Approximate Counting via the Poisson-Laplace-Mellin Method

(joint with Chung-Kuei Lee and Helmut Prodinger)

Michael Fuchs

Department of Applied Mathematics
National Chiao Tung University

Hsinchu, Taiwan

Montreal, June 18, 2012

Approximate Counting (Morris 1978)

Space needed for counting n objects: $\Theta(\log n)$.

Approximate Counting (Morris 1978)

Space needed for counting n objects: $\Theta(\log n)$.

Problem: What if space is very limited?

Approximate Counting (Morris 1978)

Space needed for counting n objects: $\Theta(\log n)$.

Problem: What if space is very limited?

Answer: Allow an error tolerance: **approximate counting**.

Counter C_n with $C_0 = 0$ and $(0 < q < 1)$

$$C_{n+1} = \begin{cases} C_n + 1, & \text{with probability } q^{C_n}; \\ C_n, & \text{with probability } 1 - q^{C_n}. \end{cases}$$

Approximate Counting (Morris 1978)

Space needed for counting n objects: $\Theta(\log n)$.

Problem: What if space is very limited?

Answer: Allow an error tolerance: **approximate counting**.

Counter C_n with $C_0 = 0$ and $(0 < q < 1)$

$$C_{n+1} = \begin{cases} C_n + 1, & \text{with probability } q^{C_n}; \\ C_n, & \text{with probability } 1 - q^{C_n}. \end{cases}$$

Easy to show:

$$\mathbb{E}(q^{-C_n}) = n\left(q^{-1} - 1\right) + 1.$$

Now, only $\Theta(\log \log n)$ space is needed.

Applications

Approximate counting has found many applications:

Applications

Approximate counting has found many applications:

- Analysis of the Webgraph.
- Monitoring network traffic.
- Finding patterns in protein and DNA sequencing.
- Computing frequency moments of data streams.
- Data storage in flash memory.
- Etc.

Applications

Approximate counting has found many applications:

- Analysis of the Webgraph.
- Monitoring network traffic.
- Finding patterns in protein and DNA sequencing.
- Computing frequency moments of data streams.
- Data storage in flash memory.
- Etc.

Many refinements have been proposed.

The Markov Chain C_n

Other problems leading to the same Markov chain:

The Markov Chain C_n

Other problems leading to the same Markov chain:

- Width of greedy decomposition of random acyclic digraphs into node-disjoint paths.
- Size of greedy independent set in random graphs.
- Size of greedy clique in random graphs.
- Length of leftmost path in random digital search trees.

The Markov Chain C_n

Other problems leading to the same Markov chain:

- Width of greedy decomposition of random acyclic digraphs into node-disjoint paths.
- Size of greedy independent set in random graphs.
- Size of greedy clique in random graphs.
- Length of leftmost path in random digital search trees.

Variations of this Markov chain were also studied:

Simon (1988); Crippa and Simon (1997); Bertoin, Biane and Yor (2003); Guillemin, Robert and Zwart (2004); Louchard and Prodinger (2008)

Analysis of Approximate Counting

Flajolet (1985):

$$\mathbb{E}(C_n) \sim \log_{1/q} n + C_{\text{mean}} + F(\log_{1/q} n),$$

where $F(z)$ is a 1-periodic function

Analysis of Approximate Counting

Flajolet (1985):

$$\mathbb{E}(C_n) \sim \log_{1/q} n + C_{\text{mean}} + F(\log_{1/q} n),$$

where $F(z)$ is a 1-periodic function and

$$\text{Var}(C_n) \sim C_{\text{var}} + G(\log_{1/q} n),$$

where $G(z)$ is a 1-periodic function and

$$C_{\text{var}} = \frac{\pi^2}{6\log^2(1/q)} - \alpha - \beta + \frac{1}{12} - \frac{1}{\log(1/q)} \sum_{l\geq 1} \frac{1}{l \sinh(2l\pi^2/\log(1/q))}$$

with $\alpha = \sum_{l\geq 1} q^l/(1-q^l)$ and $\beta = \sum_{l\geq 1} q^{2l}/(1-q^l)^2$.

