Asymptotic and Exact Poissonized Variance in the Analysis of Random Digital Trees

(joint with Hsien-Kuei Hwang and Vytas Zacharovas)

Michael Fuchs

Institute of Applied Mathematics
National Chiao Tung University

June 22nd, 2017

Digital Trees

Three standard types:

- **Trie**: René de la Briandais, 1959;
- **PATRICIA Trie**: Donald R. Morrison, 1968;
- **Digital Search Tree**: Coffman and Eve, 1970.

Digital Trees

Three standard types:

- **Trie**: René de la Briandais, 1959;
- **PATRICIA Trie**: Donald R. Morrison, 1968;
- **Digital Search Tree**: Coffman and Eve, 1970.

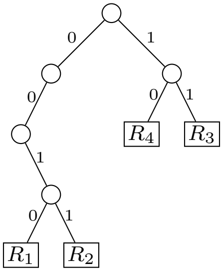
Example: $R_1{=}0010\cdots,\ R_2{=}0011\cdots,\ R_3{=}1100\cdots,\ R_4{=}1010\cdots$

Digital Trees

Three standard types:

- **Trie**: René de la Briandais, 1959;
- **PATRICIA Trie**: Donald R. Morrison, 1968;
- **Digital Search Tree**: Coffman and Eve, 1970.

Example: R_1=0010$\cdots$, R_2=0011$\cdots$, R_3=1100$\cdots$, R_4=1010$\cdots$



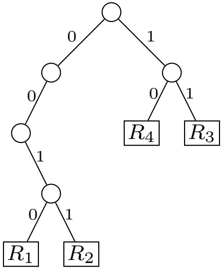
Trie

Digital Trees

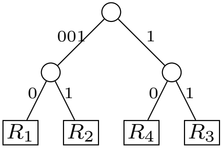
Three standard types:

- **Trie**: René de la Briandais, 1959;
- **PATRICIA Trie**: Donald R. Morrison, 1968;
- **Digital Search Tree**: Coffman and Eve, 1970.

Example: R_1=0010⋯, R_2=0011⋯, R_3=1100⋯, R_4=1010⋯



Trie



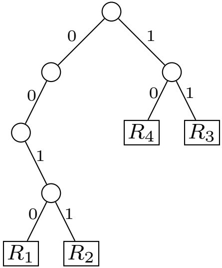
PATRICIA trie

Digital Trees

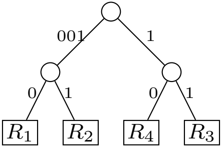
Three standard types:

- **Trie**: René de la Briandais, 1959;
- **PATRICIA Trie**: Donald R. Morrison, 1968;
- **Digital Search Tree**: Coffman and Eve, 1970.

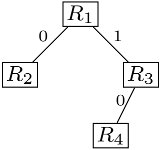
Example: $R_1 = 0010\cdots$, $R_2 = 0011\cdots$, $R_3 = 1100\cdots$, $R_4 = 1010\cdots$



Trie



PATRICIA trie



DST

Random Digital Trees

Bernoulli model:

Bits are iid with

$$\mathbb{P}(\text{bit} = 1) = p \qquad \text{and} \qquad \mathbb{P}(\text{bit} = 0) = q := 1 - p.$$

Random Digital Trees

Bernoulli model:

Bits are iid with

$$\mathbb{P}(\text{bit} = 1) = p \qquad \text{and} \qquad \mathbb{P}(\text{bit} = 0) = q := 1 - p.$$

Gives two types:

- $p = 1/2$: *symmetric digital tree*;
- $p \neq 1/2$: *asymmetric digital tree*.

Random Digital Trees

Bernoulli model:

Bits are iid with

$$\mathbb{P}(\text{bit} = 1) = p \qquad \text{and} \qquad \mathbb{P}(\text{bit} = 0) = q := 1 - p.$$

Gives two types:

- $p = 1/2$: *symmetric digital tree*;
- $p \neq 1/2$: *asymmetric digital tree*.

Bernoulli model is simple and prototypical but not very realistic!

J. Clément, P. Flajolet, B. Vallée (2001). Dynamical sources in information theory: A general analysis of trie structures, Algorithmica, 29:1-2, 307–369.

Shape Parameters

- Size S_n

  Number of internal nodes in tries; for the other two standard types the size is deterministic.

Shape Parameters

- Size S_n

  Number of internal nodes in tries; for the other two standard types the size is deterministic.

- Path Lengths L_n

  Internal path length, external path length, total path length, weighted path length, peripheral path lengths, etc.

Shape Parameters

- Size S_n

  Number of internal nodes in tries; for the other two standard types the size is deterministic.

- Path Lengths L_n

  Internal path length, external path length, total path length, weighted path length, peripheral path lengths, etc.

- Depth

Shape Parameters

- Size S_n

  Number of internal nodes in tries; for the other two standard types the size is deterministic.

- Path Lengths L_n

  Internal path length, external path length, total path length, weighted path length, peripheral path lengths, etc.

- Depth

- Number of leaves and patterns

Shape Parameters

- Size S_n

  Number of internal nodes in tries; for the other two standard types the size is deterministic.

- Path Lengths L_n

  Internal path length, external path length, total path length, weighted path length, peripheral path lengths, etc.

- Depth
- Number of leaves and patterns
- External and Internal Profiles

Shape Parameters

- Size S_n

  Number of internal nodes in tries; for the other two standard types the size is deterministic.

- Path Lengths L_n

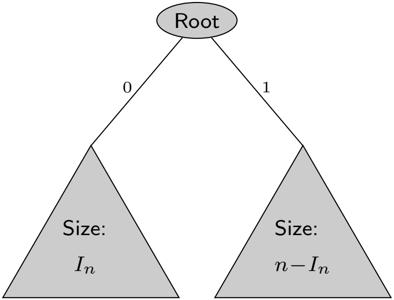
  Internal path length, external path length, total path length, weighted path length, peripheral path lengths, etc.

- Depth
- Number of leaves and patterns
- External and Internal Profiles
- Height
- Etc.

Additive Shape Parameters

$$X_{n+b} \stackrel{d}{=} X_{I_n} + X^*_{n-I_n} + T_n$$

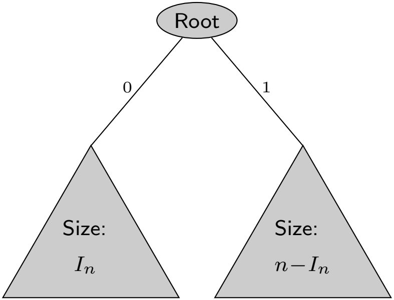
- $I_n \stackrel{d}{=} \text{Binomial}(n,p)$;
- $X_n \stackrel{d}{=} X^*_n$;
- $X_n, X^*_n, (I_n, T_n)$ independent.
- T_n toll-function.



Additive Shape Parameters

$$X_{n+b} \stackrel{d}{=} X_{I_n} + X^*_{n-I_n} + T_n$$

- $I_n \stackrel{d}{=}$ Binomial(n, p);
- $X_n \stackrel{d}{=} X^*_n$;
- $X_n, X^*_n, (I_n, T_n)$ independent.
- T_n toll-function.



Thus, moments satisfy the recurrence:

$$a_{n+b} = \sum_{j=0}^{n} \binom{n}{j} p^j q^{n-j} (a_j + a_{n-j}) + b_n.$$

Two (Main) Approaches

- **Rice-Noerlund Method**:

  Exercise 54 in Section 5.2.2 of Knuth's book. Developed into a systematic tool by Flajolet & Sedgewick (1986, 1995).

Two (Main) Approaches

- **Rice-Noerlund Method**:

  Exercise 54 in Section 5.2.2 of Knuth's book. Developed into a systematic tool by Flajolet & Sedgewick (1986, 1995).

- **Two-stage Approach**

  Introduced by Jacquet & Régnier (1988). Important contributions by Jacquet & Szpankowski (1998); Flajolet, Gourdon & Dumas (1995).

Two (Main) Approaches

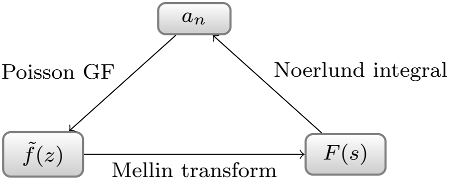
- **Rice-Noerlund Method**:

  Exercise 54 in Section 5.2.2 of Knuth's book. Developed into a systematic tool by Flajolet & Sedgewick (1986, 1995).

- **Two-stage Approach**

  Introduced by Jacquet & Régnier (1988). Important contributions by Jacquet & Szpankowski (1998); Flajolet, Gourdon & Dumas (1995).

