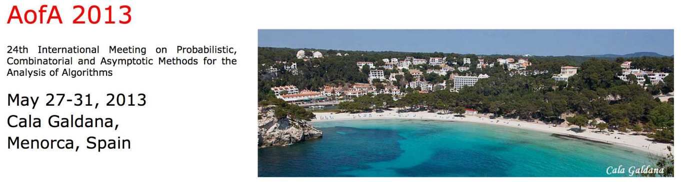
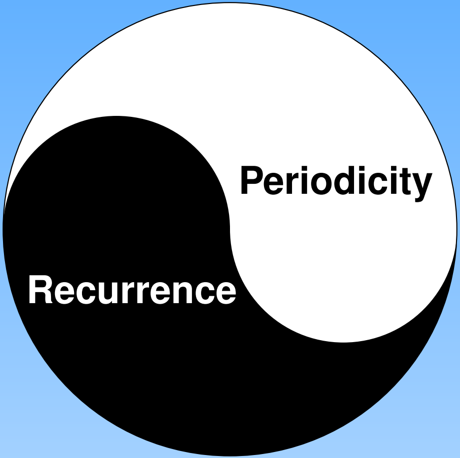
PERIODIC OSCILLATIONS OF THE VARIANCE OF TRIE STATISTICS AND RELATED STRUCTURES

Hsien-Kuei Hwang, *Academia Sinica, Taiwan*
(with Michael Fuchs, Vytas Zacharovas)

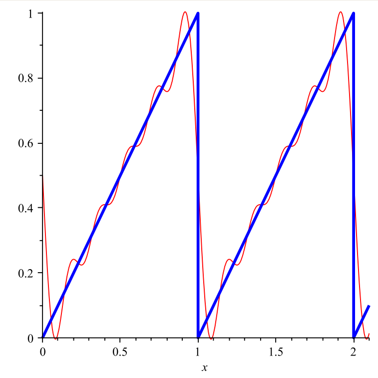


PERIODICITY EVERYWHERE IN ALGORITHMS



FLOOR & SAW-TOOTH FUNCTIONS

Fractional part



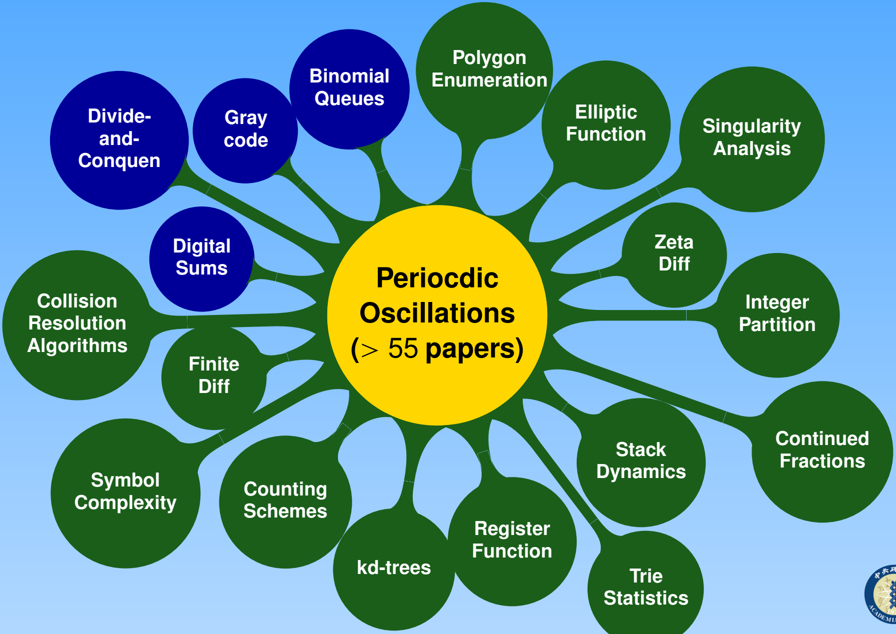
Fourier series

$$x - \lfloor x \rfloor = \frac{1}{2} - \frac{1}{\pi} \sum_{k \geqslant 1} \frac{\sin(2k\pi x)}{k}$$

($\Longleftarrow$ use five terms)

$$f_n = f_{\lfloor n/2 \rfloor} + f_{\lceil n/2 \rceil} + g_n$$

PERIODIC OSCILLATIONS IN PF'S WORKS



PERIODIC OSCILLATIONS EVERYWHERE

Two common types of periodic functions

- C^0 but ***nowhere differentiable***:

$$\frac{1}{n}\sum_{0\leqslant k<n} \#(\textbf{1's in } k\textbf{'s binary expansion})$$

$$= \tfrac{1}{2}\log_2 n + \underbrace{F(\log_2 n)}_{\textbf{fractal}} \qquad (\textbf{identity!})$$

- C^∞ with ***very small amplitude***:

$$\textbf{E}(\textbf{internal nodes in a random trie of } n \textbf{ keys})$$

$$= \frac{n}{\log 2} + \underbrace{G(\log_2 n)}_{|\cdot|\leqslant 10^{-6}} n + O(1)$$

How to compute the Fourier expansion?