Methods

Many different methods have been used:

Methods

Many different methods have been used:

- **Mellin Transform**: Flajolet (1985); Prodinger (1992)

Methods

Many different methods have been used:

- **Mellin Transform**: Flajolet (1985); Prodinger (1992)
- **Rice Method**: Kirschenhofer and Prodinger (1991)

Methods

Many different methods have been used:

- **Mellin Transform**: Flajolet (1985); Prodinger (1992)
- **Rice Method**: Kirschenhofer and Prodinger (1991)
- **Euler Transform**: Prodinger (1994)

Methods

Many different methods have been used:

- **Mellin Transform**: Flajolet (1985); Prodinger (1992)
- **Rice Method**: Kirschenhofer and Prodinger (1991)
- **Euler Transform**: Prodinger (1994)
- **Analysis of Extreme Value Distributions**: Louchard and Prodinger (2006)

Methods

Many different methods have been used:

- **Mellin Transform**: Flajolet (1985); Prodinger (1992)
- **Rice Method**: Kirschenhofer and Prodinger (1991)
- **Euler Transform**: Prodinger (1994)
- **Analysis of Extreme Value Distributions**: Louchard and Prodinger (2006)
- **Martingale Theory**: Rosenkrantz (1987)

Methods

Many different methods have been used:

- **Mellin Transform**: Flajolet (1985); Prodinger (1992)
- **Rice Method**: Kirschenhofer and Prodinger (1991)
- **Euler Transform**: Prodinger (1994)
- **Analysis of Extreme Value Distributions**: Louchard and Prodinger (2006)
- **Martingale Theory**: Rosenkrantz (1987)
- **Probability Theory**: Robert (2005)

Digital Search Tree (DST)

Introduced by Coffman and Eve (1970).

Digital Search Tree (DST)

Introduced by Coffman and Eve (1970).

Example: a digital search tree build from 9 keys:

010111
101011
100001
011011
111110
110111
010011
011110
000100

Digital Search Tree (DST)

Introduced by Coffman and Eve (1970).

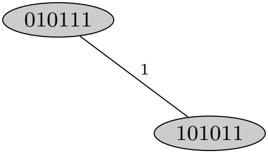
Example: a digital search tree build from 9 keys:

010111
101011
100001
011011
111110
110111
010011
011110
000100

Digital Search Tree (DST)

Introduced by Coffman and Eve (1970).

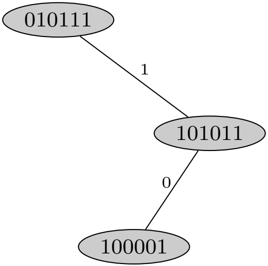
Example: a digital search tree build from 9 keys:



010111
101011
100001
011011
111110
110111
010011
011110
000100

Introduced by Coffman and Eve (1970).

Example: a digital search tree build from 9 keys:

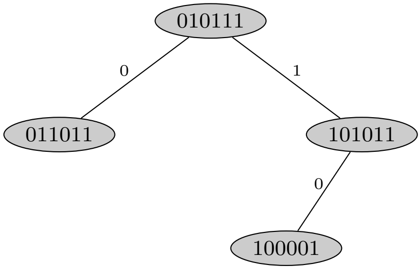


010111
101011
100001
011011
111110
110111
010011
011110
000100

Digital Search Tree (DST)

Introduced by Coffman and Eve (1970).

Example: a digital search tree build from 9 keys:

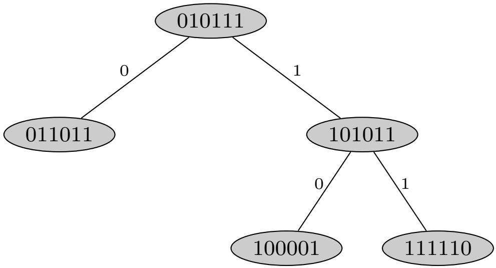


010111
101011
100001
011011
111110
110111
010011
011110
000100

Digital Search Tree (DST)

Introduced by Coffman and Eve (1970).

Example: a digital search tree build from 9 keys:

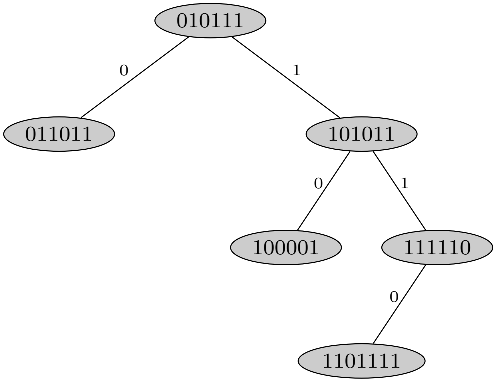


010111
101011
100001
011011
111110
110111
010011
011110
000100

Digital Search Tree (DST)

Introduced by Coffman and Eve (1970).

Example: a digital search tree build from 9 keys:

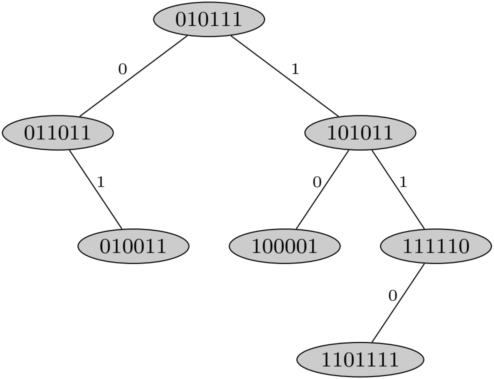


010111
101011
100001
011011
111110
110111
010011
011110
000100

Digital Search Tree (DST)

Introduced by Coffman and Eve (1970).