Poisson-Mellin-Newton cycle:



Rice-Noerlund Method

Lemma (Noerlund)

Let C be a positive oriented, closed curve encircling $0, \dots, n$ and $f(z)$ be analytic within C. Then,

$$\sum_{k=0}^{n}(-1)^k\binom{n}{k}f(k) = -\frac{1}{2\pi i}\int_C f(s)\frac{n!\Gamma(-s)}{\Gamma(n+1-s)}\mathrm{d}s.$$

Rice-Noerlund Method

Lemma (Noerlund)

Let C be a positive oriented, closed curve encircling $0, \dots, n$ and $f(z)$ be analytic within C. Then,

$$\sum_{k=0}^{n}(-1)^k\binom{n}{k}f(k) = -\frac{1}{2\pi i}\int_C f(s)\frac{n!\Gamma(-s)}{\Gamma(n+1-s)}\mathrm{d}s.$$

For additive trie parameters:

$$f(s) = \frac{F(-s)}{\Gamma(-s)} = \frac{G(-s)}{(1-p^s-q^s)\Gamma(-s)}.$$

where

$$F(s) = \mathscr{M}\left[e^{-z}\sum_{n\geq 0}a_n z^n/n!; s\right], \quad G(s) = \mathscr{M}\left[e^{-z}\sum_{n\geq 0}b_n z^n/n!; s\right].$$

Mean of Size in Symmetric Tries (i)

We have, $f(s) = (s-1)/(1-2^{1-s})$.

Mean of Size in Symmetric Tries (i)

We have, $f(s) = (s-1)/(1-2^{1-s})$.

Theorem

For $n \geq 2$*,*

$$\mathbb{E}(S_n) = \frac{n}{\log 2} - \frac{1}{\log 2} \sum_{k \neq 0} c_k \frac{n!}{\Gamma(n+\chi_k)} - 1,$$

where

$$c_k = \chi_k \Gamma(-1+\chi_k).$$

with $\chi_k = 2k\pi i/(\log 2)$*.*

Mean of Size in Symmetric Tries (i)

We have, $f(s) = (s-1)/(1-2^{1-s})$.

Theorem

For $n \geq 2$,

$$\mathbb{E}(S_n) = \frac{n}{\log 2} - \frac{1}{\log 2} \sum_{k \neq 0} c_k \frac{n!}{\Gamma(n+\chi_k)} - 1,$$

where

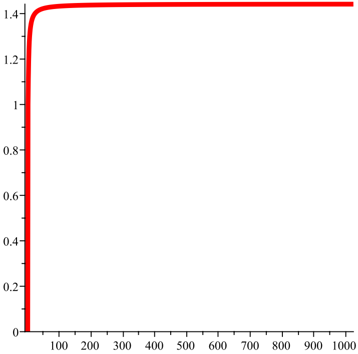
$$c_k = \chi_k \Gamma(-1+\chi_k).$$

with $\chi_k = 2k\pi i/(\log 2)$.

Note that

$$\frac{n!}{\Gamma(n+\chi_k)} \sim n^{1-\chi_k} + \frac{(1-\chi_k)\chi_k}{2} n^{-\chi_k} + \cdots .$$

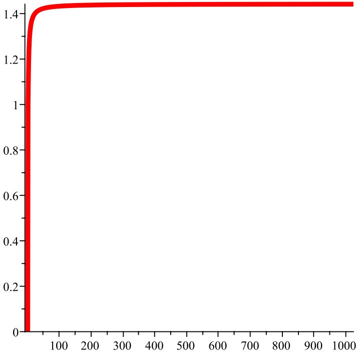
Mean of Size in Symmetric Tries (ii)



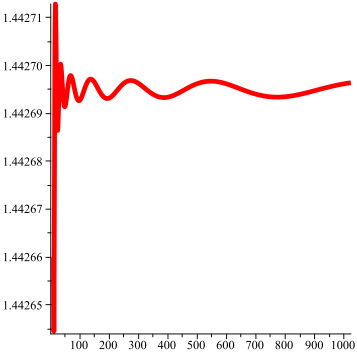
$$\frac{\mathbb{E}(S_n)}{n}$$

$$(n = 1, \ldots, 1024)$$

Mean of Size in Symmetric Tries (ii)



$$\frac{\mathbb{E}(S_n)}{n}$$
$$(n = 1, \ldots, 1024)$$



$$\frac{\mathbb{E}(S_n)+1}{n}$$
$$(n = 1, \ldots, 1024)$$

Variance via Rice-Noerlund Method

Symmetric digital trees: done by Kirschenhofer & Prodinger & Szpankowski in a series of papers in the beginning of 90s.

Variance via Rice-Noerlund Method

Symmetric digital trees: done by Kirschenhofer & Prodinger & Szpankowski in a series of papers in the beginning of 90s.

Second Moment Method:

- Use Rice-Noerlund twice for mean and second moment.

Variance via Rice-Noerlund Method

Symmetric digital trees: done by Kirschenhofer & Prodinger & Szpankowski in a series of papers in the beginning of 90s.

Second Moment Method:

- Use Rice-Noerlund twice for mean and second moment.
- Compute the variance via $\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.

Variance via Rice-Noerlund Method

Symmetric digital trees: done by Kirschenhofer & Prodinger & Szpankowski in a series of papers in the beginning of 90s.

Second Moment Method:

- Use Rice-Noerlund twice for mean and second moment.
- Compute the variance via $\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.

  Here, one needs identities of Ramanjuan: e.g. for α and β with $\alpha\beta = \pi^2$, we have

$$\sum_{k\geq 1} \frac{1}{k(e^{2\alpha k}-1)} - \frac{1}{4}\log\alpha + \frac{\alpha}{12} = \sum_{k\geq 1} \frac{1}{k(e^{2\beta k}-1)} - \frac{1}{4}\log\beta + \frac{\beta}{12}$$

Variance of Size in Symmetric Tries (i)

Theorem (Kirschenhofer & Prodinger)

We have,

$$\mathrm{Var}(S_n) \sim \frac{n}{\log 2} \sum_k G(-1+\chi_k) n^{-\chi_k},$$

Variance of Size in Symmetric Tries (i)

Theorem (Kirschenhofer & Prodinger)

We have,

$$\mathrm{Var}(S_n) \sim \frac{n}{\log 2} \sum_k G(-1+\chi_k) n^{-\chi_k},$$

where

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

Variance of TPL in Symmetric DSTs (i)

Theorem (Kirschenhofer, Prodinger, Szpankowski)

We have,

$$\mathrm{Var}(L_n) \sim n(C_{\mathit{var}} + P(\log_2 n)),$$

where $P(z)$ is a 1-periodic function with zero average value.

Variance of TPL in Symmetric DSTs (i)

Theorem (Kirschenhofer, Prodinger, Szpankowski)

We have,

$$\mathrm{Var}(L_n) \sim n(C_{var} + P(\log_2 n)),$$

where $P(z)$ is a 1-periodic function with zero average value.

To describe C_{var} we need the following:

- Let $L = \log 2$, $Q_j = \prod_{i=1}^{j}(1-2^{-i})$, $Q(1) = \lim_{j\to\infty} Q_j$.
- Put

$$\omega(z) = \frac{1}{L}\sum_{\ell\neq 0}\Gamma(-1-\chi_\ell)e^{2\ell\pi i z};$$

$$\psi(z) = -\frac{1}{L}\sum_{\ell\neq 0}\left(1-\frac{\chi_\ell}{2}\right)\Gamma(-\chi_\ell)e^{2\ell\pi i z}.$$