TWO MAJOR ANALYTIC APPROACHES

Mellin transform: **E**(size of tries)

$$\begin{aligned}\tilde{f}(z) &= 2\tilde{f}(z/2) + 1 - (1+z)e^{-z} \\ &= -\frac{1}{2\pi i}\int_{(-\frac{3}{2})} \frac{(s+1)\Gamma(s)}{1-2^{s+1}}\, z^{-s}\,\mathbf{d}s\end{aligned}$$

$$\Longrightarrow a_n = \frac{n}{\log 2}\left(1 - \sum_{k\neq 0} \chi_k \Gamma(-1+\chi_k) n^{-\chi_k}\right) - 1 + \cdots$$

Singularity analysis: **E**(size of *m*-ary search trees)

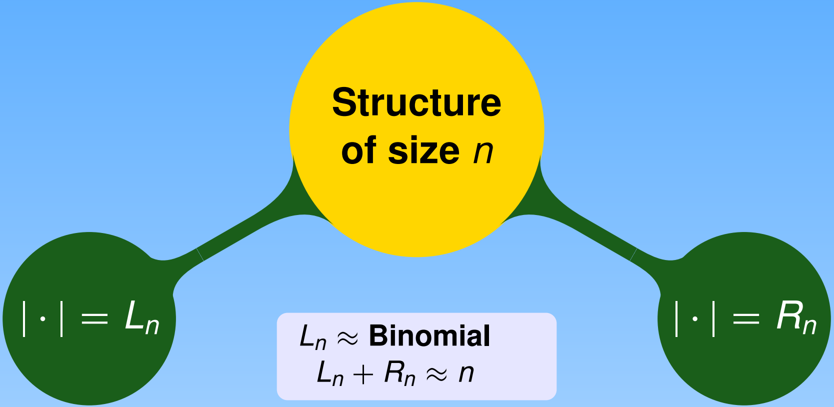
$$(1-z)^{m-1} f^{(m-1)}(z) - m! f(z) = \frac{(m-1)!}{1-z}$$

$$f(z) = \sum_{1\leqslant j<m} c_j (1-z)^{-\lambda_j} \quad (\theta^{\overline{m-1}} - m! = \prod_{1\leqslant j<m} (\theta - \lambda_j))$$

$$\Longrightarrow a_n = \frac{n+1}{2(H_m-1)} + \sum_{2\leqslant j<m} c_j \binom{n+\lambda_j-1}{n}$$

BINOMIAL SPLITTING PROCESS

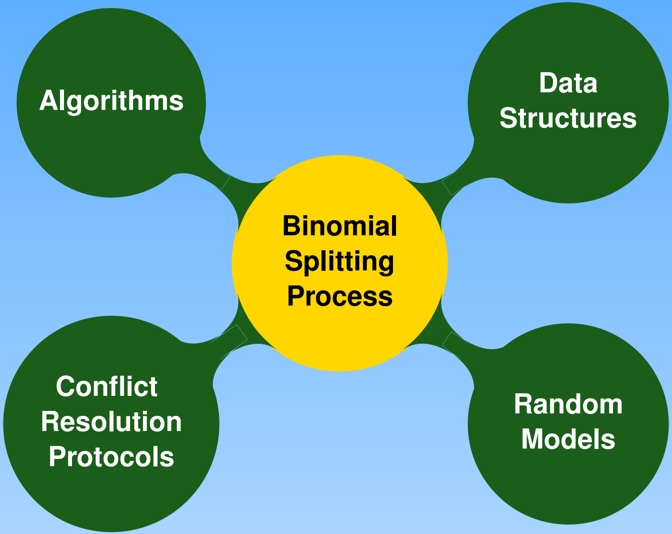
BINOMIAL SPLITTING PROCESS



$$X_n \stackrel{d}{=} X_{L_n} + X^*_{R_n} + T_n$$

Binomial Splitting Processes in PF's works
Algorithms
Polynomial factorization
Probabilistic counting schemes
Random variable generation
Bucket sort
Bit complexity
Methodology
Saddle-point method
Complex analysis
Mellin transform
Singularity analysis
Bernoulli sums
Models
Random graphs
Urn models
Random words
Dynamical systems
Buffon machines
Random walks
Applications
Digital sorting & searching
Hashing
Multi-processor systems
Collision resolution protocols
Group testing

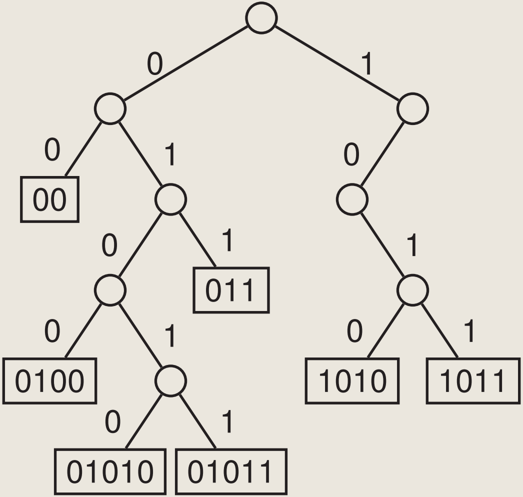
A HUGE NUMBER OF EXAMPLES




Binomial Splitting Process
Data Structures
Quadtries
Skip-lists
Suffix trees
Digital search trees
Hashing
Bucket tries
Tries, PATRICIA tries
Random Models
Random walks
Spreading rumors
Buffon machine
Urn models
Diffusion limited aggregates
Geometric IID RVs
Random graphs
Evolutionary trees
Random words
Algorithms
Huffman algorithm
Group testing
Radix & bucket sort
Exp-variate generation
Polynomial factorization
Symbol-complexity of sorting
Prob. counting schemes
Collision Resolution Protocols
Max-finding in broadcast comm. model
Mutual exclusion in mobil n/w
Tree algorithms in RFID systems
Initialization in radio n/w
Tree algos. in multiaccess channel

TRIES

TRIE (0 $\longrightarrow$ LEFT, 1 $\longrightarrow$ RIGHT)



USEFULNESS OF TRIES

Simple, efficient data structures

Widely used in diverse applications: data compression, suffix trees, internet IP addresses lookup, automatically correcting words in texts, taxonomies of regular language, . . .

