The Node Profile of Symmetric Digital Search Trees

(joint with M. Drmota, H.-K. Hwang and R. Neininger)

Michael Fuchs

Department of Applied Mathematics
National Chiao Tung University

June 8th, 2015

Node Profile of (Rooted) Trees

$B_{n,k}$ = number of external nodes at level k;

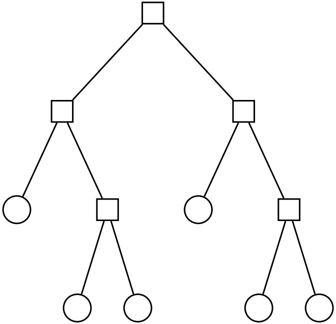
$I_{n,k}$ = number of internal nodes at level k.

Node Profile of (Rooted) Trees

$B_{n,k}$ = number of external nodes at level k;

$I_{n,k}$ = number of internal nodes at level k.

Example:

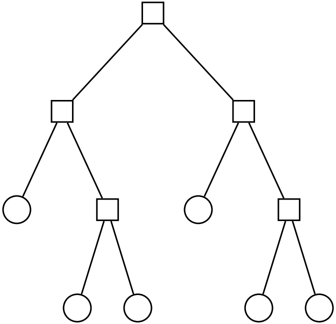


Node Profile of (Rooted) Trees

$B_{n,k}$ = number of external nodes at level k;

$I_{n,k}$ = number of internal nodes at level k.

Example:



$B_{5,0} = 0,$

$B_{5,1} = 0,$

$B_{5,2} = 2,$

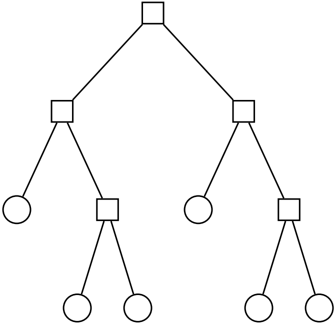
$B_{5,3} = 4,$

Node Profile of (Rooted) Trees

$B_{n,k}$ = number of external nodes at level k;

$I_{n,k}$ = number of internal nodes at level k.

Example:



$B_{5,0} = 0,$ $I_{5,0} = 1;$

$B_{5,1} = 0,$ $I_{5,1} = 2;$

$B_{5,2} = 2,$ $I_{5,2} = 2;$

$B_{5,3} = 4,$ $I_{5,3} = 0.$

Relations to Other Shape Parameters

Many shape parameters can by analyzed through the profile.

Relations to Other Shape Parameters

Many shape parameters can by analyzed through the profile.

- Depth: $P(D_n = k) = B_{n,k}/(n+1)$;
- Width: $\max\{B_{n,k} \ : \ k \geq 0\}$;
- Total Path Length: $\sum_k kB_{n,k}$;
- Height: $\max\{k \ : \ B_{n,k} > 0\}$;
- Shortest Path: $\min\{k \ : \ B_{n,k} > 0\}$;
- Fill-up Level: $\max\{k \ : \ I_{n,k} = 2^k\}$;
- Etc.

Profile of Random Trees

- $\sqrt{n}$-Trees:

  Aldous (1991); Drmota and Gittenberger (1997); Kersting (1998); Pitman (1999); etc.

Profile of Random Trees

- $\sqrt{n}$-Trees:

  Aldous (1991); Drmota and Gittenberger (1997); Kersting (1998); Pitman (1999); etc.

- $\log n$-Trees:
  - Binary Search Trees: Chauvin, Drmota, Jabbour-Hattab (2001); Drmota and Hwang (2005); F., Hwang, Neininger (2006).
  - Recursive Trees: Drmota and Hwang (2005); F., Hwang, Neininger (2006).
  - Plane-oriented Recursive Trees: Hwang (2007).
  - m-ary Seach Trees: Drmota, Janson, Neininger (2008).

Tries

René de la Briandais (1959)

Name from data retrieval (suggested by Fredkin).

Tries

René de la Briandais (1959)

Name from data retrieval (suggested by Fredkin).

Example:

011011
010101
101110
010000
101010
001100

Tries

René de la Briandais (1959)

Name from data retrieval (suggested by Fredkin).

Example:

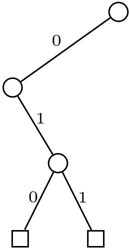
011011
010101
101110
010000
101010
001100

Tries

René de la Briandais (1959)

Name from data retrieval (suggested by Fredkin).

Example:



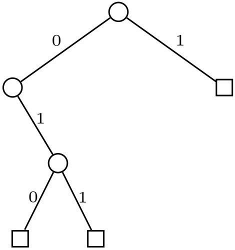
011011
010101
101110
010000
101010
001100

Tries

René de la Briandais (1959)

Name from data retrieval (suggested by Fredkin).