$$
\begin{aligned}
C_{\mathsf{var}} = &-\frac{28}{3L} - \frac{39}{4} - 2\sum_{\ell\ge1}\frac{\ell 2^\ell}{(2^\ell-1)^2} + \frac{2}{L}\sum_{\ell\ge1}\frac{1}{2^\ell-1} + \frac{\pi^2}{2L^2} + \frac{2}{L^2} \\
&-\frac{2}{L}\sum_{\ell\ge3}\frac{(-1)^{\ell+1}(\ell-5)}{(\ell+1)\ell(\ell-1)(2^\ell-1)} \\
&+\frac{2}{L}\sum_{\ell\ge1}(-1)^\ell 2^{-\binom{\ell+1}{2}}\left(\frac{L(1-2^{-\ell+1})/2-1}{1-2^{-\ell}} - \sum_{r\ge2}\frac{(-1)^{r+1}}{r(r-1)(2^{r+\ell}-1)}\right) \\
&-\frac{2Q(1)}{L} + \sum_{\ell\ge2}\frac{1}{2^\ell Q_\ell}\sum_{r\ge0}\frac{(-1)^r 2^{-\binom{r+1}{2}}}{Q_r}Q_{r+\ell-2}\cdot \\
&\cdot\left(-\sum_{j\ge1}\frac{1}{2^{j+r+\ell+2}-1}\left(2^{\ell+1}-2\ell-4+2\sum_{i=2}^{\ell-1}\binom{\ell+1}{i}\frac{1}{2^{r+i-1}-1}\right)\right. \\
&\quad+\frac{2}{(1-2^{-\ell-r})^2} + \frac{2\ell+2}{(1-2^{1-\ell-r})^2} - \frac{2}{L}\frac{1}{1-2^{1-\ell-r}} + \frac{2}{L}\sum_{j=1}^{\ell+1}\binom{\ell+1}{j}\frac{1}{2^{r+j}-1} \\
&\quad\left.-2\sum_{j=2}^{\ell+1}\binom{\ell+1}{j}\frac{1}{2^{r+j-1}-1} + \frac{2}{L}\sum_{j=0}^{\ell+1}\binom{\ell+1}{j}\sum_{i\ge1}\frac{(-1)^i}{(i+1)(2^{r+j+i}-1)}\right) \\
&+\sum_{\ell\ge3}\sum_{r=2}^{\ell-1}\binom{\ell+1}{r}\frac{Q_{r-2}Q_{\ell-r-1}}{2^\ell Q_\ell}\sum_{j\ge\ell+1}\frac{1}{2^j-1} - 2[\omega\psi]_0 - [\omega^2]_0.
\end{aligned}
$$

Bernoulli and Poisson Model

Bernoulli Model:

Digital tree is built from n records.

Bernoulli and Poisson Model

Bernoulli Model:

Digital tree is built from n records.

Poisson Model:

Digital tree is built from $\mathrm{Poi}(n)$ records.

Bernoulli and Poisson Model

Bernoulli Model:

Digital tree is built from n records.

Poisson Model:

Digital tree is built from $\mathrm{Poi}(n)$ records.

We have,

$$\mathbb{E}(X_{\mathrm{Poi}(n)}) = e^{-n} \sum_{j \geq 0} \mathbb{E}(X_j) \frac{n^j}{j!} = \tilde{f}_1(n).$$

Bernoulli and Poisson Model

Bernoulli Model:

Digital tree is built from n records.

Poisson Model:

Digital tree is built from $\mathrm{Poi}(n)$ records.

We have,

$$\mathbb{E}(X_{\mathrm{Poi}(n)}) = e^{-n} \sum_{j \geq 0} \mathbb{E}(X_j) \frac{n^j}{j!} = \tilde{f}_1(n).$$

Poisson Heuristic:

$$\tilde{f}_1(n) \text{ sufficiently smooth} \quad \Longrightarrow \quad \mathbb{E}(X_{\mathrm{Poi}(n)}) \approx \mathbb{E}(X_n).$$

Poisson-Charlier Expansion

Let

$$\tilde{f}(z) = e^{-z} \sum_{n \geq 0} a_n \frac{z^n}{n!}.$$

Poisson-Charlier Expansion

Let

$$\tilde{f}(z) = e^{-z} \sum_{n \geq 0} a_n \frac{z^n}{n!}.$$

Lemma (F., Hwang, Zacharovas)

Let $\tilde{f}(z)$ *be entire, then*

$$a_n = \sum_{j \geq 0} \frac{\tau_j(n)}{j!} \tilde{f}^{(j)}(n),$$

where

$$\tau_j(n) = n![z^n](z-n)^j e^z.$$

Poisson-Charlier Expansion

Let

$$\tilde{f}(z) = e^{-z} \sum_{n \geq 0} a_n \frac{z^n}{n!}.$$

Lemma (F., Hwang, Zacharovas)

Let $\tilde{f}(z)$ be entire, then

$$a_n = \sum_{j \geq 0} \frac{\tau_j(n)}{j!} \tilde{f}^{(j)}(n),$$

where

$$\tau_j(n) = n![z^n](z-n)^j e^z.$$

$\tau_0(z)$	$\tau_1(z)$	$\tau_2(z)$	$\tau_3(z)$	$\tau_4(z)$	$\tau_5(z)$
1	0	$-n$	$2n$	$3n(n-2)$	$-4n(5n-6)$

$\mathscr{J}\mathscr{S}$-admissibility

We say $\tilde{f}(z) \in \mathscr{J}\mathscr{S}_{\alpha,\beta}$ if

(**I**) Uniformly for $|\arg(z)| \le \epsilon$,

$$\tilde{f}(z) = \mathcal{O}\big(|z|^{\alpha} \log^{\beta} |z|\big).$$

(**O**) Uniformly for $\epsilon < |\arg(z)| \le \pi$,

$$f(z) := e^{z} \tilde{f}(z) = \mathcal{O}\big(e^{(1-\epsilon)|z|}\big).$$

$\mathscr{J}\mathscr{S}$-admissibility

We say $\tilde{f}(z) \in \mathscr{J}\mathscr{S}_{\alpha,\beta}$ if

(I) Uniformly for $|\arg(z)| \le \epsilon$,

$$\tilde{f}(z) = \mathcal{O}\big(|z|^{\alpha} \log^{\beta} |z|\big).$$

(O) Uniformly for $\epsilon < |\arg(z)| \le \pi$,

$$f(z) := e^{z}\tilde{f}(z) = \mathcal{O}\Big(e^{(1-\epsilon)|z|}\Big).$$

Theorem (Jacquet & Szpankowski)

If $\tilde{f}(z) \in \mathscr{J}\mathscr{S}_{\alpha,\beta}$, *then*

$$a_n \sim \sum_{j\ge 0} \frac{\tau_j(n)}{j!} \tilde{f}^{(j)}(n).$$

Closure Properties of $\mathscr{J}\mathscr{S}$-admissibility

Proposition

Let $\alpha \in (0,1)$*.*

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

(ii) *If* $\tilde{f} \in \mathscr{J}\mathscr{S}$*, then* $\tilde{f}(\alpha z), z^m\tilde{f} \in \mathscr{J}\mathscr{S}$.

(iii) *If* $\tilde{f}, \tilde{g} \in \mathscr{J}\mathscr{S}$*, then* $\tilde{f} + \tilde{g} \in \mathscr{J}\mathscr{S}$.

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

(v) *If* $\tilde{f} \in \mathscr{J}\mathscr{S}$*, then* $\tilde{f}' \in \mathscr{J}\mathscr{S}$.

Closure Properties of $\mathscr{J}\mathscr{S}$-admissibility

Proposition

Let $\alpha \in (0,1)$.

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

(ii) *If* $\tilde{f} \in \mathscr{J}\mathscr{S}$, *then* $\tilde{f}(\alpha z), z^m \tilde{f} \in \mathscr{J}\mathscr{S}$.

(iii) *If* $\tilde{f}, \tilde{g} \in \mathscr{J}\mathscr{S}$, *then* $\tilde{f} + \tilde{g} \in \mathscr{J}\mathscr{S}$.

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

(v) *If* $\tilde{f} \in \mathscr{J}\mathscr{S}$, *then* $\tilde{f}' \in \mathscr{J}\mathscr{S}$.

Proposition

Let $\sum_{j=0}^{b} \binom{b}{j} \tilde{f}(z)^{(j)} = 2\tilde{f}(z/2) + \tilde{g}(z)$. *Then,*

$$\tilde{g}(z) \in \mathscr{J}\mathscr{S} \qquad \Longleftrightarrow \qquad \tilde{f}(z) \in \mathscr{J}\mathscr{S}.$$

Two Stage Approach for Tries (and PATRICIA Tries)

- Uses first poissonization:

$$\tilde{f}(z) = e^{-z} \sum_{n \geq 0} a_n \frac{z^n}{n!}.$$

Two Stage Approach for Tries (and PATRICIA Tries)

- Uses first poissonization:

$$\tilde{f}(z) = e^{-z} \sum_{n\geq 0} a_n \frac{z^n}{n!}.$$

- $\tilde{f}(z)$ satisfies:

$$\tilde{f}(z) = 2\tilde{f}(z/2) + \tilde{g}(z)$$

which is solved with the Mellin transform:

$$F(s) = \int_0^\infty \tilde{f}(z) z^{s-1} \mathrm{d}z = \frac{G(s)}{1 - 2^{1+s}}.$$

Two Stage Approach for Tries (and PATRICIA Tries)

- Uses first poissonization:

$$\tilde{f}(z) = e^{-z} \sum_{n\geq 0} a_n \frac{z^n}{n!}.$$

- $\tilde{f}(z)$ satisfies:

$$\tilde{f}(z) = 2\tilde{f}(z/2) + \tilde{g}(z)$$

  which is solved with the Mellin transform:

$$F(s) = \int_0^\infty \tilde{f}(z) z^{s-1} \mathrm{d}z = \frac{G(s)}{1-2^{1+s}}.$$

- Asymptotics of $\tilde{f}(z)$ is obtained via inverse Mellin transform (Flajolet & Gourdon & Dumas, 1995).