Fundamental, prototype data structures

- **have a large number of variations, extensions**
- **closely connected to several *splitting procedures using coin-flipping*: collision resolution in multi-access (or broadcast) communication models, loser selection or leader election, etc.**
- **have direct *combinatorial interpretations* in terms of words, urn models, extendible hashing, etc.**

TRIES: LOG-SHAPE CHARACTERISTICS

- ***Depth***: Devroye (1982, 1992, 1999), Mendelson (1982), Pittel (1986), Jacquet & Régnier (1986), Szpankowski (1987, 1988), Kirschenhofer & Prodinger (1988), Jacquet & Szpankowski (1994), Louchard (1994), Schachinger (2001), Fayolle & Ward (2005), Christohpi & Mahmoud (2008), Janson (2012).
- ***Height***: Mendelson (1982), Flajolet & Steyaert (1982), Flajolet (1983), Devroye (1984), Pittel (1985, 1986), Jacquet & Régnier (1986), Szpankowski (1991), Devroye (1992, 1999, 2002), Clément, Flajolet, Vallée (2001), Broutin & Devroye (2008), Devroye, Lugosi, Park & Szpankowski (2009), Janson (2012).
- ***Horton-Strahler # & stack-size***: Devroye & Kruszewski (1996), Nebel (2000, 2002), Bourdon, Nebel & Vallée (2001).
- ***One-sided height (or leader election or loser selection)***: Prodinger (1995), Fill, Mahmoud, Szpankowski (1996), Janson & Szpankowski (1997), Ward & Szpankowski (2004, 2005), Louchard & Prodinger (2006), Mahmoud & Ward (2008).

TRIES: LINEAR SHAPE CHARACTERISTICS

- **Sum(internal nodes)**: Jacquet & Régnier (1988), Régnier & Jacquet (1989), Kirschenhofer & Prodinger (1991), Jacquet & Szpankowski (1994), Rachev & Rüschendorf (1995), Schachinger (1995), Knuth (1998), Clément, Flajolet, Vallée (2001), Schachinger (2001, 2004), Neininger & Rüschendorf (2004), Wagner (2009).
- **Node sorts**: Tsybakov & Mikhailov (1978), Capetanakis (1979), Mendelson (1982), Flajolet (1983), Mathys & Flajolet (1985), Kaplan & Gulko (1985), Szpankowski (1988), Kirschenhofer & Prodinger (1988, 1991), Janssen & de Jong (2000), Schachinger (2001), Nguyễn-Thế (2003), Mahmoud & Ward (2008), Wagner (2010), Janson (2012), Gaither, Homma, Sellke & Ward (2012).
- **External path length**: Devroye (1984), Szpankowski (1987), Kirschenhofer, Prodinger, Szpankowski (1989), Schachinger (1995, 2001, 2004), Clément, Flajolet, Vallée (2001), Nguyễn-Thế (2003), Neininger & Rüschendorf (2004).

RANDOM TRIES

The simplest Bernoulli model

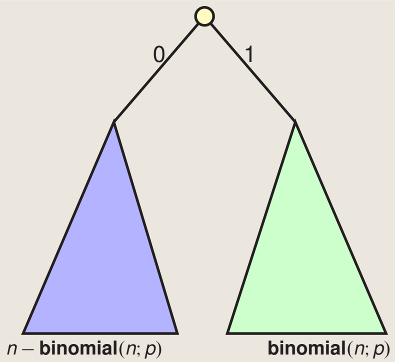
Input $= \{Y_1, Y_2, \ldots, Y_n\}$ **iid**

$$Y_i = \{Y_{i,j}\}_{j \geqslant 1},$$

$$\begin{cases} \mathbf{P}(Y_{i,j} = 1) = p \\ \mathbf{P}(Y_{i,j} = 0) = q = 1 - p. \end{cases}$$

***Random trie*: the trie constructed from this sequence of RVs.**

$$X_n \stackrel{d}{=} X_{L_n} + X^*_{R_n} + T_n$$



Many other general models in the literature

SIZE OF TRIES

SIZE OF RANDOM TRIES (BERNOULLI MODEL)

$$X_n \stackrel{d}{=} X_{\mathrm{Binom}(n;p)} + X_{n-\mathrm{Binom}(n;p)} + 1$$

$\mu_n := \mathbf{E}(X_n)$ **satisfies** $\mu_0 = \mu_1 = 0$ **and**

$$\mu_n = \sum_{0\leqslant k\leqslant n} \underbrace{\binom{n}{k} p^k q^{n-k}}_{\pi_{n,k}} (\mu_k + \mu_{n-k}) + 1 \qquad (n \geqslant 2).$$

Poisson generating function $\tilde{f}(z) := e^{-z} \sum_n \mu_n z^n/n!$

$$\tilde{f}(z) = \tilde{f}(pz) + \tilde{f}(qz) + 1 - (1+z)e^{-z}.$$

Mellin transform ($G_1(s) := -(s+1)\Gamma(s)$)

$$\tilde{f}(z) = \frac{1}{2\pi i} \int_{(-\frac{3}{2})} \frac{G_1(s)}{1 - p^{-s} - q^{-s}} z^{-s}\, \mathbf{d}s$$

EXPECTED SIZE OF RANDOM TRIES

Growth order + calculus of residues
($h := -p \log p - q \log q$)