Example:



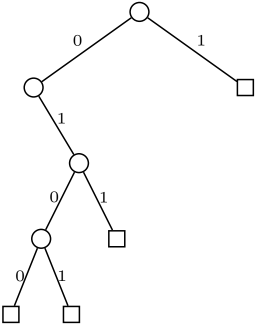
011011
010101
101110
010000
101010
001100

Tries

René de la Briandais (1959)

Name from data retrieval (suggested by Fredkin).

Example:



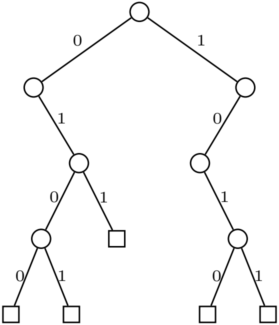
011011
010101
101110
010000
101010
001100

Tries

René de la Briandais (1959)

Name from data retrieval (suggested by Fredkin).

Example:



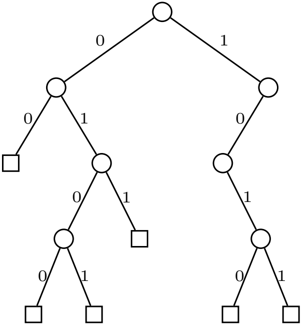
011011
010101
101110
010000
101010
001100

Tries

René de la Briandais (1959)

Name from data retrieval (suggested by Fredkin).

Example:



011011
010101
101110
010000
101010
001100

Digital Search Trees (DSTs)

Edward G. Coffman & James Eve (1970)

Closely related to Lempel-Ziv compression scheme.

Digital Search Trees (DSTs)

Edward G. Coffman & James Eve (1970)

Closely related to Lempel-Ziv compression scheme.

Example:

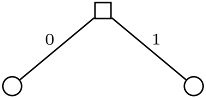
011011
010101
101110
010000
101010
001100

Digital Search Trees (DSTs)

Edward G. Coffman & James Eve (1970)

Closely related to Lempel-Ziv compression scheme.

Example:



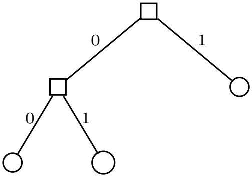
011011
010101
101110
010000
101010
001100

Digital Search Trees (DSTs)

Edward G. Coffman & James Eve (1970)

Closely related to Lempel-Ziv compression scheme.

Example:



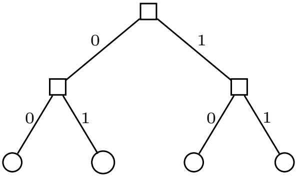
011011
010101
101110
010000
101010
001100

Digital Search Trees (DSTs)

Edward G. Coffman & James Eve (1970)

Closely related to Lempel-Ziv compression scheme.

Example:



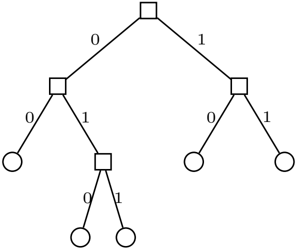
011011
010101
101110
010000
101010
001100

Digital Search Trees (DSTs)

Edward G. Coffman & James Eve (1970)

Closely related to Lempel-Ziv compression scheme.

Example:



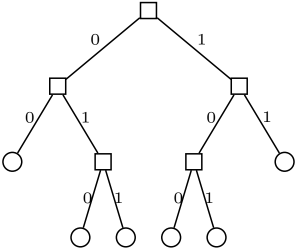
011011
010101
101110
010000
101010
001100

Digital Search Trees (DSTs)

Edward G. Coffman & James Eve (1970)

Closely related to Lempel-Ziv compression scheme.

Example:



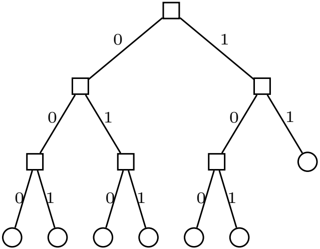
011011
010101
101110
010000
101010
001100

Digital Search Trees (DSTs)

Edward G. Coffman & James Eve (1970)

Closely related to Lempel-Ziv compression scheme.

Example:



011011
010101
101110
010000
101010
001100

Random Model

Bits generated by iid Bernoulli random variables with mean p

$\longrightarrow$ *Bernoulli model*

Random Model

Bits generated by iid Bernoulli random variables with mean p

$\longrightarrow$ *Bernoulli model*

Two types:

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

Random Model

Bits generated by iid Bernoulli random variables with mean p

$\longrightarrow$ *Bernoulli model*

Two types:

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

Question: What can be said about the profile?