Example: a digital search tree build from 9 keys:

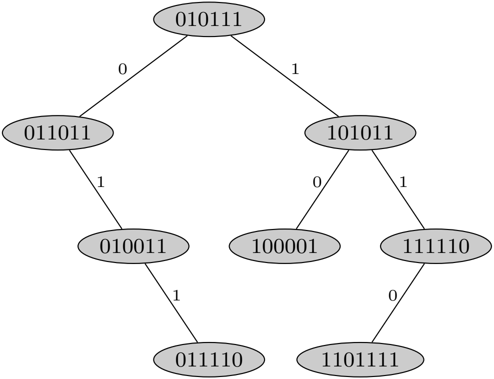


010111
101011
100001
011011
111110
110111
010011
011110
000100

Digital Search Tree (DST)

Introduced by Coffman and Eve (1970).

Example: a digital search tree build from 9 keys:

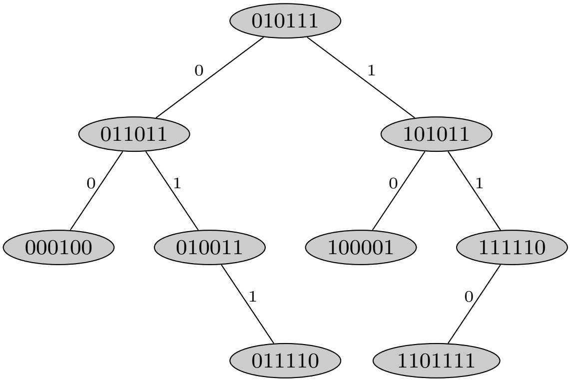


010111
101011
100001
011011
111110
110111
010011
011110
000100

Digital Search Tree (DST)

Introduced by Coffman and Eve (1970).

Example: a digital search tree build from 9 keys:



010111
101011
100001
011011
111110
110111
010011
011110
000100

Random Model and Leftmost Path

Random Model:

Bits are generated by independent Bernoulli random variables with mean p.

Random Model and Leftmost Path

Random Model:

Bits are generated by independent Bernoulli random variables with mean p.

Two types of trees:

- $p = 1/2$: *symmetric digital search tree*;

Random Model and Leftmost Path

Random Model:

Bits are generated by independent Bernoulli random variables with mean p.

Two types of trees:

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

Random Model and Leftmost Path

Random Model:

Bits are generated by independent Bernoulli random variables with mean p.

Two types of trees:

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

Length of the Leftmost Path:

X_n: number of vertices on leftmost path.

Random Model and Leftmost Path

Random Model:

Bits are generated by independent Bernoulli random variables with mean p.

Two types of trees:

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

Length of the Leftmost Path:

X_n: number of vertices on leftmost path.

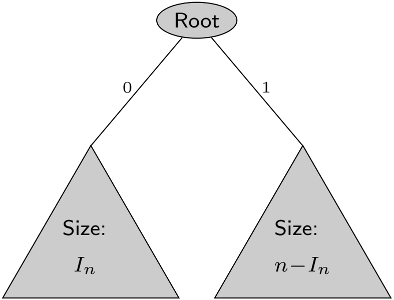
Note that:

$$X_n \stackrel{d}{=} C_n.$$

Distributional Recurrence of X_n

$$X_{n+1} \stackrel{d}{=} X_{I_n} + 1$$

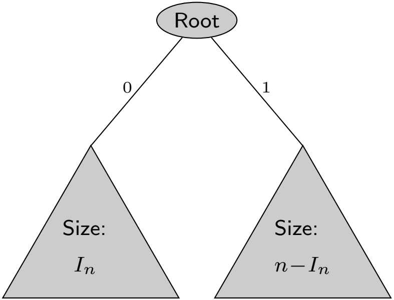
- $I_n \stackrel{d}{=} \text{Binomial}(n, q)$;
- X_n, I_n independent.



Distributional Recurrence of X_n

$$X_{n+1} \stackrel{d}{=} X_{I_n} + 1$$

- $I_n \stackrel{d}{=} \text{Binomial}(n, q)$;
- X_n, I_n independent.



