國立交通大學

應用數學系

碩 士 論 文

數位搜尋樹的機率演算分析

Probabilistic Analysis of Digital Search Trees
– Old and New Results

研 究 生：曾柏翰

指導教授：符麥可　教授

中 華 民 國 九 十 八 年 六 月

數位搜尋樹的機率演算分析

Probabilistic Analysis of Digital Search Trees – Old and New Results

研 究 生：曾柏翰　　Student：Po-Han Tseng

指導教授：符麥可　　Advisor：Michael Fuchs

國 立 交 通 大 學

應 用 數 學 系

碩 士 論 文

A Thesis

Submitted to Department of Applied Mathematics

College of Electrical Engineering and Computer Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Applied Mathematics

June 2009

Hsinchu, Taiwan, Republic of China

中華民國九十八年六月

Master Thesis

Probabilistic Analysis of Digital Search Trees - Old and New Results

Po-Han Tseng

Department of Applied Mathematics,
National Chiao Tung University

This thesis was supervised by Dr. Michael Fuchs

摘　　要

數位搜尋樹(digital search trees, DSTs for short)與桶型數位搜尋樹(bucket DSTs,每個的節點最多可儲存 b 筆資料,b-DSTs for short)為電腦科學中基本的資料結構。這兩種資料結構由具有 0-1 數列的儲存資料所組成。此篇論文中，我們考慮隨機生成的 DSTs。

在這十年來，幾乎所有關於隨機 DSTs 的重要參數(parameters)都有研究結果出現。如：深度(Depth)，距離(Distance)，外部-內部節點(External-internal nodes)，內節點路徑長度(Internal path length)和大小(Size)。這些研究結果中有用到許多的分析方法，其中最主要都在解析組合的範疇內。

在此論文中，我們主要著重於探討 DSTs 的內節點路徑長度。我們將介紹近年發展出來之研究結果，與其使用到之分析技術。除此之外，還會介紹一個全新的方法，此法將由 Fuchs、Hwang 和 Zacharovas 在以後的研究中發表。此方法將會改進對 b-DST 上內節點路徑長度的分析。

這份論文的主要目地有兩個：第一，我們給出近年來關於 DSTs 之內節點路徑長度的分析方法與研究結果，和其他參數的研究結果整理。此外我們也給了一些分析技術上的改進。第二，我們提出一個全新的分析方法，也得到一個對於 b-DSTs 上內節點路徑長度更加簡單的結果。

論文組織如下：第一章介紹內節點路徑長度研究使用之分析技術。第

二章為內節點路徑長度與其他參數的期望值(mean)與變異數(variance)之研究結果整理。第三章中介紹新的方法並給出我們的主要結果。

Preface

Digital search trees (DSTs for short) and their generalizations such as bucket digital search trees (b-DSTs) are fundamental data structures in computer science. These trees are built from records whose keys consist of 0-1 strings. In this thesis, we will consider random DSTs which are obtained by assuming that the bits of the keys are randomly generated.

Characteristic parameters of random DSTs are random variables and their analysis has attracted a lot of attention in recent decades. Examples of parameters considered in previous works include: the depth of a random node [5, 12, 15, 17, 18, 19, 20, 22], the distance of two random nodes [1], the number of external-internal nodes [5, 9, 15, 21, 13], the internal path length [5, 8, 10, 14], and the size of the tree [4, 9]. For the analysis, several interesting methods have been proposed, most of them belonging to the field of analytic combinatorics.

In this thesis, we focus on the internal path length of DSTs. We will introduce the techniques which have been devised for the analysis of the internal path length. Moreover, we will give a new method, which will appear in a forthcoming work of Fuchs, Hwang, and Zacharovas, to improve the analysis of the internal path length of b-DSTs.

The purpose of this thesis is twofold. First, we want to give a self-contained survey of the techniques used in the analysis of DSTs and the results achieved. Here, we will mainly follow previous works, but also introduce some technical improvements. Secondly, we are going to use the new approach of Fuchs, Hwang, and Zacharovas mentioned above to obtain exact and numerical results concerning the leading constant in the asymptotic expansion of the variance. In particular, our results will simplify and improve previous results.