Random Model

Bits generated by iid Bernoulli random variables with mean p

$\longrightarrow$ *Bernoulli model*

Two types:

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

Question: What can be said about the profile?

In this talk, we are interested in mean, variance and limit laws of the profile for symmetric DSTs.

Profile of Digital Trees

- Tries:

  Mean, variance, limit laws: Hwang, Nicodéme, Park and Szpankowski (2009).

Profile of Digital Trees

- Tries:

  Mean, variance, limit laws: Hwang, Nicodéme, Park and Szpankowski (2009).

- PATRICIA tries:

  Mean: Magner, Knessl, Szpankowski (2014); Variance & limit laws: Szpankowkski & Magner ($\rightarrow$ Thursday).

Profile of Digital Trees

- Tries:

  Mean, variance, limit laws: Hwang, Nicodéme, Park and Szpankowski (2009).

- PATRICIA tries:

  Mean: Magner, Knessl, Szpankowski (2014); Variance & limit laws: Szpankowkski & Magner ($\rightarrow$ Thursday).

- Asymmetric DSTs:

  Mean: Drmota and Szpankowski (2011); Variance: Kazemi and Vahidi-Asl (2011); so far no limit laws.

Profile of Digital Trees

- Tries:

  Mean, variance, limit laws: Hwang, Nicodéme, Park and Szpankowski (2009).

- PATRICIA tries:

  Mean: Magner, Knessl, Szpankowski (2014); Variance & limit laws: Szpankowkski & Magner ($\rightarrow$ Thursday).

- Asymmetric DSTs:

  Mean: Drmota and Szpankowski (2011); Variance: Kazemi and Vahidi-Asl (2011); so far no limit laws.

- Symmetric DSTs:

  Variance & limit laws: Drmota, F., Hwang, Neininger ($\rightarrow$ this talk).

Profile of Tries

Hwang, Nicodéme, Park, Szpankowski (2009)

Profile of Tries

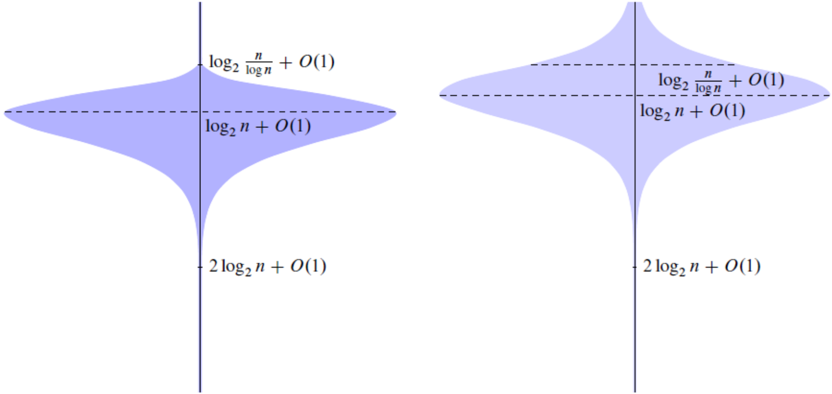
G. Park[1], H.-K Hwang[2], P. Nicodème[3], and W. Szpankowski[4]

[1] Department of Computer Science, State University of New York at Geneseo, Geneseo, 14554, USA
park@geneseo.edu
[2] Institute of Statistical Science, Academia Sinica, 11529 Taipei, Taiwan
hkhwang@stat.sinica.edu.tw
[3] Laboratory LIX, École polytechnique, 91128 Palaiseau Cedex, France
nicodeme@lix.polytechnique.fr
[4] Department of Computer Sciences, Purdue University, 250 N. University Street, West Lafayette, Indiana, 47907-2066, USA
spa@cs.purdue.edu

Abstract. The *profile* of a trie, the most popular data structures on words, is a parameter that represents the number of nodes (either internal or external) with the same distance to the root. Several, if not

Plot of Mean Profile of Symmetric Tries

Hwang, Nicodéme, Park, Szpankowski (2009):