Two Stage Approach for Tries (and PATRICIA Tries)

- Uses first poissonization:

$$\tilde{f}(z) = e^{-z} \sum_{n \geq 0} a_n \frac{z^n}{n!}.$$

- $\tilde{f}(z)$ satisfies:

$$\tilde{f}(z) = 2\tilde{f}(z/2) + \tilde{g}(z)$$

  which is solved with the Mellin transform:

$$F(s) = \int_0^\infty \tilde{f}(z) z^{s-1} \mathrm{d}z = \frac{G(s)}{1 - 2^{1+s}}.$$

- Asymptotics of $\tilde{f}(z)$ is obtained via inverse Mellin transform (Flajolet & Gourdon & Dumas, 1995).
- Asymptotics of a_n is obtained via (analytic) depoissonization (Jacquet & Szpankowski, 1998).

Asymptotic Poissonized Variances

Let

$$\tilde{f}_1(z) = e^{-z} \sum_{n \geq 0} \mathbb{E}(X_n) \frac{z^n}{n!}, \quad \tilde{f}_2(z) = e^{-z} \sum_{n \geq 0} \mathbb{E}(X_n^2) \frac{z^n}{n!}.$$

Asymptotic Poissonized Variances

Let

$$\tilde{f}_1(z) = e^{-z} \sum_{n\geq 0} \mathbb{E}(X_n) \frac{z^n}{n!}, \quad \tilde{f}_2(z) = e^{-z} \sum_{n\geq 0} \mathbb{E}(X_n^2) \frac{z^n}{n!}.$$

Variance of Poisson model:

$$\tilde{V}_1(n) := \mathrm{Var}(X_{\mathrm{Poi}(n)}) = \tilde{f}_2(n) - \tilde{f}_1(n)^2.$$

Asymptotic Poissonized Variances

Let

$$\tilde{f}_1(z) = e^{-z} \sum_{n\geq 0} \mathbb{E}(X_n) \frac{z^n}{n!}, \quad \tilde{f}_2(z) = e^{-z} \sum_{n\geq 0} \mathbb{E}(X_n^2) \frac{z^n}{n!}.$$

Variance of Poisson model:

$$\tilde{V}_1(n) := \mathrm{Var}(X_{\mathrm{Poi}(n)}) = \tilde{f}_2(n) - \tilde{f}_1(n)^2.$$

However, for many shape parameters:

$$\mathrm{Var}(X_n) \not\sim \tilde{V}_1(n),$$

i.e., Poisson model unsuitable for obtaining asymptotics of the variance!

Asymptotic Poissonized Variances

Let

$$\tilde{f}_1(z) = e^{-z} \sum_{n\ge 0} \mathbb{E}(X_n) \frac{z^n}{n!}, \quad \tilde{f}_2(z) = e^{-z} \sum_{n\ge 0} \mathbb{E}(X_n^2) \frac{z^n}{n!}.$$

Variance of Poisson model:

$$\tilde{V}_1(n) := \mathrm{Var}(X_{\mathrm{Poi}(n)}) = \tilde{f}_2(n) - \tilde{f}_1(n)^2.$$

However, for many shape parameters:

$$\mathrm{Var}(X_n) \not\sim \tilde{V}_1(n),$$

i.e., Poisson model unsuitable for obtaining asymptotics of the variance!

F., Hwang, Zacharovas (2010):

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

$\tilde{V}_1(z)$ vs. $\tilde{V}_2(z)$

$\tilde{V}_1(z)$	– depth; – leader/loser selection; – Paper-Stone-Scissors and variants; – approximate counting; – corner parking; – partial match queries in k-d digital trees; – node profile of asymmetric digital trees.
$\tilde{V}_2(z)$	– size; – path length; – occurrence of pattern; – node profile of symmetric digital trees; – Wiener and Steiner index.

Variance of TPL in Symmetric DSTs (ii)

Theorem (F., Hwang, Zacharovas)

We, have

$$\mathrm{Var}(L_n) \sim n(C_{var} + P(\log_2 n)),$$

where $P(z)$ *is a 1-periodic function with zero average value and*

$$C_{var} = \frac{Q(1)}{L} \sum_{j,h,\ell \geq 0} \frac{(-1)^j 2^{-\binom{j+1}{2}}}{Q_j Q_h Q_\ell 2^{h+\ell}} \varphi(2^{-j-h} + 2^{-j-\ell}),$$

where

$$\varphi(x) := \begin{cases} (x - \log x - 1)/(x-1)^2, & \text{if } x \neq 1; \\ 1/2, & \text{if } x = 1. \end{cases}$$

$$
\begin{aligned}
C_{\mathsf{var}} = &-\frac{28}{3L} - \frac{39}{4} - 2\sum_{\ell\ge1}\frac{\ell 2^\ell}{(2^\ell-1)^2} + \frac{2}{L}\sum_{\ell\ge1}\frac{1}{2^\ell-1} + \frac{\pi^2}{2L^2} + \frac{2}{L^2}\\
&-\frac{2}{L}\sum_{\ell\ge3}\frac{(-1)^{\ell+1}(\ell-5)}{(\ell+1)\ell(\ell-1)(2^\ell-1)}\\
&+\frac{2}{L}\sum_{\ell\ge1}(-1)^\ell 2^{-\binom{\ell+1}{2}}\left(\frac{L(1-2^{-\ell+1})/2-1}{1-2^{-\ell}} - \sum_{r\ge2}\frac{(-1)^{r+1}}{r(r-1)(2^{r+\ell}-1)}\right)\\
&-\frac{2Q(1)}{L} + \sum_{\ell\ge2}\frac{1}{2^\ell Q_\ell}\sum_{r\ge0}\frac{(-1)^r 2^{-\binom{r+1}{2}}}{Q_r}Q_{r+\ell-2}\cdot\\
&\cdot\left(-\sum_{j\ge1}\frac{1}{2^{j+r+\ell+2}-1}\left(2^{\ell+1}-2\ell-4+2\sum_{i=2}^{\ell-1}\binom{\ell+1}{i}\frac{1}{2^{r+i-1}-1}\right)\right.\\
&\quad+\frac{2}{(1-2^{-\ell-r})^2} + \frac{2\ell+2}{(1-2^{1-\ell-r})^2} - \frac{2}{L}\frac{1}{1-2^{1-\ell-r}} + \frac{2}{L}\sum_{j=1}^{\ell+1}\binom{\ell+1}{j}\frac{1}{2^{r+j}-1}\\
&\quad\left.-2\sum_{j=2}^{\ell+1}\binom{\ell+1}{j}\frac{1}{2^{r+j-1}-1} + \frac{2}{L}\sum_{j=0}^{\ell+1}\binom{\ell+1}{j}\sum_{i\ge1}\frac{(-1)^i}{(i+1)(2^{r+j+i}-1)}\right)\\
&+\sum_{\ell\ge3}\sum_{r=2}^{\ell-1}\binom{\ell+1}{r}\frac{Q_{r-2}Q_{\ell-r-1}}{2^\ell Q_\ell}\sum_{j\ge\ell+1}\frac{1}{2^j-1} - 2[\omega\psi]_0 - [\omega^2]_0.
\end{aligned}
$$

General Framework for Symmetric Tries

Consider $X_n \stackrel{d}{=} X_{I_n} + X^*_{n-I_n} + T_n$ and let

$$\tilde{g}_1(z) = e^{-z} \sum_{n\geq 0} \mathbb{E}(T_n) \frac{z^n}{n!}, \quad \tilde{g}_2(z) = e^{-z} \sum_{n\geq 0} \mathbb{E}(T_n^2) \frac{z^n}{n!}$$

and $\tilde{W}_2(z) := \tilde{g}_2(z) - \tilde{g}_1(z)^2 - z\tilde{g}_1'(z)$.

General Framework for Symmetric Tries

Consider $X_n \overset{d}{=} X_{I_n} + X^*_{n-I_n} + T_n$ and let

$$\tilde{g}_1(z) = e^{-z} \sum_{n\geq 0} \mathbb{E}(T_n) \frac{z^n}{n!}, \quad \tilde{g}_2(z) = e^{-z} \sum_{n\geq 0} \mathbb{E}(T_n^2) \frac{z^n}{n!}$$

and $\tilde{W}_2(z) := \tilde{g}_2(z) - \tilde{g}_1(z)^2 - z\tilde{g}_1'(z)$.