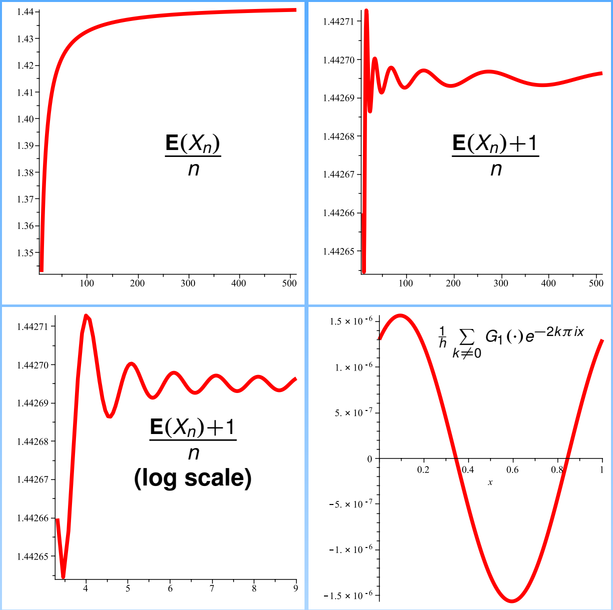
$$\frac{\tilde{f}(z)}{z} \sim \frac{1}{h} + \begin{cases} \frac{1}{h} \overbrace{\sum_{k \neq 0} G_1(-1+\chi_k) z^{-\chi_k}}^{\textbf{Periodic}}, & \textbf{if } \frac{\log p}{\log q} = \frac{r}{\ell}, (r,\ell) = 1 \\ 0, & \textbf{if } \frac{\log p}{\log q} \notin \mathbf{Q} \end{cases}$$

Uniformly for $\Re(z) > 0$ **(**$\chi_k := 2kr\pi i / \log p$**)**

Analytic de-Poissonization (Jacquet-Szpankowski)

$$\frac{\mu_n}{n} \sim \frac{1}{h} + \begin{cases} \frac{1}{h} \overbrace{\sum_{k \neq 0} G_1(-1+\chi_k) n^{-\chi_k}}^{\textbf{Periodic}}, & \textbf{if } \frac{\log p}{\log q} = \frac{r}{\ell}, (r,\ell) = 1 \\ 0, & \textbf{if } \frac{\log p}{\log q} \notin \mathbf{Q} \end{cases}$$

PERIODIC OSCILLATIONS: $p = \frac{1}{2}$



THE MELLIN-DE-POISSONIZATION APPROACH

Recurrence relation

$$a_n = \sum_k \pi_{n,k}(a_k + a_{n-k}) + b_n$$

Poisson generating function

$$\tilde{f}(z) = \tilde{f}(pz) + \tilde{f}(qz) + \tilde{g}(z)$$

Mellin inverse

$$\tilde{f}(z) = \frac{1}{2\pi i} \int_{\uparrow} \frac{\mathscr{M}[\tilde{g}; s]\, z^{-s}}{1 - p^{-s} - q^{-s}} \,\mathrm{d}s$$

Poisson-Mellin-Newton (Rice's $\int$)

$$a_n = \frac{n!}{2\pi i} \int_{\uparrow} \frac{\mathscr{M}[\tilde{f}; s]}{\Gamma(n+1-s)} \,\mathrm{d}s$$

Analytic de-Poissonization

$$a_n = \frac{n!}{2\pi i} \oint z^{-n-1} e^{z} \tilde{f}(z) \,\mathrm{d}z$$

More systematic tools

ASYMPTOTICS OF $\mathbf{V}(X_n)$

New result: $p = \frac{1}{2}$ ($\chi_k := \frac{2k\pi i}{\log 2}$)

$$\frac{\mathbf{V}(X_n)}{n} = \frac{1}{\log 2} \sum_{k \in \mathbf{Z}} G(-1 + \chi_k) n^{-\chi_k} + o(1)$$

$$G(s) = (s+1)\Gamma(s)\left(1 - \frac{s^2 + 4s + 8}{2^{s+3}}\right) + 2 \sum_{j \geqslant 1} \frac{(-1)^j j(j(j+s+1) - 1)\Gamma(j+s+1)}{(j+1)!(2^j - 1)}$$

$$(\Re(s) > -2)$$

ASYMPTOTICS OF $\mathbf{V}(X_n)$

Average value of the periodic function

$$\frac{G(-1)}{\log 2} = \frac{1}{\log 2}\left(\frac{1}{4} + 2\sum_{j\geqslant 1}\frac{(-1)^j(j-1)}{2^j-1}\right) \approx 0.845858623076001\cdots$$

Régnier & Jacquet (1988), Kirschenhofer & Prodinger (1991)

$$\frac{G(-1)}{\log 2} = \frac{1}{2\log 2} - \frac{1}{\log^2 2} - \frac{2}{\log 2}\sum_{j\geqslant 1}\frac{(-1)^j}{2^j-1} - \frac{4\pi^2}{\log^3 2}\sum_{j\geqslant 1}\frac{j}{\sinh\frac{2j\pi^2}{\log 2}}$$

Mahmoud (1992)

$$\frac{G(-1)}{\log 2} = \frac{1}{\log 2}\left(\frac{1}{2} + 2\sum_{j\geqslant 1}\frac{1}{2^j+1}\right) - \frac{1}{\log^2 2} - \frac{4\pi^2}{\log^3 2}\sum_{j\geqslant 1}\frac{j}{\sinh\frac{2j\pi^2}{\log 2}}$$

An identity (*direct proof by residue calculus*)

$$\sum_{j\geqslant 1}\frac{(-1)^j j}{2^j-1} = \frac{1}{8} - \frac{1}{2\log 2} - \underbrace{\frac{2\pi^2}{\log^2 2}\sum_{j\geqslant 1}\frac{j}{\sinh\frac{2j\pi^2}{\log 2}}}_{|\cdot|\leqslant 1.1\times 10^{-10}}$$

Fourier coefficients: $k \neq 0$

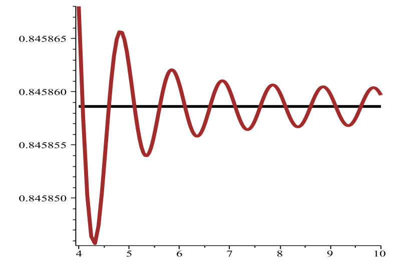
$$G(-1+\chi_k) = -\frac{\chi_k \Gamma(-1+\chi_k)(1+\chi_k)^2}{4} + 2\sum_{j\geq 1} \frac{(-1)^j j(j(j+\chi_k)-1)\Gamma(j+\chi_k)}{(j+1)!(2^j-1)}$$