This thesis is organized as follows: in Chapter 1, we introduce the techniques which are of importance in the analysis of DSTs. In Chapter 2, we present results concerning mean value and variance of the internal path length and explain how they can be proved with the methods from Chapter 1. Moreover, we also give a short survey of results concerning other parameters. In Chapter 3, we introduce the new method and explain our new findings concerning the leading constant of the variance.

誌　　謝

首先，我最想要感謝的人，就是我的指導老師 Dr. Michael Fuchs。從這篇論文開始動工前，Michael 與我花了很多時間研讀很多關於 Digital Search Trees 的論文，慢慢的搞清楚到目前為止 DSTs 的研究結果與其中所使用的方法。除此之外，在這篇論文的撰寫中，Michael 也給了我相當大的協助與指導，讓這篇論文能夠達到讓初次接觸 DSTs 或是這個領域的人有系統的學習。

另外，我也很感謝我的兩位口試委員，交大的陳秋媛教授與海大的程華淮教授。他們兩位都給了我關於這篇論文許多的意見，讓這篇論文能更加完善與嚴謹。

我當然沒有忘記 96 應數所的同學們，哈哈。雖然研究所的日子很難熬，但我們還是一起走過來了。該畢業的終究會畢業，還沒畢業的，那只是老闆還沒點頭而已，加油！

還有，不可或缺的，就是在我生命中給我最多也最豐富的家人們。讓我在新竹用功時(雖然有時會偷懶)，還是給予我最大的支持。喔對，還有我家那兩隻可愛的小狗，奶雞與都胖(雖然他們什麼都不懂，老是想著吃還有去公園)。

最後，永遠別忘記那些曾出現在生命中的美麗。

Contents

List of Figures

List of Tables

Chapter 1

Some techniques

In this chapter, we collect some analytic techniques, such as Rice method [6] (in Section 1.1), Mellin transform [2] (in Section 1.2), Poisson transform [11] (in Section 1.3) and singularity analysis [3] (in Section 1.4). These methods will be the main tools for deriving our results in Chapter 2 and Chapter 3.

1.1 Rice Method

Rice method is fruitful for finding the asymptotic expansion of sums of the form

$$\sum_{k=0}^{n} \binom{n}{k} (-1)^k f(k). \tag{1.1}$$

The starting point is the integral representation:

Lemma 1. *Let C be a positive oriented closed curve encircling the points $0, 1, \cdots, n$, and let $f(z)$ be a function which is analytic with in C. Then, we have*

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

Proof. This follows by an application of the residue theorem: The integral equals $2\pi i$ times the sum of the residues of the simple poles at the points $0, 1, \cdots, n$. For each k, we have

$$\operatorname*{Res}_{z=k} f(z) \frac{n!}{z(z-1)\cdots(z-n)} = (-1)^{n-k} \frac{n!}{k!(n-k)!} f(k). \quad \blacksquare$$

Remark. The kernel of the integral could be written as

$$\frac{n!}{z(z-1)\cdots(z-n)} = \frac{\Gamma(n+1)\Gamma(z-n)}{\Gamma(z+1)} = (-1)^{n-1}B(n+1,-z),$$

where $B(x,y)$ is the classical Beta function.

Remark. Sometimes the sum might be taken over the integers from $n_0,\cdots,n$. Then Lemma 1 still holds when C is changed to enclose just those points.

Rice method. Suppose we have an explicit sum of type (1.1). Then the Rice method allows us to compute an asymptotic expansion by using the following steps:

Step 1. Extend f_k which is defined only on the integers to an appropriate meromorphic function $f(z)$ which is analytic at the points $0,1,\cdots,n$.

Step 2. Choose a suitable contour C which encircles the points $0,1,\cdots,n$ and consider the integral

$$\Delta = \frac{(-1)^n}{2\pi i}\int_C f(z)\frac{n!}{z(z-1)\cdots(z-n)}\,\mathrm{d}z.$$

Step 3. By the residue theorem we obtain

$$\Delta = \sum_{k=0}^{n}\binom{n}{k}(-1)^k f(k) + \left\{\begin{smallmatrix}\text{Contributions from the other}\\ \text{poles inside the contour C.}\end{smallmatrix}\right\}$$

Step 4. Estimate Δ.

To carry out Step 4 one often needs growths properties of $f(z)$. Therefore, we give the following definition:

Definition 1. *A function $f(z)$ is said to be of polynomial growth in an unbounded domain Ω if it is analytic in Ω and satisfies*

$$|f(z)| = \mathcal{O}(|z|^r), \tag{1.2}$$

for some non-negative integer r as $z\to\infty$ in Ω.