Theorem (F., Hwang, Zacharovas)

Let $\tilde{g}_1, \tilde{W}_2 \in \mathscr{J}\mathscr{S}_{\alpha,\beta}$ with $\alpha < 1$ and $\tilde{g}_2 \in \mathscr{J}\mathscr{S}$. Then,

$$\mathbb{E}(X_n) \sim nP_1(\log_2 n), \qquad \mathrm{Var}(X_n) \sim nP_2(\log_2 n),$$

where P_1, P_2 are 1-periodic, computable functions.

Variance of Size in Symmetric Tries (ii)

Theorem (F., Hwang, Zacharovas)

We have,

$$\mathrm{Var}(S_n) \sim \frac{n}{\log 2} \sum_{k} G(-1+\chi_k) n^{-\chi_k},$$

where

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

with $\chi_k = 2k\pi i/(\log 2)$.

Exact Poissonized Variance

Question: why not using the PGF of the variance, i.e.,

$$\tilde{V}_\infty(z) := e^{-z} \sum_{n\geq 0} \mathrm{Var}(X_n) \frac{z^n}{n!}?$$

Exact Poissonized Variance

Question: why not using the PGF of the variance, i.e.,

$$\tilde{V}_\infty(z) := e^{-z} \sum_{n\geq 0} \mathrm{Var}(X_n)\frac{z^n}{n!}?$$

Recall that

$$X_n \stackrel{d}{=} X_{I_n} + X^*_{n-I_n} + T_n.$$

Then,

$$\sigma_n^2 = 2^{1-n}\sum_{j=0}^{n}\binom{n}{j}\sigma_j^2 + \mathrm{Var}(T_n) + 2^{-n}\sum_{j=0}^{n}\binom{n}{j}\left(\mathbb{E}(X_j)+\mathbb{E}(X_{n-j})-\mathbb{E}(X_n)+\mathbb{E}(T_n)\right)^2,$$

where $\sigma_n^2 = \mathrm{Var}(X_n)$.

Poissonization of the Variance Recurrence

Set

$$\tilde{W}_\infty(z) := e^{-z} \sum_{n\geq 0} \mathrm{Var}(T_n) \frac{z^n}{n!}.$$

Poissonization of the Variance Recurrence

Set

$$\tilde{W}_\infty(z) := e^{-z} \sum_{n\geq 0} \mathrm{Var}(T_n) \frac{z^n}{n!}.$$

Proposition

We have,

$$\tilde{V}_\infty(z) = 2\tilde{V}_\infty(z/2) + \tilde{W}_\infty(z) + \tilde{h}(z),$$

where

$$\begin{aligned} \tilde{h}(z) =& 2(\tilde{f}_1 \odot \tilde{f}_1)(z/2) + 2\tilde{f}_1(z/2)^2 \\ & - (\tilde{f}_1 \odot \tilde{f}_1)(z) + 2(\tilde{f}_1 \odot \tilde{g}_1)(z) - (\tilde{g}_1 \odot \tilde{g}_1)(z). \end{aligned}$$

Poissonization of the Variance Recurrence

Set

$$\tilde{W}_\infty(z) := e^{-z} \sum_{n\geq 0} \mathrm{Var}(T_n) \frac{z^n}{n!}.$$

Proposition

We have,

$$\tilde{V}_\infty(z) = 2\tilde{V}_\infty(z/2) + \tilde{W}_\infty(z) + \tilde{h}(z),$$

where

$$\begin{aligned}\tilde{h}(z) =& 2(\tilde{f}_1 \odot \tilde{f}_1)(z/2) + 2\tilde{f}_1(z/2)^2 \\ & - (\tilde{f}_1 \odot \tilde{f}_1)(z) + 2(\tilde{f}_1 \odot \tilde{g}_1)(z) - (\tilde{g}_1 \odot \tilde{g}_1)(z).\end{aligned}$$

Here, $\tilde{f} \odot \tilde{g}$ denotes the **Hadamard product** of $\tilde{f}$ and $\tilde{g}$.

Closure Property of Hadamard Product

Proposition (F., Hwang, Zacharovas)

If

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

are both $\mathscr{J}\mathscr{S}$*-admissible, then*

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

is also $\mathscr{J}\mathscr{S}$*-admissible.*

Closure Property of Hadamard Product

Proposition (F., Hwang, Zacharovas)

If

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

are both $\mathscr{J}\mathscr{S}$*-admissible, then*

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

is also $\mathscr{J}\mathscr{S}$*-admissible.*

Proof uses the integral representation

$$(\tilde{f} \odot \tilde{g})(z^2) = \frac{e^{-z^2}}{4\pi i} \int_0^\infty u e^{-u^2} \oint_{|\xi|=1} e^{zu\xi + zu\bar{\xi}} \tilde{f}(zu\xi)\tilde{g}(zu\bar{\xi}) \frac{d\xi}{\xi} du.$$

An IFF Condition for $\mathscr{J}\mathscr{S}$-admissibility

Theorem (Jacquet)

$\tilde{f} \in \mathscr{J}\mathscr{S}$ iff there exists an analytic extension $\alpha(z)$ of a_n with

$$\alpha(z) = \mathcal{O}(z^p), \qquad \text{for some } p > 0,$$

as $z \to \infty$, in a cone $\{z : |\arg(z)| \le \epsilon\}$.

An IFF Condition for $\mathscr{J}\mathscr{S}$-admissibility

Theorem (Jacquet)

$\tilde{f} \in \mathscr{J}\mathscr{S}$ iff there exists an analytic extension $\alpha(z)$ of a_n with

$$\alpha(z) = \mathcal{O}(z^p), \qquad \text{for some } p > 0,$$

as $z \to \infty$, in a cone $\{z : |\arg(z)| \le \epsilon\}$.

Was proved in two stages:

An IFF Condition for $\mathscr{J}\mathscr{S}$-admissibility

Theorem (Jacquet)

$\tilde{f} \in \mathscr{J}\mathscr{S}$ iff there exists an analytic extension $\alpha(z)$ of a_n with

$$\alpha(z) = \mathcal{O}(z^p), \qquad \text{for some } p > 0,$$

as $z \to \infty$, in a cone $\{z : |\arg(z)| \leq \epsilon\}$.

Was proved in two stages:

- "If": Jacquet & Szpankowski (1999);

An IFF Condition for $\mathscr{J}\mathscr{S}$-admissibility

Theorem (Jacquet)

$\tilde{f} \in \mathscr{J}\mathscr{S}$ iff there exists an analytic extension $\alpha(z)$ of a_n with

$$\alpha(z) = \mathcal{O}(z^p), \qquad \text{for some } p > 0,$$

as $z \to \infty$, in a cone $\{z : |\arg(z)| \leq \epsilon\}$.

Was proved in two stages:

- "If": Jacquet & Szpankowski (1999);
- "Only If": Jacquet (2014).

An IFF Condition for $\mathscr{J}\mathscr{S}$-admissibility

Theorem (Jacquet)

$\tilde{f} \in \mathscr{J}\mathscr{S}$ iff there exists an analytic extension $\alpha(z)$ of a_n with

$$\alpha(z) = \mathcal{O}(z^p), \qquad \text{for some } p > 0,$$

as $z \to \infty$, in a cone $\{z : |\arg(z)| \leq \epsilon\}$.

Was proved in two stages:

- "If": Jacquet & Szpankowski (1999);
- "Only If": Jacquet (2014).

Corollary

If $\tilde{f}, \tilde{g}$ are $\mathscr{J}\mathscr{S}$-admissible, then $\tilde{f} \odot \tilde{g}$ is $\mathscr{J}\mathscr{S}$-admissible.

Asymptotic Expansion for Hadamard Product

Theorem

If $\tilde{f}, \tilde{g} \in \mathscr{J}\mathscr{S}$, *then*

$$(\tilde{f} \odot \tilde{g})(z) \sim \sum_{n \geq 0} \frac{z^k}{k!} \tilde{f}^{(k)}(z) \tilde{g}^{(k)}(z).$$

Asymptotic Expansion for Hadamard Product

Theorem

If $\tilde{f}, \tilde{g} \in \mathscr{J}\mathscr{S}$, then

$$(\tilde{f} \odot \tilde{g})(z) \sim \sum_{n\geq 0} \frac{z^k}{k!} \tilde{f}^{(k)}(z)\tilde{g}^{(k)}(z).$$

We have,

$$\tilde{V}_\infty(z) = 2\tilde{V}_\infty(z/2) + \tilde{W}_\infty(z) + \tilde{h}(z),$$

where

$$\tilde{h}(z) \sim \sum_{k\geq 2} \frac{c_k}{k!} \left(\frac{z}{2}\right)^k \tilde{f}_1^{(k)}(z/2)^2$$

with

$$c_k = 2(1 - 2^{1-k}).$$

General Framework for Symmetric Tries - Revisited

Recall: $X_n \stackrel{d}{=} X_{I_n} + X^*_{n-I_n} + T_n$.