Kirschenhofer & Prodinger (1991)

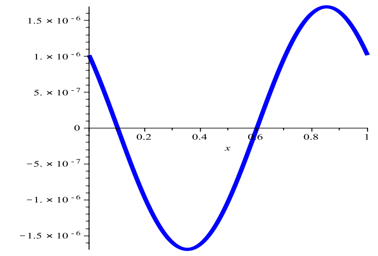
$$\begin{aligned}G(-1+\chi_k) &= -3\chi_k\Gamma(-1+\chi_k) + \frac{1}{\log 2}\sum_{\substack{j+m=k\\ j,m\neq 0}} \chi_j\Gamma(-1+\chi_j)\chi_m\Gamma(1+\chi_m)\\
&\quad - (1-\chi_k)(2-\chi_k)\Gamma(\chi_k)\left(\frac{1}{2} - \sum_{j\geqslant 1}\frac{(\chi_k+j)\binom{-\chi_k}{j-1}}{(j+1)(2^j-1)}\right)\\
&\quad - 2\Gamma(1+\chi_k)\left(\frac{5-\chi_k}{4(1-\chi_k)} - \sum_{j\geqslant 1}\frac{(\chi_k+j+1)\binom{-\chi_k-1}{j-1}}{(j+1)(2^j-1)}\right)\\
&\quad - \frac{\chi_k\Gamma(1+\chi_k)}{\log 2}\end{aligned}$$

PERIODIC OSCILLATIONS: $p = \frac{1}{2}$

$\frac{\mathbf{V}(X_n)}{n}$



$\frac{1}{\log 2}\sum_{k\in\mathbf{Z}} G(-1+\chi_k)n^{-\chi_k}$



$$\frac{1}{\log 2}\sum_{k\neq 0}|G(-1+\chi_k)| \leqslant 1.7\times 10^{-6}$$

ASYMPTOTICS OF $\mathbf{V}(X_n)$

New result: $\frac{\log p}{\log q} \notin \mathbf{Q}$

(*implicit in Jacquet & Régnier, 1989*)

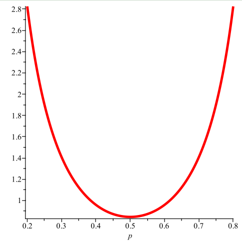
$$\frac{\mathbf{V}(X_n)}{n} = \frac{1}{h}\left(\frac{1}{2} - \frac{1}{h} + 2\sum_{j\geqslant 2} \frac{(-1)^j (p^j + q^j)}{1 - p^j - q^j}\right) + o(1)$$

$$(h := -p\log p - q\log q)$$

As $p \to \frac{1}{2}$

$$\frac{1}{h}\left(\frac{1}{2} - \frac{1}{h} + 2\sum_{j\geqslant 2} \frac{(-1)^j (p^j + q^j)}{1 - p^j - q^j}\right)$$

$$\longrightarrow \frac{1}{\log 2}\left(\frac{1}{2} - \frac{1}{\log 2} - 2\sum_{j\geqslant 1} \frac{(-1)^j}{2^j - 1}\right)$$



ASYMPTOTICS OF $\mathbf{V}(X_n)$

New result: $\frac{\log p}{\log q} = \frac{r}{\ell}, (r,\ell) = 1$ $\left(\chi_k := \frac{2kr\pi i}{\log p}\right)$

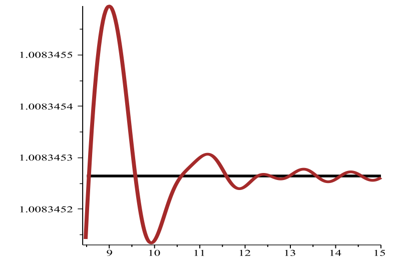
$$\frac{\mathbf{V}(X_n)}{n} = \frac{1}{h}\sum_{k\in\mathbf{Z}} G(-1+\chi_k)n^{-\chi_k} + o(1)$$

$$G(-1) = \overbrace{\frac{1}{2} - \frac{1}{h} + 2\sum_{j\geqslant 2}\frac{(-1)^j(p^j+q^j)}{1-p^j-q^j}}^{\textbf{same as irrational case}} - \frac{1}{h\log p}\sum_{j\geqslant 1}\frac{4rj\pi^2}{\sinh\frac{2rj\pi^2}{\log p}}$$

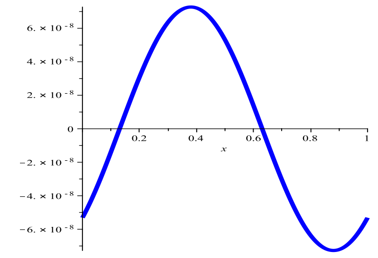
$$\begin{aligned} G(-1+\chi_k) &= \chi_k\Gamma(-1+\chi_k)\left(1-\frac{\chi_k+3}{2^{1+\chi_k}}\right) \\ &\quad - \frac{1}{h}\sum_{j\in\mathbf{Z}}\Gamma(\chi_j+1)\Gamma(\chi_{k-j}+1) \\ &\quad - 2\sum_{j\geqslant 1}\frac{(-1)^j(j+1+\chi_k)\Gamma(j+\chi_k)\left(p^{j+1}+q^{j+1}\right)}{(j-1)!(j+1)\left(1-p^{j+1}-q^{j+1}\right)} \end{aligned}$$

PERIODIC OSCILLATIONS: $q = p^2$ $\left(p = \frac{\sqrt{5}-1}{2}\right)$

$\frac{\mathbf{V}(X_n)}{n}$



$\frac{1}{h}\sum_{k\in\mathbf{Z}} G(-1+\chi_k)n^{-\chi_k}$



$$\frac{1}{h}\sum_{k\neq 0}|G(-1+\chi_k)| \leqslant 7.3\times 10^{-8}$$

METHOD OF PROOF

POISSONIZED VARIANCE

$$\tilde{f}_m(z) := e^{-z} \sum_n \frac{\mathbf{E}(X_n^m)}{n!} z^n$$

- **Mean:** $\mathbf{E}(X_n) \sim \tilde{f}_1(n)$
- **Second moment:** $\mathbf{E}(X_n^2) \sim \tilde{f}_2(n)$
- **Variance** $\boxed{\mathbf{V}(X_n) = \mathbf{E}(X_n^2) - (\mathbf{E}(X_n))^2}$

$$\tilde{D}(n) := \tilde{f}_2(n) - \tilde{f}_1(n)^2 \overset{??}{\approx} \mathbf{V}(X_n)?$$

Yes, but not good enough in certain cases.