Recurrence of moments:

$$f_{n+1} = \sum_{j=0}^{n} \binom{n}{j} q^j p^{n-j} f_j + g_n.$$

Other Shape Parameters

- **Depth**

  Konheim, Newman, Knuth, Devroye, Louchard, Szpankowski

- **Total Path Length**

  Flajolet, Sedgewick, Prodinger, Kirschenhofer, Szpankowski, Hubalek

- **Peripheral Path Length**

  Drmota, Gittenberger, Panholzer, Prodinger, Ward

- **# of Occurrences of Patterns**

  Knuth, Flajolet, Sedgewick, Prodinger, Kirschenhofer

- **Colless Index**

  Fuchs, Hwang, Zacharovas

Methods

- **Rice Method**

  Introduced by Flajolet and Sedgewick.

Methods

- **Rice Method**

  Introduced by Flajolet and Sedgewick.

- **Approach of Flajolet and Richmond**

  Based on Euler transform, Mellin transform, and singularity analysis.

Methods

- **Rice Method**

  Introduced by Flajolet and Sedgewick.

- **Approach of Flajolet and Richmond**

  Based on Euler transform, Mellin transform, and singularity analysis.

- **Approach via Analytic Depoissonization**

  Introduced by Jacquet & Regnier and Jacquet & Szpankowski. Based on saddle point method and Mellin transform.

Methods

- **Rice Method**

  Introduced by Flajolet and Sedgewick.

- **Approach of Flajolet and Richmond**

  Based on Euler transform, Mellin transform, and singularity analysis.

- **Approach via Analytic Depoissonization**

  Introduced by Jacquet & Regnier and Jacquet & Szpankowski. Based on saddle point method and Mellin transform.

- **Schachinger's Approach**

  Largely elementary.

Poisson-Laplace-Mellin Method

Poisson-Laplace-Mellin Method

- Poissonized mean and variance satisfy differential-function equations.

Poisson-Laplace-Mellin Method

- Poissonized mean and variance satisfy differential-function equations.
- Due to Jacquet-Szpankowski's theory of depoissonization only the Poisson model has to be analyzed.

Poisson-Laplace-Mellin Method

- Poissonized mean and variance satisfy differential-function equations.
- Due to Jacquet-Szpankowski's theory of depoissonization only the Poisson model has to be analyzed.
- Laplace transform to get rid of the differential operator.

Poisson-Laplace-Mellin Method

- Poissonized mean and variance satisfy differential-function equations.
- Due to Jacquet-Szpankowski's theory of depoissonization only the Poisson model has to be analyzed.
- Laplace transform to get rid of the differential operator.
- A normalization factor simplifies the functional equation satisfied by Laplace transform.

Poisson-Laplace-Mellin Method

- Poissonized mean and variance satisfy differential-function equations.
- Due to Jacquet-Szpankowski's theory of depoissonization only the Poisson model has to be analyzed.
- Laplace transform to get rid of the differential operator.
- A normalization factor simplifies the functional equation satisfied by Laplace transform.
- Mellin transform is applied which can be computed explicitly.

Poisson-Laplace-Mellin Method

- Poissonized mean and variance satisfy differential-function equations.
- Due to Jacquet-Szpankowski's theory of depoissonization only the Poisson model has to be analyzed.
- Laplace transform to get rid of the differential operator.
- A normalization factor simplifies the functional equation satisfied by Laplace transform.
- Mellin transform is applied which can be computed explicitly.
- We use inverse Mellin transform and inverse Laplace transform to obtain asymptotic expansions in the Poisson model.

Poissonization

Moments satisfy the recurrence:

$$f_{n+1} = \sum_{j=0}^{n} \binom{n}{j} q^j p^{n-j} f_j + g_n.$$

Poissonization

Moments satisfy the recurrence:

$$f_{n+1} = \sum_{j=0}^{n} \binom{n}{j} q^j p^{n-j} f_j + g_n.$$

Consider Poisson-generating function of f_n and g_n, i.e.,

$$\tilde{f}(z) := e^{-z} \sum_{n\geq 0} f_n \frac{z^n}{n!}, \qquad \tilde{g}(z) := e^{-z} \sum_{n\geq 0} g_n \frac{z^n}{n!}.$$

Poissonization

Moments satisfy the recurrence:

$$f_{n+1} = \sum_{j=0}^{n} \binom{n}{j} q^j p^{n-j} f_j + g_n.$$

Consider Poisson-generating function of f_n and g_n, i.e.,

$$\tilde{f}(z) := e^{-z} \sum_{n\geq 0} f_n \frac{z^n}{n!}, \qquad \tilde{g}(z) := e^{-z} \sum_{n\geq 0} g_n \frac{z^n}{n!}.$$

Then,

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

This is a differential-functional equation.