Theorem (F., Hwang, Zacharovas)

Let $\tilde{g}_1, \tilde{W}_2 \in \mathscr{J}\mathscr{S}_{\alpha,\beta}$ *with* $\alpha < 1$ *and* $\tilde{g}_2 \in \mathscr{J}\mathscr{S}$. *Then,*

$$\mathbb{E}(X_n) \sim nP_1(\log_2 n), \qquad \mathrm{Var}(X_n) \sim nP_2(\log_2 n),$$

where P_1, P_2 *are* 1*-periodic, computable functions.*

General Framework for Symmetric Tries - Revisited

Recall: $X_n \stackrel{d}{=} X_{I_n} + X^*_{n-I_n} + T_n$.

Theorem (F., Hwang, Zacharovas)

Let $\tilde{g}_1, \tilde{W}_2 \in \mathscr{J}\mathscr{S}_{\alpha,\beta}$ *with* $\alpha < 1$ *and* $\tilde{g}_2 \in \mathscr{J}\mathscr{S}$. *Then,*

$$\mathbb{E}(X_n) \sim nP_1(\log_2 n), \qquad \mathrm{Var}(X_n) \sim nP_2(\log_2 n),$$

where P_1, P_2 *are* 1*-periodic, computable functions.*

Theorem

Let $\tilde{g}_1, \tilde{W}_\infty \in \mathscr{J}\mathscr{S}_{\alpha,\beta}$ *with* $\alpha < 1$. *Then,*

$$\mathbb{E}(X_n) \sim nP_1(\log_2 n), \qquad \mathrm{Var}(X_n) \sim nP_2(\log_2 n),$$

where P_1, P_2 *are* 1*-periodic, computable functions.*

Variance of Size in Symmetric Tries (iii)

Theorem

We have,

$$\mathrm{Var}(S_n) \sim \frac{1}{\log 2} P_1(n) - \frac{1}{\log^2 2} \sum_{k\geq 2} \frac{1}{k!} P_k(n),$$

where

$$P_k(n) = \sum_{\ell} d_{k,\ell} \frac{n!}{\Gamma(n+k-1+\chi_\ell)}$$

with $d_{1,\ell}$ as before and for $k \geq 2$

$$d_{k,\ell} = \sum_{j} \chi_j \Gamma(k-1+\chi_j) \chi_{\ell-j} \Gamma(k-1+\chi_{\ell-j}).$$

Variance of Size in Symmetric Tries (iii)

Theorem

We have,

$$\mathrm{Var}(S_n) \sim \frac{1}{\log 2} P_1(n) - \frac{1}{\log^2 2} \sum_{k \geq 2} \frac{1}{k!} P_k(n),$$

where

$$P_k(n) = \sum_{\ell} d_{k,\ell} \frac{n!}{\Gamma(n+k-1+\chi_\ell)}$$

with $d_{1,\ell}$ *as before and for* $k \geq 2$

$$d_{k,\ell} = \sum_j \chi_j \Gamma(k-1+\chi_j) \chi_{\ell-j} \Gamma(k-1+\chi_{\ell-j}).$$

Remark: Similar results for other (additive) shape parameters of symmetric tries and PATRICIA tries.

Identity for Hadamard Product

Theorem

Under assumptions which include $\mathscr{J}\mathscr{S}$-admissibility, we have

$$(\tilde{f} \odot \tilde{g})(z) = \sum_{k\geq 0} \frac{z^k}{k!} \tilde{f}^{(k)}(z)\tilde{g}^{(k)}(z).$$

Identity for Hadamard Product

Theorem

Under assumptions which include $\mathscr{J}\mathscr{S}$-admissibility, we have

$$(\tilde{f} \odot \tilde{g})(z) = \sum_{k\geq 0} \frac{z^k}{k!} \tilde{f}^{(k)}(z) \tilde{g}^{(k)}(z).$$

Equivalently, we have:

$$\left(\sum_{j\geq 0} \frac{\tau_j(n)}{j!} \tilde{f}^{(j)}(n)\right)\left(\sum_{j\geq 0} \frac{\tau_j(n)}{j!} \tilde{g}^{(j)}(n)\right) = \sum_{j\geq 0} \frac{\tau_j(n)}{j!} \left(\sum_{k\geq 0} \frac{n^k}{k!} \tilde{f}^{(k)}(n) \tilde{g}(k)(n)\right)^{(j)}.$$

A Quick Proof (i)

Let

$$\tilde{f}(z) = \sum_{m\geq 0} \frac{A_m}{m!} z^m, \qquad \tilde{g}(z) = \sum_{m\geq 0} \frac{B_m}{m!} z^m.$$

A Quick Proof (i)

Let

$$\tilde{f}(z) = \sum_{m\geq 0} \frac{A_m}{m!} z^m, \qquad \tilde{g}(z) = \sum_{m\geq 0} \frac{B_m}{m!} z^m.$$

Then,

$$\left(\sum_{j\geq 0} \frac{\tau_j(n)}{j!} \tilde{f}^{(j)}(n)\right)\left(\sum_{j\geq 0} \frac{\tau_j(n)}{j!} \tilde{g}^{(j)}(n)\right)$$

$$= \sum_{j\geq 0} \frac{\tau_j(n)}{j!} \left(\sum_{k\geq 0} \frac{n^k}{k!} \tilde{f}^{(k)}(n) \tilde{g}(k)(n)\right)^{(j)}$$

can be rewritten to

$$\left(\sum_{0\leq m\leq n} \binom{n}{m} A_m\right)\left(\sum_{0\leq m\leq n} \binom{n}{m} B_m\right) = n![z^n]e^z \sum_{k\geq 0} \frac{z^k}{k!} \tilde{f}^{(k)}(z)\tilde{g}^{(k)}(z).$$

A Quick Proof (ii)

Note that

$$n![z^n]e^z \sum_{k\geq 0} \frac{z^k}{k!} \tilde{f}^{(k)}(z)\tilde{g}^{(k)}(z) = n! \sum_{k\geq 0} \frac{1}{k!}[z^n]z^k e^z \tilde{f}^{(k)}(z)\tilde{g}^{(k)}(z)$$

A Quick Proof (ii)

Note that

$$n![z^n]e^z\sum_{k\geq 0}\frac{z^k}{k!}\tilde{f}^{(k)}(z)\tilde{g}^{(k)}(z)=n!\sum_{k\geq 0}\frac{1}{k!}[z^n]z^ke^z\tilde{f}^{(k)}(z)\tilde{g}^{(k)}(z)$$

$$=n!\sum_{0\leq k\leq n}\frac{1}{k!}\sum_{i_1+i_2+i_3=n-k}\frac{A_{i_2+k}B_{i_3+k}}{i_1!i_2!i_3!}$$

A Quick Proof (ii)

Note that

$$n![z^n]e^z\sum_{k\ge0}\frac{z^k}{k!}\tilde{f}^{(k)}(z)\tilde{g}^{(k)}(z)=n!\sum_{k\ge0}\frac{1}{k!}[z^n]z^ke^z\tilde{f}^{(k)}(z)\tilde{g}^{(k)}(z)$$

$$=n!\sum_{0\le k\le n}\frac{1}{k!}\sum_{i_1+i_2+i_3=n-k}\frac{A_{i_2+k}B_{i_3+k}}{i_1!i_2!i_3!}$$

$$=\left(\sum_{0\le m\le n}\binom{n}{m}A_m\right)\left(\sum_{0\le m\le n}\binom{n}{m}B_m\right)$$

A Quick Proof (ii)

Note that

$$n![z^n]e^z\sum_{k\geq 0}\frac{z^k}{k!}\tilde{f}^{(k)}(z)\tilde{g}^{(k)}(z)=n!\sum_{k\geq 0}\frac{1}{k!}[z^n]z^ke^z\tilde{f}^{(k)}(z)\tilde{g}^{(k)}(z)$$

$$=n!\sum_{0\leq k\leq n}\frac{1}{k!}\sum_{i_1+i_2+i_3=n-k}\frac{A_{i_2+k}B_{i_3+k}}{i_1!i_2!i_3!}$$

$$=\left(\sum_{0\leq m\leq n}\binom{n}{m}A_m\right)\left(\sum_{0\leq m\leq n}\binom{n}{m}B_m\right)$$

and this is true since

$$\binom{n}{s}\binom{n}{t}=\sum_{0\leq k\leq n}\binom{n}{k}\binom{n-k}{s-k,t-k,n+k-s-t}$$

which counts twice the number of ways of choosing a set of size s and t from a set of size n.