Symmetric Tries: Mean

We have,

$$\mu_{n,k} := \mathbb{E}(B_{n,k}) \sim \begin{cases} n(1-2^{-k})^{n-1}, & \text{if } 2^{-k}n \to \infty; \\ \tilde{M}_{k,1}(n), & \text{if } 4^{-k}n \to 0, \end{cases}$$

where

$$\tilde{M}_{k,1}(z) = z(e^{-z/2^k} - e^{-z/2^{k-1}}).$$

Symmetric Tries: Mean

We have,

$$\mu_{n,k} := \mathbb{E}(B_{n,k}) \sim \begin{cases} n(1-2^{-k})^{n-1}, & \text{if } 2^{-k}n \to \infty; \\ \tilde{M}_{k,1}(n), & \text{if } 4^{-k}n \to 0, \end{cases}$$

where

$$\tilde{M}_{k,1}(z) = z(e^{-z/2^k} - e^{-z/2^{k-1}}).$$

In particular,

$$\tilde{M}_{k,1}(n) \sim \begin{cases} ne^{-n/2^k}, & \text{if } 2^{-k}n \to \infty; \\ \Theta(n), & \text{if } 2^{-k}n = \Theta(1); \\ 2^{-k}n^2, & \text{if } 2^{-k}n \to 0. \end{cases}$$

Symmetric Tries: Mean

We have,

$$\mu_{n,k} := \mathbb{E}(B_{n,k}) \sim \begin{cases} n(1-2^{-k})^{n-1}, & \text{if } 2^{-k}n \to \infty; \\ \tilde{M}_{k,1}(n), & \text{if } 4^{-k}n \to 0, \end{cases}$$

where

$$\tilde{M}_{k,1}(z) = z(e^{-z/2^k} - e^{-z/2^{k-1}}).$$

In particular,

$$\tilde{M}_{k,1}(n) \sim \begin{cases} ne^{-n/2^k}, & \text{if } 2^{-k}n \to \infty; \\ \Theta(n), & \text{if } 2^{-k}n = \Theta(1); \\ 2^{-k}n^2, & \text{if } 2^{-k}n \to 0. \end{cases}$$

Thus, the profile has maximum of order n (asymmetric tries: $n/\sqrt{\log n}$)

Symmetric Tries: Variance

We have,

$$\sigma_{n,k}^2 := \mathrm{Var}(B_{n,k}) \sim \begin{cases} n(1-2^{-k})^{n-1}, & \text{if } 2^{-k}n \to \infty; \\ \tilde{V}_k(n), & \text{if } 4^{-k}n \to 0, \end{cases}$$

where

$$\tilde{V}_k(z) = z(e^{-z/2^k} - e^{-z/2^{k-1}}) + 2^{-k}z^2 e^{-z/2^{k-1}} - 2^{1-k}z^2(e^{-z/2^k} - e^{-z/2^{k-1}})^2.$$

Symmetric Tries: Variance

We have,

$$\sigma_{n,k}^2 := \mathrm{Var}(B_{n,k}) \sim \begin{cases} n(1-2^{-k})^{n-1}, & \text{if } 2^{-k}n \to \infty; \\ \tilde{V}_k(n), & \text{if } 4^{-k}n \to 0, \end{cases}$$

where

$$\tilde{V}_k(z) = z(e^{-z/2^k} - e^{-z/2^{k-1}}) + 2^{-k}z^2e^{-z/2^{k-1}} - 2^{1-k}z^2(e^{-z/2^k} - e^{-z/2^{k-1}})^2.$$

In particular,

$$\tilde{V}_k(n) \sim \begin{cases} ne^{-n/2^k} \sim \tilde{M}_k(n), & \text{if } 2^{-k}n \to \infty; \\ \Theta(n), & \text{if } 2^{-k}n = \Theta(1); \\ 2^{1-k}n^2 \sim 2\tilde{M}_k(n), & \text{if } 2^{-k}n \to 0. \end{cases}$$

Poissonization and Depoissonization

Poisson Model: Build digital tree from Poisson-distributed number of records.

Poissonization and Depoissonization

Poisson Model: Build digital tree from Poisson-distributed number of records.

Poisson moments:

$$\tilde{M}_{k,\ell}(z) = \mathbb{E}(B^{\ell}_{\mathrm{Pois}(z),k}) = e^{-z}\sum_{n\geq 0}\mathbb{E}(B^{\ell}_{n,k})\frac{z^n}{n!}.$$

Poissonization and Depoissonization

Poisson Model: Build digital tree from Poisson-distributed number of records.

Poisson moments:

$$\tilde{M}_{k,\ell}(z) = \mathbb{E}(B^{\ell}_{\mathrm{Pois}(z),k}) = e^{-z}\sum_{n\geq 0}\mathbb{E}(B^{\ell}_{n,k})\frac{z^n}{n!}.$$

Poisson Heuristic:

$$\tilde{M}_{k,\ell}(z) \text{ sufficiently smooth} \quad \Longrightarrow \quad \mathbb{E}(B^{\ell}_{n,k}) \approx \tilde{M}_{k,\ell}(n).$$

Poissonization and Depoissonization

Poisson Model: Build digital tree from Poisson-distributed number of records.