Poisson Heuristic

Poisson Heuristic:

$$f_n \text{ sufficiently smooth} \quad \Longrightarrow \quad f_n \approx \tilde{f}(n).$$

Poisson Heuristic

Poisson Heuristic:

$$f_n \text{ sufficiently smooth} \quad \Longrightarrow \quad f_n \approx \tilde{f}(n).$$

More precisely: if f_n is smooth enough,

$$f_n \sim \sum_{j\geq 0} \frac{\tilde{f}^{(j)}(n)}{n!}\tau_j(n) = \tilde{f}(n) - \frac{n}{2}\tilde{f}''(n) + \dots,$$

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

This is called *Poisson-Charlier expansion* (can be already found in Ramanujan's notebooks).

Jacquet-Szpankowski-admissibility (JS-admissibility)

$\tilde{f}(z)$ is called JS-admissible if

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

$$\tilde{f}(z) = \mathcal{O}\left(|z|^{\alpha} \log^{\beta} |z|\right),$$

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

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

Jacquet-Szpankowski-admissibility (JS-admissibility)

$\tilde{f}(z)$ is called JS-admissible if

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

$$\tilde{f}(z) = \mathcal{O}\left(|z|^\alpha \log^\beta |z|\right),$$

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

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

Theorem (Jacquet and Szpankowski)

If $\tilde{f}(z)$ is JS-admissible, then

$$f_n \sim \tilde{f}(n) - \frac{n}{2}\tilde{f}''(n) + \cdots .$$

Depoissonization

JS-admissibility satisfies closure properties:

(i) $\tilde{f}, \tilde{g}$ JS-admissible, then $\tilde{f} + \tilde{g}$ JS-admissible.

(ii) $\tilde{f}$ JS-admissible, then $\tilde{f}'$ JS-admissible. Etc.

Depoissonization

JS-admissibility satisfies closure properties:

(i) $\tilde{f}, \tilde{g}$ JS-admissible, then $\tilde{f} + \tilde{g}$ JS-admissible.

(ii) $\tilde{f}$ JS-admissible, then $\tilde{f}'$ JS-admissible. Etc.

Proposition

Consider

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

We have,

$$\tilde{g}(z) \textit{ JS-admissible} \quad \Longleftrightarrow \quad \tilde{f}(z) \textit{ JS-admissible.}$$

Poissonized Mean and Second Moment

Define

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

which are poissonized mean and second moment.

Poissonized Mean and Second Moment

Define

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

which are poissonized mean and second moment.

Then,

$$\tilde{f}_1(z) + \tilde{f}_1'(z) = \tilde{f}_1(qz) + 1$$
$$\tilde{f}_2(z) + \tilde{f}_2'(z) = \tilde{f}_2(qz) + 2\tilde{f}_1(qz) + 1$$

Poissonized Mean and Second Moment

Define

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

which are poissonized mean and second moment.

Then,

$$\begin{aligned}\tilde{f}_1(z) + \tilde{f}_1'(z) &= \tilde{f}_1(qz) + 1\\ \tilde{f}_2(z) + \tilde{f}_2'(z) &= \tilde{f}_2(qz) + 2\tilde{f}_1(qz) + 1\end{aligned}$$

Previous results show that $\tilde{f}_1(z), \tilde{f}_2(z)$ are JS admissible.

Poissonized Variance

Important in order to avoid having to cope with **cancelations**!

Poissonized Variance

Important in order to avoid having to cope with **cancelations**!

- Mean, Variance $= \Theta(\log^{\beta} n)$:

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

Poissonized Variance

Important in order to avoid having to cope with **cancelations**!

- Mean, Variance $= \Theta(\log^\beta n)$:

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

- Mean, Variance $= \Theta(n \log^\beta n)$:

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

Poissonized Variance

Important in order to avoid having to cope with **cancelations**!

- Mean, Variance $= \Theta(\log^{\beta} n)$:

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

- Mean, Variance $= \Theta(n \log^{\beta} n)$:

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

More general:

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

Then one obtains even identities!

Laplace and Mellin Transform (i)

We start from,

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

Laplace and Mellin Transform (i)

We start from,

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

Applying Laplace transform,

$$(s+1)\mathscr{L}[\tilde{f}(z); s] = \frac{1}{q}\mathscr{L}\left[\tilde{f}(z); \frac{1}{q}\right] + \mathscr{L}[\tilde{g}(z); s].$$

Laplace and Mellin Transform (i)

We start from,

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

Applying Laplace transform,

$$(s+1)\mathscr{L}[\tilde{f}(z); s] = \frac{1}{q}\mathscr{L}\left[\tilde{f}(z); \frac{1}{q}\right] + \mathscr{L}[\tilde{g}(z); s].$$

Define,

$$Q(s) := \sum_{l\geq 1}\left(1 - q^l s\right)$$

and $Q_\infty := Q(1)$.