POISSONIZED VARIANCE

$$\tilde{f}_2(z) = \tilde{D}(z) + \tilde{f}_1(z)^2$$

$$\begin{aligned}\sigma_n^2 &= \mathbf{E}(X_n^2) - \mu_n^2 \\ &= \sum_{j\geqslant 0} \frac{\tilde{f}_2^{(j)}(n)}{j!}\tau_j(n) - \left(\sum_{j\geqslant 0} \frac{\tilde{f}_1^{(j)}(n)}{j!}\tau_j(n)\right)^2 \\ &= \tilde{D}(n) - \boxed{\boxed{n\tilde{f}_1'(n)^2}} - n\tilde{f}_1(n)\tilde{f}_1''(n) + \textbf{s.o.t.}\end{aligned}$$

Take $\tilde{f}_1(z) \asymp z\log z$

$$n\tilde{f}_1'(n)^2 \asymp n(\log n)^2, \qquad n\tilde{f}_1(n)\tilde{f}_1''(n) \asymp n\log n.$$

In most such case, $\sigma_n^2 \asymp n\log n$ ***or*** $\sigma_n^2 \asymp n$.

OUR APPROACH TO $\mathbf{V}(T_n)$

The crucial step

$$\tilde{V}(z) := \tilde{f}_2(z) - \tilde{f}_1(z)^2 - z\tilde{f}_1'(z)^2$$

Then (with $\tilde{f}_2(n) = \tilde{V}(n) + \tilde{f}_1(n)^2 + n\tilde{f}_1'(n)^2$)

$$\begin{aligned}\sigma_n^2 &= \sum_{j\geqslant 0} \frac{\tilde{f}_2^{(j)}(n)}{j!}\tau_j(n) - \left(\sum_{j\geqslant 0} \frac{\tilde{f}_1^{(j)}(n)}{j!}\tau_j(n)\right)^2 \\ &= \tilde{V}(n) - \underbrace{\frac{n}{2}\tilde{V}''(n) - \frac{n^2}{2}\tilde{f}_1''(n)^2}_{=O(1)} + o(1).\end{aligned}$$

$z\tilde{f}_1'(z)^2 =$ ***the right correction term when*** $\sigma_n^2 \approx n(\log n)^\beta$

Poisson-Charlier expansion

$$\boxed{\tilde{f}(z) := e^{-z} \sum_{j\geqslant 0} \frac{a_j}{j!} z^j \text{ entire}}$$

$$\Longrightarrow \boxed{a_n = \sum_{j\geqslant 0} \frac{\tilde{f}^{(j)}(n)}{j!} \tau_j(n)}$$

$$\tau_j(n) = \sum_{0\leqslant \ell \leqslant j} \binom{j}{\ell} (-1)^{j-\ell} \frac{n\cdots(n-\ell+1)}{n^\ell}$$

First few τ_j

$$\begin{aligned} \tau_0(n) &= 1 \\ \tau_1(n) &= 0 \\ \tau_2(n) &= -n \\ \tau_3(n) &= 2n \\ \tau_4(n) &= 3n(n-2) \\ \tau_5(n) &= -4n(5n-6) \\ &\cdots \end{aligned}$$

$\mathbf{deg}(\tau_j(n)) = \lfloor j/2 \rfloor$

(Charlier or Laguerre)

POISSON-CHARLIER EXPANSION

An example: $\tilde{f}(z) = e^{-2z}$

$$(-1)^n = e^{-2n} \sum_{j \geqslant 0} \frac{(-2)^j}{j!} \tau_j(n).$$

But $(-1)^n \not\sim e^{-2n}$.

Another example: $\tilde{f}(z) = e^{z}$

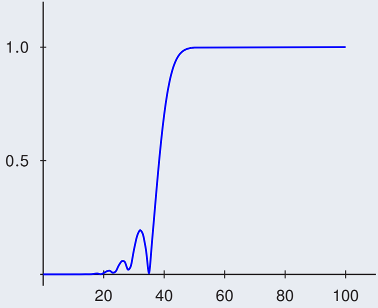
$$2^n = e^n \sum_{j \geqslant 0} \frac{\tau_j(n)}{j!}.$$

But $2^n \not\sim e^n$.