Poisson moments:

$$\tilde{M}_{k,\ell}(z) = \mathbb{E}(B^{\ell}_{\mathrm{Pois}(z),k}) = e^{-z}\sum_{n\geq 0}\mathbb{E}(B^{\ell}_{n,k})\frac{z^n}{n!}.$$

Poisson Heuristic:

$\tilde{M}_{k,\ell}(z)$ sufficiently smooth $\implies$ $\mathbb{E}(B^{\ell}_{n,k}) \approx \tilde{M}_{k,\ell}(n)$.

Poisson heuristic made precise by the **Theory of Analytic Depoissonization** (Jacquet & Szpankowski; 1998).

Poisson Variance

Correct choice is crucial!

Poisson Variance

Correct choice is crucial!

- Asymmetric Digital Trees:

$$\tilde{V}_k(z) = \tilde{M}_{k,2}(z) - \tilde{M}_{k,1}(z)^2.$$

Poisson Variance

Correct choice is crucial!

- Asymmetric Digital Trees:

$$\tilde{V}_k(z) = \tilde{M}_{k,2}(z) - \tilde{M}_{k,1}(z)^2.$$

- Symmetric Digital Trees:

$$\tilde{V}_k(z) = \tilde{M}_{k,2}(z) - \tilde{M}_{k,1}(z)^2 - z\tilde{M}'_{k,1}(z)^2.$$

Poisson Variance

Correct choice is crucial!

- Asymmetric Digital Trees:

$$\tilde{V}_k(z) = \tilde{M}_{k,2}(z) - \tilde{M}_{k,1}(z)^2.$$

- Symmetric Digital Trees:

$$\tilde{V}_k(z) = \tilde{M}_{k,2}(z) - \tilde{M}_{k,1}(z)^2 - z\tilde{M}'_{k,1}(z)^2.$$

With this choice:

$$\mathrm{Var}(B_{n,k}) \sim \tilde{V}_k(n)$$

when $4^{-k}n \to 0$.

Symmetric DSTs: Mean

Let

$$Q(z) = \prod_{\ell=1}^{\infty} \left(1 - z2^{-\ell}\right), \qquad Q_n = \prod_{\ell=1}^{n} \left(1 - 2^{-\ell}\right) = \frac{Q(2^{-n})}{Q(1)}.$$

Symmetric DSTs: Mean

Let

$$Q(z) = \prod_{\ell=1}^{\infty}\left(1 - z2^{-\ell}\right), \qquad Q_n = \prod_{\ell=1}^{n}\left(1 - 2^{-\ell}\right) = \frac{Q(2^{-n})}{Q(1)}.$$

Theorem

We have,

$$\mu_{n,k} \begin{cases} \sim \dfrac{2^k}{Q_k}\left(1 - 2^{-k}\right)^n, & \text{if } 2^{-k}n \to \infty; \\ = 2^k F(n/2^k) + \mathcal{O}(1), & \text{if } 4^{-k}n \to 0, \end{cases}$$

where $F(x)$ is the positive function

$$F(x) = \sum_{j\geq 0} \frac{(-1)^j 2^{-\binom{j}{2}}}{Q_j Q(1)} e^{-2^j x}.$$

$F(x)$ (i)

As $x \to \infty$,

$$F(x) = \frac{e^{-x}}{Q(1)} + \mathcal{O}(e^{-2x})$$

$F(x)$ (i)

As $x \to \infty$,

$$F(x) = \frac{e^{-x}}{Q(1)} + \mathcal{O}(e^{-2x})$$

and as $x \to 0$,

$$F(x) \sim \frac{X^{1/\log 2}}{\sqrt{2\pi x}} \exp\left(-\frac{(\log X \log X)^2}{\log 2} - \sum_{j\in\mathbb{Z}} c_j (X \log X)^{-\chi_j}\right),$$

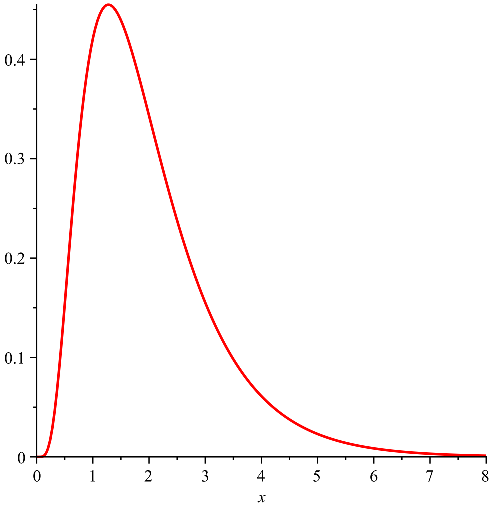
where $X = 1/(x \log 2)$, $\chi_j = 2j\pi i/\log 2$,

$$c_0 = \frac{\log 2}{12} + \frac{\pi^2}{6 \log 2}$$

and

$$c_j = \frac{1}{2j \sinh(2j\pi/\log 2)}, \qquad (j \neq 0).$$

$F(x)$ (ii)



