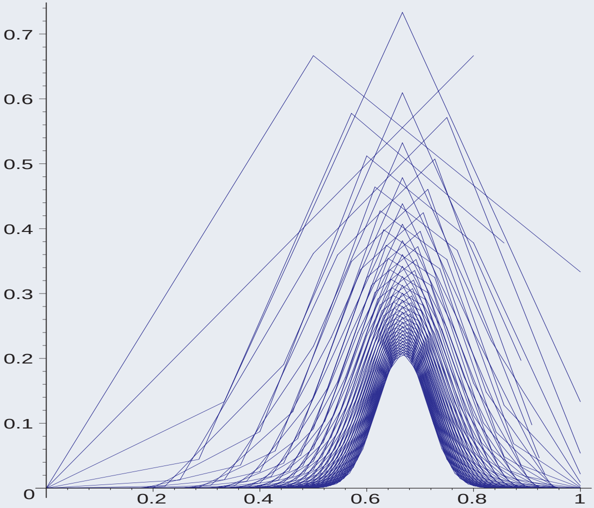
More precisely

$$Q_{n,2} = \begin{cases} 0, & \text{if } n \le 2; \\ \frac{2}{3}, & \text{if } n = 3; \\ \frac{4}{9}, & \text{if } n \ge 4. \end{cases}$$

implying that $V_n = \frac{2}{45}(n+1)$ for $n \ge 4$ and $V_3 = \frac{2}{9}$.

WHAT ABOUT THE LIMIT LAW?

Histograms of X_n: look like normal curve



ASYMPTOTICS OF CENTRAL MOMENTS OF X_n

Focus: proving asymptotic normality

By induction, we prove for $r \geq 1$

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

These will imply that ($\sigma = \sqrt{2/45}$)

$$\frac{X_n - \mu_n}{\sigma\sqrt{n}} \xrightarrow{d} \mathscr{N}(0,1).$$

$r = 1$ ***OK; it remains*** $r \geq 1$

ASYMPTOTICS OF CENTRAL MOMENTS OF X_n

By induction, the order of each term

$$Q_{n,m} := \sum_{\substack{h+k+\ell=m \\ h,k\leq m}} \binom{m}{h,k,\ell} \frac{1}{n} \sum_{0\leq j<n} \underbrace{P_{j,h}}_{\substack{|| \\ O(j^{h/2})}} \underbrace{P_{n-1-j,k}}_{\substack{|| \\ O(n-1-j)^{k/2}}} \underbrace{\Delta_{n,j}^{\ell}}_{\substack{|| \\ O(1)}}$$

The dominant part of $Q_{n,m}$

$$Q_{n,m} := \sum_{\substack{h+k=m \\ 1\leq h,k<m}} \binom{m}{h} \frac{1}{n} \sum_{0\leq j<n} P_{j,h} P_{n-1-j,k}$$

$$+ O\left(\sum_{\substack{h+k+\ell=m \\ h,k<m \\ 1\leq \ell\leq m}} \binom{m}{h,k,\ell} \frac{1}{n} \sum_{0\leq j<n} j^{h/2}(n-1-j)^{k/2} \right)$$

ASYMPTOTICS OF CENTRAL MOMENTS OF X_n

Asymptotics of the convoluted sum

$$\frac{1}{n}\sum_{0\le j<n} j^a(n-1-j)^b = O\left(\int_0^n x^a(n-x)^b dx\right)$$
$$= O\left(n^{a+b}\right) \qquad (a, b > -1).$$

$$Q_{n,m} := \sum_{\substack{h+k=m\\1\le h,k<m}} \binom{m}{h}\frac{1}{n}\sum_{0\le j<n} P_{j,h}P_{n-1-j,k} + O\left(n^{(m-1)/2}\right).$$

If $m = 2\ell - 1$, then either h or k is odd

$$Q_{n,2\ell-1} = o\left(n^{\ell-1/2}\right) \implies P_{n,2\ell-1} = o(n^{\ell-1/2}).$$