Major difficulty: prove the asymptotic nature

$$a_n = \sum_{0 \leqslant j \leqslant k} \frac{\tilde{f}^{(j)}(n)}{j!} \tau_j(n) + O(\tilde{f}^{(k)}(n)\tau_k(n))$$

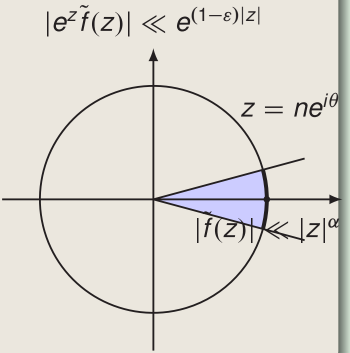
$$e^{-2n}\sum_{j\geqslant 0}\frac{(-2)^j}{j!}\tau_j(n) \longrightarrow (-1)^n \qquad (n=10,11)$$



Justification of de-Poissonization

Justify

- $\mu_n = \sum_{j\geqslant 0} \frac{\tilde{f}_1^{(j)}(n)}{j!}\tau_j(n)$;
- $\mathbf{E}(X_n^2) = \sum_{j\geqslant 0} \frac{\tilde{f}_2^{(j)}(n)}{j!}\tau_j(n)$



JACQUET AND SZPANKOWSKI (1998)

ELSEVIER

Theoretical Computer Science 201 (1998) 1–62

Theoretical Computer Science

Fundamental Study

Analytical depoissonization and its applications

Philippe Jacquet[a,1], Wojciech Szpankowski[b,*,2]

[a] *INRIA, Rocquencourt, 78153 Le Chesnay Cedex, France*

[b] *Department of Computer Science, Purdue University, W. Lafayette, IN 47907, USA*

Received October 1996; received in revised form May 1997

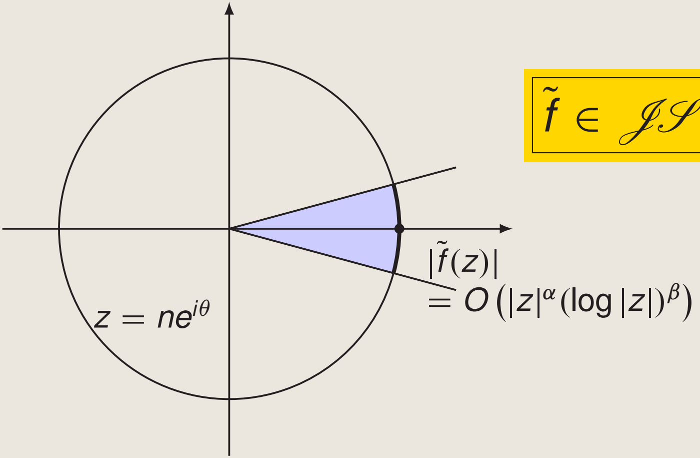
Communicated by J. Díaz

JS-ADMISSIBLE FUNCTIONS

JS-ADMISSIBLILITY

$\tilde{f}$ is JS-admissible (*combining ideas from Hayman (1956) and Jacquet-Szpankowski (1998)*)

$$|e^z \tilde{f}(z)| = O\left(e^{(1-\varepsilon)|z|}\right)$$



$$\tilde{f} \in \mathscr{JS}$$

PROPERTIES OF JS-ADMISSIBLE FUNCTIONS

If $\tilde{f} \in \mathscr{JS}_{\alpha,\beta}$, then

$$a_n = \sum_{0 \leqslant j < 2k} \frac{\tilde{f}^{(j)}(n)}{j!} \tau_j(n) + O\left(n^{\alpha-k} \log^\beta n\right),$$

Closure properties: $m \geqslant 0$, $\alpha \in (0,1)$

(i) $z^m, e^{-\alpha z} \in \mathscr{JS}$.

(ii) **If $\tilde{f} \in \mathscr{JS}$, then $\tilde{P}\tilde{f} \in \mathscr{JS}$ for any polynomial $\tilde{P}(z)$.**

(iii) **If $\tilde{f} \in \mathscr{JS}$, then $\tilde{f}(\alpha z) \in \mathscr{JS}$.**

(iv) **If $\tilde{f}, \tilde{g} \in \mathscr{JS}$, then $\tilde{f} + \tilde{g} \in \mathscr{JS}$.**

(v) **If $\tilde{f}, \tilde{g} \in \mathscr{JS}$, then $\tilde{f}(\alpha z)\tilde{g}((1-\alpha)z) \in \mathscr{JS}$.**

(vi) **If $\tilde{f} \in \mathscr{JS}$, then $\tilde{f}^{(m)} \in \mathscr{JS}$.**

TECHNICAL PROPERTIES

$$\tilde{f}(z) = \tilde{f}(pz) + \tilde{f}(qz) + \tilde{g}(z)$$

$$\boxed{\tilde{f} \in \mathscr{JS} \iff \tilde{g} \in \mathscr{JS}}$$

Closed under a Hadamard product:

$\tilde{f}(z) := e^{-z} \sum_{n\geqslant 0} \frac{a_n}{n!} z^n, \tilde{g}(z) := e^{-z} \sum_{n\geqslant 0} \frac{b_n}{n!} z^n$

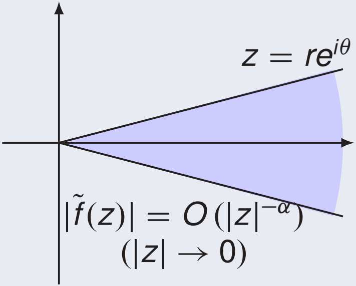
$$\tilde{h}(z) := e^{-z} \sum_{n\geqslant 0} \frac{a_n b_n}{n!} z^n$$

$\tilde{f} \in \mathscr{JS}_{\alpha_1,\beta_1}$ & $\tilde{g} \in \mathscr{JS}_{\alpha_2,\beta_1} \implies \tilde{h} \in \mathscr{JS}_{\alpha_1+\alpha_2,\beta_1+\beta_2}$

$$\tilde{h}(z) = \tilde{f}(z)\tilde{g}(z) + z\tilde{f}'(z)\tilde{g}'(z) + O\left(|z|^{\alpha_1+\alpha_2-2}(\log_+ |z|)^{\beta_1+\beta_2}\right)$$

uniformly as $|z| \to \infty$ **and** $|\arg(z)| \leqslant \theta < \pi/2$

Exponential Smallness Lemma (Flajolet et al., 1995)



$$|\tilde{f}(z)| = O\left(|z|^{-\beta}\right) \quad (|z| \to \infty)$$

$\alpha < \Re(s) < \beta$

$$\Longrightarrow \boxed{\left|\int_0^\infty e^{-zs}\tilde{f}(z)\,\mathbf{d}z\right| = O\left(e^{-\theta|\Im(s)|}\right)}$$