Some Details of the Proof (i)

We have,

$$\tilde{M}_{k,1}(z) + \tilde{M}'_{k,1}(z) = 2\tilde{M}_{k-1,1}(z/2).$$

Some Details of the Proof (i)

We have,

$$\tilde{M}_{k,1}(z) + \tilde{M}'_{k,1}(z) = 2\tilde{M}_{k-1,1}(z/2).$$

By Laplace transform and its inverse,

$$\tilde{M}_{k,1}(z) = 2^k \sum_{0\le j\le k} \frac{(-1)^j 2^{-\binom{j}{2}}}{Q_j Q_{k-j}} e^{-z/2^{k-j}}.$$

Some Details of the Proof (i)

We have,

$$\tilde{M}_{k,1}(z) + \tilde{M}'_{k,1}(z) = 2\tilde{M}_{k-1,1}(z/2).$$

By Laplace transform and its inverse,

$$\tilde{M}_{k,1}(z) = 2^k \sum_{0\le j\le k} \frac{(-1)^j 2^{-\binom{j}{2}}}{Q_j Q_{k-j}} e^{-z/2^{k-j}}.$$

From this,

$$\mu_{n,k} = 2^k \sum_{0\le j\le k} \frac{(-1)^j 2^{-\binom{j}{2}}}{Q_j Q_{k-j}} \left(1 - 2^{j-k}\right)^n.$$

This formula was first derived by Louchard (1987).

Some Details of the Proof (i)

We have,

$$\tilde{M}_{k,1}(z) + \tilde{M}'_{k,1}(z) = 2\tilde{M}_{k-1,1}(z/2).$$

By Laplace transform and its inverse,

$$\tilde{M}_{k,1}(z) = 2^k \sum_{0\le j\le k} \frac{(-1)^j 2^{-\binom{j}{2}}}{Q_j Q_{k-j}} e^{-z/2^{k-j}}.$$

From this,

$$\mu_{n,k} = 2^k \sum_{0\le j\le k} \frac{(-1)^j 2^{-\binom{j}{2}}}{Q_j Q_{k-j}} \left(1 - 2^{j-k}\right)^n.$$

This formula was first derived by Louchard (1987).

This is useful if $n2^{-k} \to \infty$.

Some Details of the Proof (ii)

If $4^{-k}n \to 0$, Poisson heuristic holds.

Some Details of the Proof (ii)

If $4^{-k}n \to 0$, Poisson heuristic holds.

Lemma

We have,

$$\tilde{M}_{k,1}(z) = 2^k \sum_{r\geq 0} \frac{2^{-\binom{r+1}{2}-kr}}{Q_r} F^{(r)}\left(\frac{z}{2^k}\right).$$

Some Details of the Proof (ii)

If $4^{-k}n \to 0$, Poisson heuristic holds.

Lemma

We have,

$$\tilde{M}_{k,1}(z) = 2^k \sum_{r\geq 0} \frac{2^{-\binom{r+1}{2}-kr}}{Q_r} F^{(r)}\left(\frac{z}{2^k}\right).$$

This gives,

$$\tilde{M}_{k,1}(z) = 2^k F\left(\frac{z}{2^k}\right) + \mathcal{O}(1).$$

Result follows from depoissonization.

Symmetric DSTs: Variance

Theorem (Drmota, F., Hwang, Neininger)

We have,

$$\sigma_{n,k}^2 \begin{cases} \sim \dfrac{2^k}{Q_k}\left(1-2^{-k}\right)^n, & \text{if } 2^{-k}n \to \infty; \\ = 2^k H(n/2^k) + \mathcal{O}(1), & \text{if } 4^{-k}n \to 0, \end{cases}$$

where $H(x)$ is a function with

$$H(x) = \frac{e^{-x}}{Q(1)} + \mathcal{O}(xe^{-2x}), \qquad (x \to \infty)$$

and

$$H(x) \sim 2F(x), \qquad (x \to 0).$$

$H(x)$ (i)

We have,

$$H(x) = \sum_{j,r=0}^{\infty} \sum_{0\le h,\ell\le j} \frac{2^{-j}(-1)^{r+h+\ell}2^{-\binom{r}{2}-\binom{h}{2}-\binom{\ell}{2}+2h+2\ell}}{Q_r Q(1) Q_h Q_{j-h} Q_\ell Q_{j-\ell}} \varphi(2^{r+j}, 2^h + 2^\ell; x),$$

where

$$\varphi(u,v;x) = \begin{cases} \dfrac{e^{-ux} - ((v-u)x+1)e^{-vx}}{(v-u)^2}, & \text{if } u \neq v; \\ x^2 e^{-ux}/2, & \text{if } u = v. \end{cases}$$