ASYMPTOTICS OF EVEN CENTRAL MOMENTS

$m = 2\ell$: dominant terms are those with both h and k even

$$
\begin{aligned}
Q_{n,2\ell} &\sim \sum_{1\le h<\ell} \binom{2\ell}{2h} \frac{1}{n} \sum_{0\le j<n} P_{j,2h} P_{n-1-j,2\ell-2h} \\
&\sim \sum_{1\le h<\ell} \binom{2\ell}{2h} \frac{(2h)!(2\ell-2h)!}{h!2^h(\ell-h)!2^{\ell-h}} \cdot \frac{\sigma^{2\ell}}{n} \sum_{0\le j<n} j^h (n-1-j)^{\ell-h} \\
&\sim \frac{(2\ell)!\sigma^{2\ell}}{2^\ell \ell!} \sum_{1\le<\ell} \binom{\ell}{h} \frac{1}{n} \int_0^n x^h (n-x)^{\ell-h} dx \\
&= \frac{(2\ell)!\sigma^{2\ell}}{2^\ell \ell!} \sum_{1\le<\ell} \binom{\ell}{h} \cdot \frac{h!(\ell-h)!}{(\ell+1)!} n^\ell \\
&= \frac{\ell-1}{\ell+1} \cdot \frac{(2\ell)!\sigma^{2\ell}}{2^\ell \ell!} n^\ell .
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{n} \int_0^n x^h (n-x)^{\ell-h} dx
&\overset{x \mapsto nt}{=} n^\ell \int_0^1 t^h (1-t)^{\ell-h} dt \\
&= n^\ell \frac{h!(\ell-h)!}{(\ell+1)!}
\end{aligned}
$$

ASYMPTOTICS OF EVEN CENTRAL MOMENTS

By the transfer: $a_n \sim cn^\alpha$ iff $b_n \sim c\frac{\alpha+1}{\alpha-1} n^\alpha$

$$P_{n,2\ell} \sim \frac{(2\ell)!\sigma^{2\ell}}{2^\ell \ell!} n^\ell$$

This completes the proof of the CLT.

Features

- **Completely elementary, no advanced probability, no complex analysis**
- **Straightforward and rather mechanical in some sense**
- **All asymptotics reduced to "asymptotic transfer"**

ADVANTAGES OF THE METHOD OF PROOF (I)

Despite messy details

- **The same proof extends easily to**

$$X_n \stackrel{d}{=} X_{\text{uniform}[0,n-1]} + X^*_{n-1-\text{uniform}[0,n-1]} + T_n$$

with

$$T_n = O(n^{1/2}(\log n)^{-1/2-\varepsilon}),$$

and

$$T_n \sim cn^{1/2}(\log n)^{\beta},$$

for which X_n is asymptotically normally distributed

$$\frac{X_n - \mathbb{E}(X_n)}{\sqrt{\mathbb{V}(X_n)}} \xrightarrow{d} \mathscr{N}(0,1).$$

ADVANTAGES OF THE METHOD OF PROOF (I)

$$X_n \stackrel{d}{=} X_{\text{uniform}[0,n-1]} + X^*_{n-1-\text{uniform}[0,n-1]} + T_n$$

- **Also extendable to** $T_n \sim cn^{\alpha}(\log n)^{\beta}$ **with** $\alpha > 1/2$ **for which** X_n **is asymptotically *non-normal***

$$\frac{X_n - \xi_n}{T_n} \xrightarrow{d} Y_\alpha,$$

where $\xi_n = \begin{cases} Cn, & \text{if } 1/2 < \alpha < 1; \\ \mathbb{E}(X_n), & \text{if } \alpha = 1; \\ 0, & \text{if } \alpha > 1, \end{cases}$ **and**

$(U = \textbf{uniform}(0,1))$

$$Y_\alpha \stackrel{d}{=} \begin{cases} U^\alpha Y_\alpha + (1-U)^\alpha Y^*_\alpha + 1, \ \textbf{if } \alpha > 1/2, \alpha \neq 1; \\ UY + (1-U)Y^* + 2U\log U + 2(1-U)\log(1-U) + 1, \\ \qquad \textbf{if } \alpha = 1, \end{cases}$$

EXAMPLE: TOTAL PATH LENGTH OF BSTs

$T_n = n - 1$: Essentially the major cost used by Quicksort

Received much attention in the last 15 years, but many problems remain open.

- **Characterization of X? (Fill and Janson, 2000–2002)**
- **Optimal Berry-Esseen bound? (Neininger and Rüschendorf, 2002)**
- **Local limit theorem?**
- **Large deviations?**
- **A purely analytic approach to this problem?**

ADVANTAGES OF THE METHOD OF PROOF (I)

Phase change

From normal to non-normal: $n^{1/2}$ ***is the threshold***

Alternative approaches

- **contraction method (under more general settings) (H. & Neininger)**
- **decomposition into subtree-functionals and Stein's method (Devroye)**
- **special cases doable by PDE $+$ analytic (Flajolet et al.)**
- **special cases doable by urn models (Mahmoud, Janson)**

EXAMPLE: LEAVES IN RANDOM BSTs

$X_n \overset{\mathscr{D}}{=} X_{\text{uniform}[0,n-1]} + X^*_{n-1-\text{uniform}[0,n-1]} + \delta_{n,1}$