Remark. Suppose $f(z)$ is of polynomial growth. Then, the integral

$$\int_C f(z)\frac{n!}{z(z-1)\cdots(z-n)}\,\mathrm{d}z \to 0$$

as C becomes large (for instance if we choose larger and larger circles).

The following are two examples to demonstrate the Rice method.

Example 1. Consider the sum

$$S_n = \sum_{k=1}^{n} \binom{n}{k} \frac{(-1)^k}{k}.$$

Step 1. $f(z) = 1/z$ is obviously a suitable extension of the sequence $1/k$.

Step 2. We choose the curve C to be a circle with radius larger than n centered at 0.

Step 3. The kernel of Δ has a double pole at 0, simple poles at $1, 2, \cdots, n$, and is analytic everywhere else. Thus

$$\begin{aligned}\Delta =& S_n + (-1)^n \operatorname*{Res}_{z=0}\Big[\frac{n!}{z^2(z-1)\cdots(z-n)}\Big] \\ =& S_n + (-1)^n \Big[\sum_{k=1}^{n} \frac{n!}{(z-1)\cdots(z-k)^2\cdots(z-n)}\Big]_{z=0} \\ =& S_n + \sum_{k=1}^{n} \frac{1}{k}.\end{aligned}$$

Step 4. Clearly, $f(z)$ is of polynomial growth, thus Δ converges to 0 as soon as C becomes large.

Hence, we have

$$-S_n = \sum_{k=1}^{n} \frac{1}{k} = H_n = \log n + \gamma + \mathcal{O}(n^{-1}),$$

where H_n are the harmonic numbers and $\gamma = 0.57721\cdots$ is the Euler number. The asymptotics of the harmonic numbers is well-known (see Example 5 in Section 1.2 for a proof).

Example 2. Consider the sum

$$A_n = \sum_{k\geqslant 2} \binom{n}{k} (-1)^k Q_{k-2}, \qquad n \geqslant 1.$$

where $Q_n = \prod_{1\leq j\leq n} (1 - 2^{-j})$.

Step 1. We introduce the function

$$Q(x) = \Big(1 - \frac{x}{2}\Big)\Big(1 - \frac{x}{4}\Big)\Big(1 - \frac{x}{8}\Big)\cdots .$$

Note that $Q_n = Q_\infty/Q(2^{-n})$ where $Q_\infty := Q(1) = \lim_{n\to\infty} Q_n = 0.288788\cdots$, and $Q_\infty/Q(2^{-z+2})$ is analytic on $[2,\infty)$ which gives the appropriate extension.

Step 2. Take as C a large segment of the line $\Re(s) = 1/2$ closed to the right by a large semi-circle which encloses the points $2, 3, \cdots, n$.

Step 3. Note that the zeros of $Q(2^{-z+2})$ all satisfy $2^{-z+j} = 1$ with $j \le 1$. Thus, the kernel of Δ has poles at $1 \pm (2\pi ik)/\log 2$ (one double pole at $k = 0$ and single poles for all k with $k \neq 0$) inside C. To find the contribution at 1 we use Taylor expansion.

Here the following fact will turn out to be useful:

If $G(z) = \prod_{k\in R} g_k(z)$, then $G'(z)/G(z) = \sum_{k\in R} g_k'(z)/g_k(z)$. From this it follows that if $F(z) = \prod_{j\in\ R}\big(1 - f_j(z)\big)^{-1}$ for some index set R, then the Taylor series expansion of F at a, if it exists, is given by

$$F(z) = F(a)\Big(1 + \sum_{j\in R}\frac{f_j'(a)}{1 - f_j(a)}(z-a) + \mathcal{O}(z-a)^2\Big).$$

Consequently we obtain the series expansions

$$\begin{aligned}\frac{n!}{z(z-1)\cdots(z-n)} =& \frac{1}{z(z-1)}\prod_{2\le j\le n}(1 - z/j)^{-1}\\ =& \frac{n}{z-1}\Big(1 + (H_{n-1} - 1)(z-1) + \mathcal{O}\big((z-1)^2\big)\Big)\\ =& \frac{n}{z-1} + n(H_{n-1} - 1) + \mathcal{O}(z-1).\end{aligned}$$