$H(x)$ (i)

We have,

$$H(x) = \sum_{j,r=0}^{\infty} \sum_{0\le h,\ell\le j} \frac{2^{-j}(-1)^{r+h+\ell}2^{-\binom{r}{2}-\binom{h}{2}-\binom{\ell}{2}+2h+2\ell}}{Q_r Q(1) Q_h Q_{j-h} Q_\ell Q_{j-\ell}} \varphi(2^{r+j}, 2^h+2^\ell; x),$$

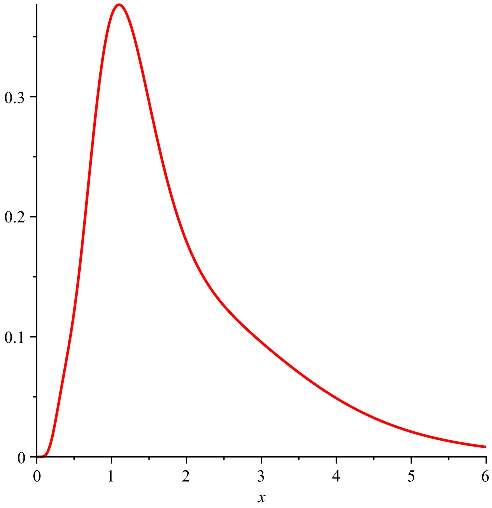
where

$$\varphi(u,v;x) = \begin{cases} \dfrac{e^{-ux} - ((v-u)x+1)e^{-vx}}{(v-u)^2}, & \text{if } u \neq v; \\ x^2 e^{-ux}/2, & \text{if } u = v. \end{cases}$$

Proposition (Drmota, F., Hwang, Neininger)

$H(x)$ is a positive function on $(0,\infty)$.

$H(x)$ (ii)



Some Details of the Proof (i)

We have,

$$\tilde{V}_k(z) + \tilde{V}'_k(z) = 2\tilde{V}_{k-1}(z/2) + z\tilde{M}''_{k,2}(z)^2.$$

Some Details of the Proof (i)

We have,

$$\tilde{V}_k(z) + \tilde{V}_k'(z) = 2\tilde{V}_{k-1}(z/2) + z\tilde{M}_{k,2}''(z)^2.$$

By Laplace transform and its inverse,

$$\tilde{V}_k(z) = \sum_{(j,r,h,\ell)\in\mathscr{V}} \frac{2^{k-j}(-1)^{r+h+\ell}2^{-\binom{r}{2}-\binom{h}{2}-\binom{\ell}{2}+2h+2\ell}}{Q_rQ_{k-j-r}Q_hQ_{j-h}Q_\ell Q_{j-\ell}}\varphi\left(2^{r+j}, 2^h+2^\ell, \frac{z}{2^k}\right)$$

with

$$\mathscr{V} = \{(j,r,h,\ell)\ :\ 0 \le j \le k, 0 \le r \le k-j, 0 \le h, \ell \le j\}$$

and

$$\varphi(u,v;x) = \begin{cases} \dfrac{e^{-ux} - ((v-u)x+1)e^{-vx}}{(v-u)^2}, & \text{if } u \ne v; \\ x^2e^{-ux}/2, & \text{if } u = v. \end{cases}$$

Some Details of the Proof (ii)

Lemma

We have,

$$\tilde{V}_k(z) = 2^k \sum_{m\geq 0} \frac{2^{-\binom{m+1}{2}-km}}{Q_m} H^{(m)}\left(\frac{z}{2^k}\right).$$

Some Details of the Proof (ii)

Lemma

We have,

$$\tilde{V}_k(z) = 2^k \sum_{m\ge 0} \frac{2^{-\binom{m+1}{2}-km}}{Q_m} H^{(m)}\left(\frac{z}{2^k}\right).$$

The Laplace transform of $H(z)$:

$$\mathscr{L}[H(z); s] = \sum_{j\ge 0} 4^{-j} \frac{\tilde{g}_j^*(2^{-j}s)}{Q(-2^{1-j}s)}$$

where

$$\tilde{g}_j^*(s) = \sum_{0\le k,\ell\le j} \frac{(-1)^{h+\ell} 2^{-\binom{h}{2}-\binom{\ell}{2}+2h+2\ell}}{Q_k Q_{j-k} Q_\ell Q_{j-\ell}} \frac{1}{(2^j s + 2^h + 2^\ell)^2}.$$