Exact vs. Asymptotic Poissonized Variance

Recall

$$\tilde{V}_\infty(z) = e^{-z} \sum_{n \geq 0} \mathrm{Var}(X_n) \frac{z^n}{n!}.$$

Exact vs. Asymptotic Poissonized Variance

Recall

$$\tilde{V}_{\infty}(z) = e^{-z} \sum_{n \geq 0} \mathrm{Var}(X_n) \frac{z^n}{n!}.$$

Corollary

We have,

$$\tilde{V}_{\infty}(z) = \tilde{f}_2(z) - \sum_{k \geq 0} \frac{z^k}{k!} \tilde{f}_1^{(k)}(z)^2.$$

Exact vs. Asymptotic Poissonized Variance

Recall

$$\tilde{V}_{\infty}(z) = e^{-z} \sum_{n\geq 0} \mathrm{Var}(X_n) \frac{z^n}{n!}.$$

Corollary

We have,

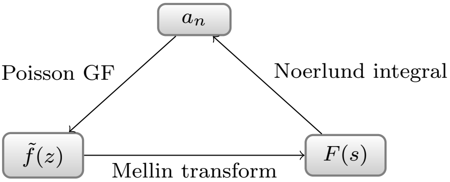
$$\tilde{V}_{\infty}(z) = \tilde{f}_2(z) - \sum_{k\geq 0} \frac{z^k}{k!} \tilde{f}_1^{(k)}(z)^2.$$

This gives:

$$\tilde{V}_{\infty}(z) = \underbrace{\underbrace{\tilde{f}_2(z) - \tilde{f}_1(z)^2}_{=\tilde{V}_1(z)} - z\tilde{f}_1'(z)^2}_{=\tilde{V}_2(z)} - \frac{z^2}{2}\tilde{f}_1''(z)^2 - \cdots$$

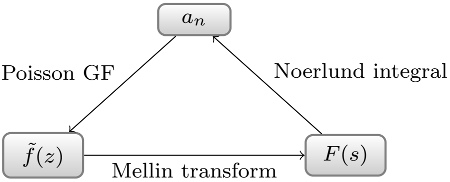
Summary

Poisson-Mellin-Newton cycle:



Summary

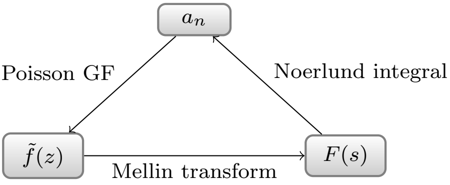
Poisson-Mellin-Newton cycle:



- We used exact poissonized variance in contrast to asymptotic poissonized variances.

Summary

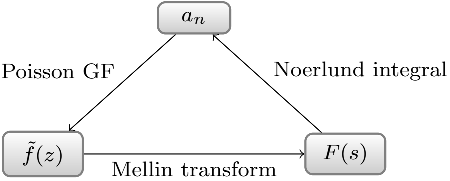
Poisson-Mellin-Newton cycle:



- We used exact poissonized variance in contrast to asymptotic poissonized variances.
- This yields general frameworks for asymptotics of mean and variance of additive shape parameter in tries and PATRICIA tries under natural conditions.

Summary

Poisson-Mellin-Newton cycle:



- We used exact poissonized variance in contrast to asymptotic poissonized variances.
- This yields general frameworks for asymptotics of mean and variance of additive shape parameter in tries and PATRICIA tries under natural conditions.
- This also yields full asymptotic expansions of the variance for symmetric tries and PATRICIA tries.

Inverse Mellin and De-poissonization Combined

Formally, we have

$$a_n = \frac{n!}{2\pi i}\int_{|z|=r}\frac{e^z\tilde{f}(z)}{z^{n+1}}$$

Inverse Mellin and De-poissonization Combined

Formally, we have

$$
\begin{aligned}
a_n &= \frac{n!}{2\pi i}\int_{|z|=r}\frac{e^z\tilde{f}(z)}{z^{n+1}}\\
&= \frac{n!}{2\pi i}\int_{|z|=r}\frac{e^z}{z^{n+1}}\left(\frac{1}{2\pi i}\int_{(c)}F(s)z^{-s}\mathrm{d}s\right)\mathrm{d}z
\end{aligned}
$$

Inverse Mellin and De-poissonization Combined

Formally, we have

$$\begin{aligned} a_n &= \frac{n!}{2\pi i}\int_{|z|=r}\frac{e^z\tilde{f}(z)}{z^{n+1}} \\ &= \frac{n!}{2\pi i}\int_{|z|=r}\frac{e^z}{z^{n+1}}\left(\frac{1}{2\pi i}\int_{(c)}F(s)z^{-s}\mathrm{d}s\right)\mathrm{d}z \\ &= \frac{n!}{2\pi i}\int_{(c)}F(s)\left(\frac{1}{2\pi i}\int_{|z|=r}e^z z^{-n-s-1}\mathrm{d}z\right)\mathrm{d}s \end{aligned}$$

Inverse Mellin and De-poissonization Combined

Formally, we have

$$
\begin{aligned}
a_n &= \frac{n!}{2\pi i}\int_{|z|=r}\frac{e^z\tilde{f}(z)}{z^{n+1}}\\
&= \frac{n!}{2\pi i}\int_{|z|=r}\frac{e^z}{z^{n+1}}\left(\frac{1}{2\pi i}\int_{(c)}F(s)z^{-s}\mathrm{d}s\right)\mathrm{d}z\\
&= \frac{n!}{2\pi i}\int_{(c)}F(s)\left(\frac{1}{2\pi i}\int_{|z|=r}e^z z^{-n-s-1}\mathrm{d}z\right)\mathrm{d}s\\
&= \frac{1}{2\pi i}\int_{(c)}\frac{F(s)}{\Gamma(s)}\frac{n!\Gamma(s)}{\Gamma(n+1+s)}\mathrm{d}s
\end{aligned}
$$

Inverse Mellin and De-poissonization Combined

Formally, we have

$$\begin{aligned} a_n &= \frac{n!}{2\pi i}\int_{|z|=r} \frac{e^z \tilde{f}(z)}{z^{n+1}} \\ &= \frac{n!}{2\pi i}\int_{|z|=r} \frac{e^z}{z^{n+1}} \left(\frac{1}{2\pi i}\int_{(c)} F(s) z^{-s}\mathrm{d}s\right)\mathrm{d}z \\ &= \frac{n!}{2\pi i}\int_{(c)} F(s) \left(\frac{1}{2\pi i}\int_{|z|=r} e^z z^{-n-s-1}\mathrm{d}z\right)\mathrm{d}s \\ &= \frac{1}{2\pi i}\int_{(c)} \frac{F(s)}{\Gamma(s)} \frac{n!\Gamma(s)}{\Gamma(n+1+s)}\mathrm{d}s \end{aligned}$$

which after $s \longleftrightarrow -s$ is the Noerlund integral.

Inverse Mellin and De-poissonization Combined

Formally, we have

$$
\begin{aligned}
a_n &= \frac{n!}{2\pi i} \int_{|z|=r} \frac{e^z \tilde{f}(z)}{z^{n+1}} \\
&= \frac{n!}{2\pi i} \int_{|z|=r} \frac{e^z}{z^{n+1}} \left(\frac{1}{2\pi i} \int_{(c)} F(s) z^{-s} \mathrm{d}s \right) \mathrm{d}z \\
&= \frac{n!}{2\pi i} \int_{(c)} F(s) \left(\frac{1}{2\pi i} \int_{|z|=r} e^z z^{-n-s-1} \mathrm{d}z \right) \mathrm{d}s \\
&= \frac{1}{2\pi i} \int_{(c)} \frac{F(s)}{\Gamma(s)} \frac{n!\Gamma(s)}{\Gamma(n+1+s)} \mathrm{d}s
\end{aligned}
$$

which after $s \longleftrightarrow -s$ is the Noerlund integral.

Remark: justification needs e.g. polynomial growth of $F(-s)/\Gamma(-s)$!