Mellin transform then becomes easy

POISSON GENERATING FUNCTIONS

The first two moments ($\tilde{V} := \tilde{f}_2 - \tilde{f}_1^2 - z(\tilde{f}_1')^2$)

$$\tilde{f}_1(z) = \tilde{f}_1(pz) + \tilde{f}_1(qz) + 1 - (1+z)e^{-z}$$

$$\begin{aligned}\tilde{V}(z) = \tilde{V}(pz) + \tilde{V}(qz) &+ pqz\left(\tilde{f}_1'(pz) - \tilde{f}_1'(qz)\right)^2 \\ &+ e^{-z}(1 + z - (1 + 2z + z^2 + z^3)e^{-z}) \\ &+ 2e^{-z}(1+z)(\tilde{f}_1(pz) + \tilde{f}_1(qz)) \\ &- 2z^2e^{-z}\left(p\tilde{f}_1'(pz) + q\tilde{f}_1'(qz)\right)\end{aligned}$$

Both Mellin-doable + JS-admissible

Advantages of the approach

- ***General, simple to use, easily-extendible***
- ***Simpler Fourier expansions*: NOT of the form** $F_1(x) - F_0^2(x)$.

OTHER APPROACHES USED FOR VARIANCE

A few approaches used in the literature

- ***Complex-analytic*: two-dimensional saddle-point method**

- ***Second moment*: compute the asymptotics of the 2nd moment, then handle the cancelation. (i) *more precise expansion* for μ_n is needed (ii) handle *analytically the convolution term* $\sum_j \pi_{n,j}\mu_j\mu_{n-j}$;**

- ***Poissonized variance*: relies on Poisson heuristic**

Further manipulation of cancelations needed

Other trie statistics

– **External path length, internal path length, node sorts, imbalance measures, . . .**

	$p = \frac{1}{2}$	$\frac{\log p}{\log q} \in \mathbf{Q}$	$\frac{\log p}{\log q} \notin \mathbf{Q}$
$\frac{\mathbf{EPL}_n}{n} \sim$	**Periodic**	$c \log n +$ **Periodic**	$c \log n + c'$
$\frac{\mathbf{IPL}_n}{n} \sim$	**Periodic**$(\log n)^2$	**(Periodic)**$(\log n)^2$	$c(\log n)^2$

– **Patricia tries: path length, patterns, . . .**

$$\begin{aligned}\mathscr{E}(z, y) &= \mathscr{E}(pe^y z, y)\mathscr{E}(qe^y z, y) + \mathscr{E}(pz, y) \\ &\quad + \mathscr{E}(qz, y) - \mathscr{E}(pe^y z, y) - \mathscr{E}(qe^y z, y)\end{aligned}$$

(Kirschenhofer et al., 1988)

GENERALITY

Other BSPs

– **Radix sort (Mahmoud et al., 2000)**

$$\mathscr{E}(z,y) = (1-e^y)z + \mathscr{E}\left(\frac{e^y z}{b}, y\right)^b$$

– **Conflict resolution algorithms (many papers)**

$$\tilde{P}(z,y) = e^y \prod_{1\leqslant j\leqslant r} \tilde{P}(p_j z, y) + (1-e^y)(1+z)e^{-z}$$

where $\sum_{1\leqslant j\leqslant r} p_j = 1$, $\tilde{P}(z,y) = e^{-z}\sum_{n\geqslant 0} \frac{\mathbf{E}(e^{X_n y})}{n!} z^n$.

CONFLICT RESOLUTION ALGORITHMS

Rational: $p_j = \rho^{e_j}$, $1 \leqslant j \leqslant r$; irrational otherwise

$$\frac{\mathbf{E}(X_n)}{n} = \frac{1}{h}\left(1 + \begin{cases} \sum_{k\neq 0} \chi_k \Gamma(-1+\chi_k) n^{-\chi_k}, & \textbf{rational} \\ 0, & \textbf{irrational} \end{cases}\right) + o(1)$$

$$\frac{\mathbf{V}(X_n)}{n} = \frac{1}{h}\left(G(-1) + \begin{cases} \sum_{k\neq 0} G(-1+\chi_k) n^{-\chi_k}, & \textbf{rational} \\ 0, & \textbf{irrational} \end{cases}\right) + o(1)$$

$$G(s) = (s+1)\Gamma(s)\left(1 - \frac{s^2+4s+8}{2^{s+3}}\right) \qquad \left(P(s) := \sum_j p_j^s\right)$$
$$+ 2\sum_{j\geqslant 1} \frac{(-1)^j j(j(j+s+1)-1)\Gamma(j+s+1)P(j+1)}{(j+1)!(1-P(j+1))}$$
$$+ \frac{\Gamma(3+s)}{2^{3+s}} - 2\sum_{j\geqslant 1} \frac{(-1)^j \Gamma(j+3+s)P(j+1)}{(j-1)!(1-P(j+1))}$$
$$- \begin{cases} \frac{1}{h}\sum_{j\in\mathbf{Z}} \Gamma(\chi_j+1)\Gamma(3+s-\chi_j), & \textbf{rational} \\ 0, & \textbf{irrational} \end{cases}$$

REMARKS

Features of binomial recurrences

- ***O-estimate easier, but $\sim$-estimate harder***
- ***mean and limit law easier but asymptotic variance harder***

Limit laws: at least five ways

- **Contraction method (Neininger, Rösler, Rüschendorf)**
- **Martingale difference (Schacinger, 2000)**
- **Complex-analytic (Jacquet, Régnier, Szpankowski)**
- **Renewal theory (Janson, 2012)**
- **Method of moments (Fuchs, H.)**

Gracias