And

$$\begin{aligned}Q_\infty/Q(2^{-z+1}) =& Q_\infty\prod_{j<1}(1 - 2^{-z+j})^{-1}\\ =& 1 - \log 2\sum_{j<1}\frac{2^{j-1}}{1 - 2^{j-1}}(z-1) + \mathcal{O}(z-1)^2\\ =& 1 - \alpha\log 2(z-1) + \mathcal{O}(z-1)^2,\end{aligned}$$

where $\alpha = 1 + \frac{1}{3} + \frac{1}{7} + \cdots$. Thus, we obtain

$$\begin{aligned}\frac{Q_\infty}{Q(2^{-z+2})}\frac{n!}{z(z-1)\cdots(z-n)} =& \frac{1}{1-2^{-z+1}}\frac{Q_\infty}{Q(2^{-z+1})}\frac{n!}{z(z-1)\cdots(z-n)}\\ =& \Big(\frac{1}{(z-1)\log 2} + \frac{1}{2} + \mathcal{O}(z-1)\Big)\\ & \times (1 - \alpha\log 2(z-1) + \mathcal{O}(z-1)^2)\\ & \times \Big(\frac{n}{z-1} + n(H_{n-1}-1) + \mathcal{O}(z-1)\Big).\end{aligned}$$

The residue at $z = 1$ is the coefficient of $1/(z-1)$ in the above product:

$$\frac{n}{\log 2}(H_{n-1}-1) - n\Big(\alpha - \frac{1}{2}\Big) = n\log_2 n + n\Big(\frac{\gamma-1}{\log 2} - \alpha + \frac{1}{2}\Big) + \mathcal{O}(1).$$

The poles at $1 \pm 2\pi i k/\log 2$ with $k \neq 0$ add a small contribution $\delta(n)$ to the linear term [5], where

$$\delta(n) = \frac{1}{\log 2}\sum_{k\neq 0}\Gamma\Big(-1 - \frac{2k\pi i}{\log 2}\Big)e^{2k\pi i\log_2 n}.$$

Step 4. On the right semi-circle, Δ converges to 0 as C becomes large since

$$|Q^{-1}(2^{-z+2})| = \prod_{j\le 1}(1-2^{-|z|+j})^{-1} = \mathcal{O}(|z|^0)$$

as $|z| \to \infty$. On the left segment we have the bound

$$\mathcal{O}\left(\int_{-\infty}^{\infty}\frac{\Gamma(n+1)}{\Gamma(n+1/2-iy)}\,\mathrm{d}y\right) = \mathcal{O}(n^{1/2}).$$

Thus we have

$$A_n = n\log_2 n + n\left(\frac{\gamma-1}{\log 2} - \alpha + \frac{1}{2} + \delta(n)\right) + \mathcal{O}(n^{1/2}). \tag{1.3}$$

1.2 Mellin Transform

The Mellin transform (Hjalmar Mellin 1854–1933, Finish mathematician) is the most popular transform in the analysis of algorithms.

Definition 2. *Let $f(x)$ be a continuous function over $(0,\infty)$. Its Mellin transform $f^*(s)$ is defined by*

$$f^*(s) = \mathcal{M}[f(x);s] = \int_0^\infty f(x)x^{s-1}\,\mathrm{d}x.$$

Table 1.1: Some common Mellin transforms.

$f(x)$	$f^*(s)$	$\langle \alpha, \beta \rangle$
e^{-x}	$\Gamma(s)$	$\langle 0, +\infty \rangle$
e^{-x^2}	$\frac{1}{2}\Gamma(\frac{1}{2}s)$	$\langle 0, +\infty \rangle$
$\frac{1}{1+x}$	$\frac{\pi}{\sin \pi s}$	$\langle 0, 1 \rangle$
$\log(1+x)$	$\frac{\pi}{s \sin \pi s}$	$\langle -1, 0 \rangle$
$H(x) \equiv 1_{0<x<1}$	$\frac{1}{s}$	$\langle 0, +\infty \rangle$
$x^\alpha (\log x)^k H(x)$	$\frac{(-1)^k k!}{(s+\alpha)^{k+1}}$	$\langle -\alpha, +\infty \rangle$, k integer

Basic properties. The following lemma gives the conditions for the existence of the Mellin transform of a given function $f(x)$.