Variance of Size in Symmetric Tries (iv)

Recall $\tilde{V}_\infty(z) = 2\tilde{V}_\infty(z/2) + \tilde{h}(z)$ where

$$\tilde{h}(z) = \sum_{k\geq 2} \frac{c_k}{k!} \left(\frac{z}{2}\right)^k \tilde{f}_1^{(k)}(z/2)^2.$$

Variance of Size in Symmetric Tries (iv)

Recall $\tilde{V}_\infty(z) = 2\tilde{V}_\infty(z/2) + \tilde{h}(z)$ where

$$\tilde{h}(z) = \sum_{k\geq 2} \frac{c_k}{k!} \left(\frac{z}{2}\right)^k \tilde{f}_1^{(k)}(z/2)^2.$$

For the size:

$$\tilde{h}(z) = \sum_{j,\ell\geq 1} 2^{j+\ell-2} e^{-\frac{z}{2^j}-\frac{z}{2^\ell}} \sum_{k\geq 2} \frac{2^{-(j+\ell-1)k} c_k}{k!} z^k \left(\frac{z}{2^j} - k + 1\right)\left(\frac{z}{2^\ell} - k + 1\right)$$

Variance of Size in Symmetric Tries (iv)

Recall $\tilde{V}_\infty(z) = 2\tilde{V}_\infty(z/2) + \tilde{h}(z)$ where

$$\tilde{h}(z) = \sum_{k\geq 2} \frac{c_k}{k!}\left(\frac{z}{2}\right)^k \tilde{f}_1^{(k)}(z/2)^2.$$

For the size:

$$\tilde{h}(z) = \sum_{j,\ell\geq 1} 2^{j+\ell-2} e^{-\frac{z}{2^j}-\frac{z}{2^\ell}} \sum_{k\geq 2} \frac{2^{-(j+\ell-1)k}c_k}{k!} z^k \left(\frac{z}{2^j} - k + 1\right)\left(\frac{z}{2^\ell} - k + 1\right)$$

From this one obtains an explicit expression for the Mellin transform!

Variance of Size in Symmetric Tries (iv)

Recall $\tilde{V}_\infty(z) = 2\tilde{V}_\infty(z/2) + \tilde{h}(z)$ where

$$\tilde{h}(z) = \sum_{k\geq 2} \frac{c_k}{k!}\left(\frac{z}{2}\right)^k \tilde{f}_1^{(k)}(z/2)^2.$$

For the size:

$$\tilde{h}(z) = \sum_{j,\ell\geq 1} 2^{j+\ell-2} e^{-\frac{z}{2^j}-\frac{z}{2^\ell}} \sum_{k\geq 2} \frac{2^{-(j+\ell-1)k} c_k}{k!} z^k \left(\frac{z}{2^j} - k + 1\right)\left(\frac{z}{2^\ell} - k + 1\right)$$

From this one obtains an explicit expression for the Mellin transform!

Lemma

$\mathscr{M}[\tilde{h}(z); -s]/\Gamma(-s)$ is of polynomial growth on $\Re(s) > 0$.

Variance of Size in Symmetric Tries (v)

Start with:

$$-\frac{1}{2\pi i}\int_{(3/2)}\frac{\mathscr{M}[\tilde{h}(z);-s]}{(1-2^{1-s})\Gamma(-s)}\frac{n!\Gamma(-s)}{\Gamma(n+1-s)}$$

Variance of Size in Symmetric Tries (v)

Start with:

$$-\frac{1}{2\pi i}\int_{(3/2)}\frac{\mathscr{M}[\tilde{h}(z);-s]}{(1-2^{1-s})\Gamma(-s)}\frac{n!\Gamma(-s)}{\Gamma(n+1-s)}$$

Moving the line of integration to:

Variance of Size in Symmetric Tries (v)

Start with:

$$-\frac{1}{2\pi i}\int_{(3/2)}\frac{\mathscr{M}[\tilde{h}(z);-s]}{(1-2^{1-s})\Gamma(-s)}\frac{n!\Gamma(-s)}{\Gamma(n+1-s)}$$

Moving the line of integration to:

- $+\infty$: gives $\mathrm{Var}(S_n)$;

Variance of Size in Symmetric Tries (v)

Start with:

$$-\frac{1}{2\pi i}\int_{(3/2)}\frac{\mathscr{M}[\tilde{h}(z);-s]}{(1-2^{1-s})\Gamma(-s)}\frac{n!\Gamma(-s)}{\Gamma(n+1-s)}$$

Moving the line of integration to:

- $+\infty$: gives $\mathrm{Var}(S_n)$;
- inside $(0,1)$: gives main term in the asymptotics of $\mathrm{Var}(S_n)$.

Variance of Size in Symmetric Tries (v)

Start with:

$$-\frac{1}{2\pi i}\int_{(3/2)}\frac{\mathscr{M}[\tilde{h}(z);-s]}{(1-2^{1-s})\Gamma(-s)}\frac{n!\Gamma(-s)}{\Gamma(n+1-s)}$$

Moving the line of integration to:

- $+\infty$: gives $\mathrm{Var}(S_n)$;
- inside $(0,1)$: gives main term in the asymptotics of $\mathrm{Var}(S_n)$.

$\longrightarrow$ Rice-Noerlund method for variance without cancellations!

Variance of Size in Symmetric Tries (v)

Start with:

$$-\frac{1}{2\pi i}\int_{(3/2)}\frac{\mathscr{M}[\tilde{h}(z);-s]}{(1-2^{1-s})\Gamma(-s)}\frac{n!\Gamma(-s)}{\Gamma(n+1-s)}$$

Moving the line of integration to:

- $+\infty$: gives $\mathrm{Var}(S_n)$;
- inside $(0,1)$: gives main term in the asymptotics of $\mathrm{Var}(S_n)$.

$\longrightarrow$ Rice-Noerlund method for variance without cancellations!

Longer expansions by moving the line beyond the imaginary axis but for this we need polynomial growth beyond $\Re(s) > 0$

Variance of Size in Symmetric Tries (v)

Start with:

$$-\frac{1}{2\pi i}\int_{(3/2)}\frac{\mathscr{M}[\tilde{h}(z);-s]}{(1-2^{1-s})\Gamma(-s)}\frac{n!\Gamma(-s)}{\Gamma(n+1-s)}$$

Moving the line of integration to:

- $+\infty$: gives $\mathrm{Var}(S_n)$;
- inside $(0,1)$: gives main term in the asymptotics of $\mathrm{Var}(S_n)$.

$\longrightarrow$ Rice-Noerlund method for variance without cancellations!

Longer expansions by moving the line beyond the imaginary axis but for this we need polynomial growth beyond $\Re(s) > 0$

$\longrightarrow$ tameness of $\mathscr{M}[\tilde{h}(z);-s]/\Gamma(-s)$? (**Vallée**; 2017)

Open Problems

- How to show e.g.

$$\mathrm{Var}(S_n) \sim \frac{1}{\log 2} P_1(n) - \frac{1}{\log^2 2} \sum_{k \geq 2} \frac{1}{k!} P_k(n)$$

via the Rice-Noerlund method?

Open Problems

- How to show e.g.

$$\mathrm{Var}(S_n) \sim \frac{1}{\log 2} P_1(n) - \frac{1}{\log^2 2} \sum_{k\ge 2} \frac{1}{k!} P_k(n)$$

  via the Rice-Noerlund method?

- Does the above hold as identity?

Open Problems

- How to show e.g.

$$\mathrm{Var}(S_n) \sim \frac{1}{\log 2} P_1(n) - \frac{1}{\log^2 2} \sum_{k \geq 2} \frac{1}{k!} P_k(n)$$

  via the Rice-Noerlund method?

- Does the above hold as identity?

  For this one would need polynomial growth of $\mathscr{M}[\tilde{h}(z); -s]/\Gamma(-s)$ on $\mathbb{C}$ (when staying uniformly away from the singularities).

Open Problems

- How to show e.g.

$$\mathrm{Var}(S_n) \sim \frac{1}{\log 2} P_1(n) - \frac{1}{\log^2 2} \sum_{k\geq 2} \frac{1}{k!} P_k(n)$$

  via the Rice-Noerlund method?

- Does the above hold as identity?

  For this one would need polynomial growth of $\mathscr{M}[\tilde{h}(z); -s]/\Gamma(-s)$ on $\mathbb{C}$ (when staying uniformly away from the singularities).

- Conditions on T_n

$$X_n \stackrel{d}{=} X_{I_n} + X^*_{n-I_n} + T_n$$

  such that $\mathrm{Var}(X_n)$ can be obtained with Rice-Noerlund method?

Open Problems

- How to show e.g.

$$\mathrm{Var}(S_n) \sim \frac{1}{\log 2} P_1(n) - \frac{1}{\log^2 2} \sum_{k \geq 2} \frac{1}{k!} P_k(n)$$

  via the Rice-Noerlund method?

- Does the above hold as identity?

  For this one would need polynomial growth of $\mathscr{M}[\tilde{h}(z); -s]/\Gamma(-s)$ on $\mathbb{C}$ (when staying uniformly away from the singularities).

- Conditions on T_n

$$X_n \stackrel{d}{=} X_{I_n} + X^*_{n-I_n} + T_n$$

  such that $\mathrm{Var}(X_n)$ can be obtained with Rice-Noerlund method?

  B. Vallée has a general framework for mean but how about variance?