Laplace and Mellin Transform (ii)

Set

$$\bar{f}(s) := \frac{\mathscr{L}[\tilde{f}(z); s]}{Q(-s)}, \qquad \bar{g}(s) := \frac{\mathscr{L}[\tilde{g}(z); s]}{Q(-s/q)}.$$

Laplace and Mellin Transform (ii)

Set

$$\bar{f}(s) := \frac{\mathscr{L}[\tilde{f}(z); s]}{Q(-s)}, \qquad \bar{g}(s) := \frac{\mathscr{L}[\tilde{g}(z); s]}{Q(-s/q)}.$$

Then,

$$\bar{f}(s) = \frac{1}{q}\bar{f}\left(\frac{s}{q}\right) + \bar{g}(s).$$

Laplace and Mellin Transform (ii)

Set

$$\bar{f}(s) := \frac{\mathscr{L}[\tilde{f}(z); s]}{Q(-s)}, \qquad \bar{g}(s) := \frac{\mathscr{L}[\tilde{g}(z); s]}{Q(-s/q)}.$$

Then,

$$\bar{f}(s) = \frac{1}{q}\bar{f}\left(\frac{s}{q}\right) + \bar{g}(s).$$

Applying Mellin transform,

$$\mathscr{M}[\bar{f}(s); \omega] = \frac{\mathscr{M}[\bar{g}(s); \omega]}{1 - q^{\omega - 1}}.$$

Laplace and Mellin Transform (ii)

Set

$$\bar{f}(s) := \frac{\mathscr{L}[\tilde{f}(z); s]}{Q(-s)}, \qquad \bar{g}(s) := \frac{\mathscr{L}[\tilde{g}(z); s]}{Q(-s/q)}.$$

Then,

$$\bar{f}(s) = \frac{1}{q}\bar{f}\left(\frac{s}{q}\right) + \bar{g}(s).$$

Applying Mellin transform,

$$\mathscr{M}[\bar{f}(s); \omega] = \frac{\mathscr{M}[\bar{g}(s); \omega]}{1 - q^{\omega-1}}.$$

From this, an asymptotic expansion of $\tilde{f}(z)$ as $z \to \infty$ is obtained via inverse Mellin transform and inverse Laplace transform.

Inverse Laplace Transform

Theorem (F., Hwang, Zacharovas)

Let the Laplace transform of $\tilde{f}(z)$ exist and be analytic in $\mathbb{C} \setminus (-\infty, 0]$.

Assume that

$$\mathscr{L}[\tilde{f}; s] = \mathcal{O}\left(|s|^{-\alpha}\right)$$

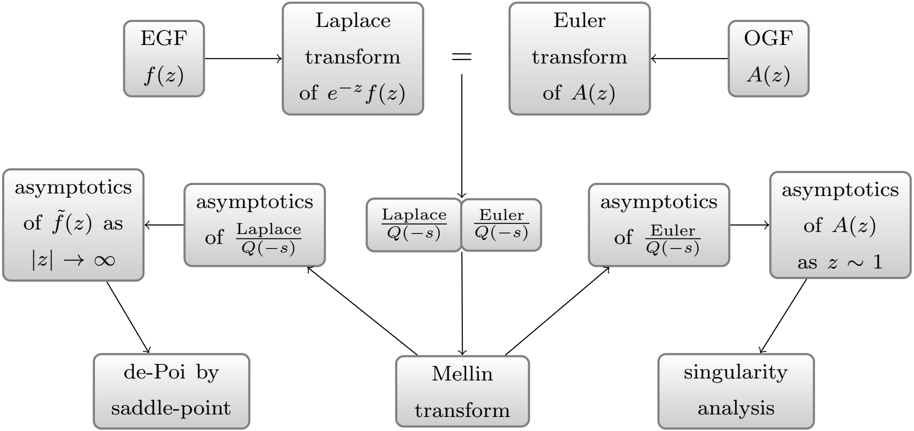
uniformly for $|s| \to 0$ and $|\arg(s)| \leq \pi - \epsilon$.

Then,

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

uniformly for $|z| \to \infty$ and $|\arg(z)| \leq \pi/2 - \epsilon$.