Lemma 2. *The conditions*

$$f(x) \underset{x \to 0^+}{=} \mathcal{O}(x^u); \qquad f(x) \underset{x \to +\infty}{=} \mathcal{O}(x^v),$$

when $u > v$, guarantee that $f^*(s)$ exists in the strip $-u < \Re(s) < -v$.

Proof. From the decomposition

$$\begin{aligned} \left| \int_0^\infty f(x) x^{s-1} \, \mathrm{d}x \right| &\le \int_0^1 |f(x)| x^{\Re(s)-1} \, \mathrm{d}x + \int_1^\infty |f(x)| x^{\Re(s)-1} \, \mathrm{d}x \\ &\le \alpha \int_0^1 x^{u+\Re(s)-1} \, \mathrm{d}x + \beta \int_1^\infty x^{v+\Re(s)-1} \, \mathrm{d}x, \end{aligned}$$

where α, β are some constants. The first integral exists for $u + \Re(s) > 0$ and the second for $v + \Re(s) < 0$. Thus $f^*(s)$ exists in the strip $-u < \Re(s) < -v$. ∎

Remark. From the above lemma we see that the domain of existence of a Mellin transform is a complex strip, and the largest one is called the fundamental strip. We introduce the notation $\langle \alpha, \beta \rangle$ for the open strip of complex numbers s such that $\alpha < \Re(s) < \beta$.

Table 1.1 presents some common Mellin transforms with their corresponding fundamental strips. These formulas are simple and easy to check.

Moreover, some basic transformation rules are given in Table 1.2. These rules are also easy to confirm.

Table 1.2: Functional properties of Mellin transform.

	$f(x)$	$f^*(s)$	$\langle \alpha, \beta \rangle$		
F_1	$x^\nu f(x)$	$f^*(s+v)$	$\langle \alpha - \nu, \beta - \nu \rangle$		Shift
F_2	$f(x^\rho)$	$\frac{1}{\rho} f^*(\frac{s}{\rho})$	$\langle \rho\alpha, \rho\beta \rangle$	$\rho > 0$	Multiple
	$f(1/x)$	$-f^*(-s)$	$\langle -\beta, -\alpha \rangle$		
F_3	$f(\mu x)$	$\frac{1}{\mu^s} f^*(s)$	$\langle \alpha, \beta \rangle$	$\mu > 0$	
	$\sum_k \lambda_k f(\mu_k x)$	$(\sum_k \lambda_k \mu_k^{-s}) \cdot f^*(s)$			By linearity
F_4	$f(x) \log x$	$\frac{\mathrm{d}}{\mathrm{d}s} f^*(s)$	$\langle \alpha, \beta \rangle$		Differential
F_5	$\Theta f(x)$	$-s f^*(s)$	$\langle \alpha', \beta' \rangle$	$\Theta = x\frac{\mathrm{d}}{\mathrm{d}x}$	
	$\frac{\mathrm{d}}{\mathrm{d}x} f(x)$	$-(s-1) f^*(s-1)$	$\langle \alpha' + 1, \beta' + 1 \rangle$		
	$\int_0^x f(t)\,\mathrm{d}t$	$-\frac{1}{s} f^*(s+1)$			

Inversion. We can see that the Mellin transform is closely related to the Fourier transforms (as well as the Laplace transform): Let $x = e^{-y}$ and $s = \sigma + it$, we obtain

$$f^*(s) = \int_0^\infty f(x) x^{s-1}\,\mathrm{d}x = \int_{-\infty}^\infty f(e^{-y}) e^{-\sigma y} e^{-ity}\,\mathrm{d}y.$$

Thus the Mellin transform turns into a Fourier transform, and the inversion theorem for the Mellin transform follows from that for the Fourier transform.

Theorem 1. *Let $f(x)$ be continuous on $(0, \infty)$ and assume that its Mellin transform has fundamental strip $\langle a, b \rangle$. Then*

$$f(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} f^*(s) x^{-s}\,\mathrm{d}s, \tag{1.4}$$

where $a < c < b$.

Asymptotic properties. The usefulness of the Mellin transform comes from its asymptotic properties as we will see below. In particular we have two important results, namely, the direct and converse mapping theorem.