Some Details of the Proof (iii)

Lemma

We have, as $s \to \infty$,

$$\frac{\tilde{g}_0^*(s)}{Q(-2s)} \sim \frac{1}{s^2 Q(-2s)}, \qquad 4^{-1}\frac{\tilde{g}_1^*(2^{-1}s)}{Q(-s)} \sim \frac{9}{sQ(-2s)}$$

and, for $j \geq 2$,

$$4^{-j}\frac{\tilde{g}_j^*(2^{-j}s)}{Q(-2^{1-j}s)} \sim \frac{(2j-3)!}{((j-2)!)^2}\frac{2^{\binom{j}{2}}}{s^{j-2}Q(-2s)}.$$

Thus,

$$\mathscr{L}[H(z);s] \sim \frac{2}{Q(-2s)}$$

and hence, $H(x) \sim 2F(x)$ *as* $x \to 0$.

Symmetric DSTs: Limit Laws

Corollary (Drmota, F., Hwang, Neininger)

We have,

$$\mu_{n,k} \longrightarrow \infty \quad \textit{iff} \quad \sigma_{n,k}^2 \longrightarrow \infty.$$

Theorem (Drmota, F., Hwang, Neininger)

Assume that $\mu_{n,k} \longrightarrow \infty$. Then,

$$\frac{B_{n,k} - \mu_{n,k}}{\sigma_{n,k}} \stackrel{d}{\longrightarrow} N(0,1),$$

where $N(0,1)$ denotes a standard normal distribution.

Application to the Height

H_n =height of a symmetric DST of size n.

Application to the Height

H_n =height of a symmetric DST of size n.

Theorem (Drmota, F., Hwang, Neininger)

Set

$$k_n = \min\{k \geq \log_2 n \ : \ 2^k F(n/2^k) \leq 1\}.$$

Then,

$$k_n = \log_2 n + \sqrt{2 \log_2 n} - \log_2 \left(\sqrt{\log_2 n}\right) + \mathcal{O}(1).$$

Moreover,

$$P(H_n = k_n - 2 \text{ or } H_n = k_n - 1) \to 1.$$

Application to the Height

H_n =height of a symmetric DST of size n.

Theorem (Drmota, F., Hwang, Neininger)

Set

$$k_n = \min\{k \geq \log_2 n \ : \ 2^k F(n/2^k) \leq 1\}.$$

Then,

$$k_n = \log_2 n + \sqrt{2 \log_2 n} - \log_2 \left(\sqrt{\log_2 n}\right) + \mathcal{O}(1).$$

Moreover,

$$P(H_n = k_n - 2 \text{ or } H_n = k_n - 1) \to 1.$$

This solves an open problem of Aldous & Shields.

Summary of Results for Symmetric DSTs

- Mean profile tends to infinity when k is roughly in the range

$$\log_2 n - \log_2 \log n \leq k \leq \log_2 n + \sqrt{2 \log_2 n};$$

otherwise it is bounded.

Summary of Results for Symmetric DSTs

- Mean profile tends to infinity when k is roughly in the range

$$\log_2 n - \log_2 \log n \leq k \leq \log_2 n + \sqrt{2 \log_2 n};$$

otherwise it is bounded.

- Maximum of mean profile is of linear order.

Summary of Results for Symmetric DSTs

- Mean profile tends to infinity when k is roughly in the range

$$\log_2 n - \log_2 \log n \le k \le \log_2 n + \sqrt{2 \log_2 n};$$

otherwise it is bounded.

- Maximum of mean profile is of linear order.
- Variance has same order as the mean. Thus, it tends to infinity iff mean tends to infinity.

Summary of Results for Symmetric DSTs

- Mean profile tends to infinity when k is roughly in the range

$$\log_2 n - \log_2 \log n \leq k \leq \log_2 n + \sqrt{2 \log_2 n};$$

  otherwise it is bounded.

- Maximum of mean profile is of linear order.

- Variance has same order as the mean. Thus, it tends to infinity iff mean tends to infinity.

- If mean tends to infinity, a central limit theorem holds.

Summary of Results for Symmetric DSTs

- Mean profile tends to infinity when k is roughly in the range

$$\log_2 n - \log_2 \log n \leq k \leq \log_2 n + \sqrt{2 \log_2 n};$$

  otherwise it is bounded.

- Maximum of mean profile is of linear order.

- Variance has same order as the mean. Thus, it tends to infinity iff mean tends to infinity.

- If mean tends to infinity, a central limit theorem holds.

- Our results have many applications, e.g., they allow us to solve a problem of Aldous & Shields.