Our Approach vs. Flajolet-Richmond



Main Result

Theorem (F., Lee, Prodinger)

We have,

$$\mathrm{Var}(C_n) \sim \sum_k g_k n^{\chi_k},$$

where

$$g_k = \frac{Q_\infty}{L\Gamma(1+\chi_k)} \sum_{h,l,j\ge 0} \frac{(-1)^j q^{h+l+\binom{j+1}{2}}}{Q_h Q_l Q_j} \varphi(\chi_k; q^{h+j} + q^{l+j}).$$

Here, $\chi_k = 2k\pi i/L, L = \log(1/q), Q_j = \prod_{l=1}^{j}(1-q^l)$ *and*

$$\varphi(\chi; x) = \begin{cases} \pi(x^\chi - 1)/(\sin(\pi\chi)(x-1)), & \text{if } x \neq 1; \\ \pi\chi/\sin(\pi\chi), & \text{if } x = 1. \end{cases}$$

An Identity

Corollary (F., Lee, Prodinger)

We have,

$$\frac{Q_\infty}{L}\sum_{h,l,j\geq 0}\frac{(-1)^j q^{h+l+\binom{j+1}{2}}}{Q_hQ_lQ_j}\psi(q^{h+j}+q^{l+j})$$
$$=\frac{\pi^2}{6L^2}-\alpha-\beta+\frac{1}{12}-\frac{1}{L}\sum_{l\geq 1}\frac{1}{l\sinh(2l\pi^2/L)},$$

where

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

We have a direct proof for this using tools from q-analysis.

Total Path Length (i)

T_n: total path length in symmetric digital search tree.

Theorem (F., Hwang, Zacharovas)

We have

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

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

$$C_{\mathit{var}} = \frac{Q_\infty}{L} \sum_{j,h,l \geq 0} \frac{(-1)^j 2^{-\binom{j+1}{2}}}{Q_j Q_h Q_l 2^{h+l}} \delta(2^{-j-h} + 2^{-j-l}),$$

where

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

Total Path Length (ii)

Variance of total path length was also derived by Kirschenhofer, Prodinger and Szpankowski with different expression for C_{var}.

Total Path Length (ii)

Variance of total path length was also derived by Kirschenhofer, Prodinger and Szpankowski with different expression for C_{var}.

To describe their expression we need:

- Let $[FG]_0$ denote the 0-th Fourier coefficient of the product of the two Fourier series $F(z)$ and $G(z)$.
- Put

$$F(z) = \frac{1}{L}\sum_{l\neq 0}\Gamma(-1-\chi_l)e^{2l\pi iz}$$

  and

$$H(z) = -\frac{1}{L}\sum_{l\neq 0}\left(1-\frac{\chi_l}{2}\right)\Gamma(-\chi_l)e^{2l\pi iz}.$$

$$
\begin{aligned}
C_{\mathsf{var}} = &-\frac{28}{3L}-\frac{39}{4}-2\sum_{l\ge1}\frac{l2^l}{(2^l-1)^2}+\frac{2}{L}\sum_{l\ge1}\frac{1}{2^l-1}+\frac{\pi^2}{2L^2}+\frac{2}{L^2}\\
&-\frac{2}{L}\sum_{l\ge3}\frac{(-1)^{l+1}(l-5)}{(l+1)l(l-1)(2^l-1)}\\
&+\frac{2}{L}\sum_{l\ge1}(-1)^l2^{-\binom{l+1}{2}}\left(\frac{L(1-2^{-l+1})/2-1}{1-2^{-l}}-\sum_{r\ge2}\frac{(-1)^{r+1}}{r(r-1)(2^{r+l}-1)}\right)\\
&-\frac{2Q(1)}{L}+\sum_{l\ge2}\frac{1}{2^lQ_l}\sum_{r\ge0}\frac{(-1)^r2^{-\binom{r+1}{2}}}{Q_r}Q_{r+l-2}\cdot\\
&\cdot\left(-\sum_{j\ge1}\frac{1}{2^{j+r+l+2}-1}\left(2^{l+1}-2l-4+2\sum_{i=2}^{l-1}\binom{l+1}{i}\frac{1}{2^{r+i-1}-1}\right)\right.\\
&\qquad+\frac{2}{(1-2^{-l-r})^2}+\frac{2l+2}{(1-2^{1-l-r})^2}-\frac{2}{L}\frac{1}{1-2^{1-l-r}}+\frac{2}{L}\sum_{j=1}^{l+1}\binom{l+1}{j}\frac{1}{2^{r+j}-1}\\
&\qquad\left.-2\sum_{j=2}^{l+1}\binom{l+1}{j}\frac{1}{2^{r+j-1}-1}+\frac{2}{L}\sum_{j=0}^{l+1}\binom{l+1}{j}\sum_{i\ge1}\frac{(-1)^i}{(i+1)(2^{r+j+i}-1)}\right)\\
&+\sum_{l\ge3}\sum_{r=2}^{l-1}\binom{l+1}{r}\frac{Q_{r-2}Q_{l-r-1}}{2^lQ_l}\sum_{j\ge l+1}\frac{1}{2^j-1}-2[FH]_0-[F^2]_0.
\end{aligned}
$$

Approximate Counting with m Counters (i)

Cichoń and Macyna:

Consider m counters. When counting the n-th object choose one uniform at random.