Let $F(z,y) := \sum_{n,m} \mathbb{P}(X_n = m) y^m z^n$**. Then** $F(0,y) = 1$ **and**

$$\frac{\partial}{\partial z} F(z,y) = F^2(z,y) + y - 1,$$

(a Riccati DE). The exact solution (Flajolet et al., 1997)

$$F(z,y) = \frac{1 - \sqrt{1-y}\tanh(\sqrt{1-y}\,z)}{1 - \frac{1}{\sqrt{1-y}}\tanh(\sqrt{1-y}\,z)}.$$

For $y \sim 1$**, the dominant singularity lies at the zero** $z = \frac{1}{2\sqrt{1-y}} \log \frac{1+\sqrt{1-y}}{1-\sqrt{1-y}}$ **of the denominator. Then we get** ***CLT*** **and** ***LLT*** **for** X_n**.**

ADVANTAGES OF THE METHOD OF PROOF (II)

Further refinements of the method possible

- **Consider** $\phi_n(y) := E(e^{X_n y})e^{-\mu_n y - \sigma_n^2 y^2/2}$**, which satisfies a recurrence of the same type.**
- **Show that**

$$|\phi_n^{(k)}(0)| \leq A^k k! n^{k\beta}$$

  for all $k \geq 0$**.**
- **Derive a uniform estimate for the characteristic function.**
- **CLT with error bound: Berry-Esseen smoothing inequality:**

$$\sup_x |P(X_n^* < x) - \Phi(x)| = O\left(T^{-1} + \int_{-T}^{T} \left|\frac{\varphi_n(y) - e^{-y^2/2}}{y}\right| dy\right).$$

ADVANTAGES OF THE METHOD OF PROOF (II)

$X_n \stackrel{d}{=} X_{\text{uniform}[0,n-1]} + X^*_{n-1-\text{uniform}[0,n-1]} + cn^\beta$

If $\beta < 1/2$, then

$$\sup_{-\infty<x<\infty} \left| P\left(\frac{X_n - \mathbb{E}(X_n)}{\sqrt{\mathrm{Var}(Z_n)}} < x \right) - \Phi(x) \right| = \begin{cases} O(n^{-1/2}), & \text{if } \beta < 1/3; \\ O(n^{-1/2}\log n), & \text{if } \beta = 1/3; \\ O(n^{-3(1/2-\beta)}), & \text{if } 1/3 < \beta < 1/2. \end{cases}$$

ADVANTAGES OF THE METHOD OF PROOF (III)

The *moments-transfer* approach applicable to more general recurrences

- **phase changes in random *m*-ary search trees, generalized quicksort, quadtrees, . . .**
- **maxima in triangles**
- **digital tree structures: tries, digital search trees, conflict resolution algorithms, . . .**
- **bivariate shape parameters in random log-trees**
- **recursive heuristics for random graphs**

Hard part

The development of asymptotic transfer results

METHOD OF MOMENTS FOR RECURSIVE RVs

Recurrence of MGF

↓

Asymptotics of Mean and/or Variance

↓

Proper Scaling

↓

Recurrence of Scaled MGF

↓

Guess Limit Law

↓

Asymptotics of Higher Moments

↓

Limit Law

MAIN STEPS

1. **Recurrence of MGF** $P_n = \Pi_n[P_0, \ldots, P_{n-1}]$**;**
2. **Recurrence of mean**

$$a_n = c\sum_j \pi_{n,j} a_j + b_n;$$

3. **Asymptotic transfers: if** $b_n \sim Kn^\alpha$**, then** $a_n \sim ?$ **if** $b_n = O(n^\alpha)$**, then** $a_n = O(?)$**; etc.**
4. **Mean** $\sim ?$ **Variance** $\sim ?$
5. $\phi_n := e^{-a_n y} P_n = \Pi_n[\phi_0, \ldots, \phi_{n-1}, \Delta]$**;**
6. **Recurrence for higher central moments:** $\phi_{n,k} := \phi_n^{(k)}(0)$ **satisfies**

$$\phi_{n,k} = c\sum_j \pi_{n,j}\phi_{j,k} + \psi_{n,k}.$$

7. **Induction and limit law.**

METHOD OF MOMENTS IN ANALYSIS OF ALGORITHMS

Some examples

- **Height in binary trees (Flajolet and Odlyzko)**
- **Path length in binary trees (Takács)**
- **Log-product of subtree sizes in binary search trees (Fill)**
- **Cost of linear probing hashing (Flajolet et al.)**
- **Tries (Schachinger)**
- **Maximum degree in triangulations (Gao and Wormald)**
- **Random trees, urn models (Kuba, Panholzer)**
- **Random graphs (Janson, Rucinski, et al.)**
- **etc.**