Approximate Counting with m Counters (i)

Cichoń and Macyna:

Consider m counters. When counting the n-th object choose one uniform at random.

D_n: sum of all counters after counting n objects.

Approximate Counting with m Counters (i)

Cichoń and Macyna:

Consider m counters. When counting the n-th object choose one uniform at random.

D_n: sum of all counters after counting n objects.

Then,

$$D_n \stackrel{d}{=} C_{I_1}^{(1)} + \cdots + C_{I_m}^{(m)},$$

where

$$P(I_1 = n_1, \ldots, I_m = n_m) = \frac{1}{m^n}\binom{n}{n_1, \ldots, n_m}.$$

Approximate Counting with m Counters (i)

Cichoń and Macyna:

Consider m counters. When counting the n-th object choose one uniform at random.

D_n: sum of all counters after counting n objects.

Then,

$$D_n \stackrel{d}{=} C_{I_1}^{(1)} + \cdots + C_{I_m}^{(m)},$$

where

$$P(I_1 = n_1, \ldots, I_m = n_m) = \frac{1}{m^n} \binom{n}{n_1, \ldots, n_m}.$$

Let $\tilde{f}_D, \tilde{f}_C$ denote Poisson mean of D_n and C_n. Similar, let $\tilde{V}_D$ and $\tilde{V}_C$ denote Poisson variance of D_n and C_n.

Approximate Counting with m Counters (ii)

Poisson model:

$$\tilde{f}_D(z) = m\tilde{f}_C(z/m),$$
$$\tilde{V}_D(z) = m\tilde{V}_C(z/m).$$

Approximate Counting with m Counters (ii)

Poisson model:

$$\tilde{f}_D(z) = m\tilde{f}_C(z/m),$$
$$\tilde{V}_D(z) = m\tilde{V}_C(z/m).$$

From this, we obtain the following result.

Theorem (F., Lee, Prodinger)

We have,

$$\mathbb{E}(D_n) \sim m\log_{1/q}(n/m) + mC_{mean} + mF(\log_{1/q}(n/m)),$$
$$\mathrm{Var}(D_n) \sim mC_{var} + mG(\log_{1/q}(n/m)),$$

where C_{mean}, C_{var} and $F(z), G(z)$ are as before.

Approximate Counting with m Counters (iii)

Another variant of approximate counting with m counters:

Consider m counters. Use a counter until it is increased; then cyclically move on to the next.

Approximate Counting with m Counters (iii)

Another variant of approximate counting with m counters:

Consider m counters. Use a counter until it is increased; then cyclically move on to the next.

D_n: sum of all counters after counting n objects.

Approximate Counting with m Counters (iii)

Another variant of approximate counting with m counters:

Consider m counters. Use a counter until it is increased; then cyclically move on to the next.

D_n: sum of all counters after counting n objects.

Then,

$$D_n \stackrel{d}{=} X_n,$$

where

$$X_{n+m} \stackrel{d}{=} X_{I_n} + m.$$

X_n is the length of the leftmost path in random bucket digital search tree with bucket size m.

Approximate Counting with m Counters (iv)

Recurrence of moments:

$$f_{n+m} = \sum_{j=0}^{n} \binom{n}{j} q^j p^{n-j} f_j + g_n.$$

Approximate Counting with m Counters (iv)

Recurrence of moments:

$$f_{n+m} = \sum_{j=0}^{n} \binom{n}{j} q^j p^{n-j} f_j + g_n.$$

Poissonized variance $\tilde{V}(z)$ satisfies the differential function equation:

$$\sum_{i=0}^{m} \binom{m}{i} \tilde{V}^{(i)}(z) = \tilde{V}(qz) + \tilde{g}(z),$$

where

$$\tilde{g}(z) = \left(\sum_{i=0}^{m} \binom{m}{i} \tilde{f}_1^{(i)}(z) \right)^2 - \sum_{i=0}^{m} \binom{m}{i} \left(\tilde{f}_1(z)^2 \right)^{(i)}.$$

Approximate Counting with m Counters (v)

Theorem (F., Lee, Prodinger)

We have,

$$\mathrm{Var}(D_n) \sim \sum_k g_k n^{\chi_k},$$

where

$$g_k = \frac{1}{L\Gamma(1+\chi_k)} \int_0^\infty \frac{s^{\chi_k}}{Q(-s/q)^m} \left(p(s) + \int_0^\infty e^{-sz} \tilde{g}(z) \mathrm{d}z \right) \mathrm{d}s$$

and

$$p(s) = \frac{(s+1)^m - 1 - ms}{s^2}.$$