Before we can give these results, we give the notation of the singular expansion: For a meromorphic function $\phi(s)$ with poles in Ω, the singular expansion is

$$\phi(s) \asymp \sum_{k \in \Omega} \triangle_k(s),$$

where $\triangle_k(s)$ is the Laurent expansion of ϕ around $s = k$ up to at most $\mathcal{O}(1)$ term. For example, since

$$\frac{1}{s(s-1)} = -\frac{1}{s} - 1 + \mathcal{O}(s) \quad (s \to 0), \quad \text{and}$$
$$\frac{1}{s(s-1)} = \frac{1}{s-1} - 1 + \mathcal{O}((s-1)) \quad (s \to 1),$$

then we write

$$\frac{1}{s(s-1)} \asymp \left[-\frac{1}{s} - 1\right]_{s=0} + \left[\frac{1}{s-1} - 1\right]_{s=1}$$

for the singular expansion of $1/s(s-1)$.

The prototype of the direct mapping is the function e^{-x}: we know its Taylor expansion at 0 is

$$e^{-x} = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} x^k,$$

and its Mellin transform

$$\int_0^\infty e^{-x} x^{s-1}\, \mathrm{d}x = \Gamma(s) = \frac{\Gamma(s+k+1)}{s(s+1)(s+2)\cdots(s+k)}.$$

That means $\Gamma(s)$ has poles at the points $s = -k$ with positive integer k, and hence we have the singular expansion

$$\Gamma(s) \asymp \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \frac{1}{s+k} \qquad (s \in \mathbb{C}).$$

We can observe that one can map the Taylor expansion to coincide with the singular expansion by the rule

$$x^k \mapsto \frac{1}{s+k}.$$

In fact, this is a general phenomenon.

Theorem 2. *Let $f(x)$ be continuous with its Mellin transform $f^*(s)$ having nonempty fundamental strip $\langle \alpha, \beta \rangle$.*

(i) [Asymptotics for $x \to 0$] *Assume that $f(x)$ has the following asymptotic expansion as $x \to 0$*

$$f(x) = \sum_{\xi,k} c_{\xi,k} x^\xi (\log x)^k + \mathcal{O}(x^\gamma), \tag{1.5}$$

where $-\gamma < -\xi \leq \alpha$ *and* k *is non-negative. Then* $f^*(s)$ *is continuable to the strip* $\langle -\gamma, \beta \rangle$ *and*

$$f^*(s) \asymp \sum_{\xi,k} c_{\xi,k} \frac{(-1)^k k!}{(s+\xi)^{k+1}} \qquad (s \in \langle -\gamma, \beta \rangle). \tag{1.6}$$

(ii) [Asymptotics for $x \to \infty$] *Assume that* $f(x)$ *has the asymptotic expansion of form* (1.5) *where now* $\beta \leq -\xi < -\gamma$ *as* $x \to \infty$. *Then* $f^*(s)$ *is continuable to the strip* $\langle \alpha, -\gamma \rangle$ *and*

$$f^*(s) \asymp -\sum_{\xi,k} c_{\xi,k} \frac{(-1)^k k!}{(s+\xi)^{k+1}} \qquad (s \in \langle \alpha, -\gamma \rangle). \tag{1.7}$$

Proof. Since $\mathcal{M}[f(1/x); s] = -\mathcal{M}[f(x); -s]$, we only need to prove the case $x \to 0$. By assumption, the function

$$g(x) = f(x) - \sum_{\xi,k} c_{\xi,k} x^{\xi} (\log x)^k$$

is $\mathcal{O}(x^{\gamma})$. In the fundamental strip we also have

$$f^*(s) = \int_0^1 g(x) x^{s-1} \, \mathrm{d}x + \int_0^1 \sum_{\xi,k} c_{\xi,k} x^{s+\xi-1} (\log x)^k \, \mathrm{d}x + \int_1^{\infty} f(x) x^{s-1} \, \mathrm{d}x.$$

The first integral is analytic in $\langle -\gamma, \infty \rangle$ and the third one in $\langle -\infty, \beta \rangle$. Thus the sum of those two is analytic in the strip $\langle -\gamma, \beta \rangle$. After integrating the second integral becomes

$$\sum_{\xi,k} c_{\xi,k} \frac{(-1)^k k!}{(s+\xi)^{k+1}}.$$

Hence, $f^*(s)$ exists in $\langle -\gamma, \beta \rangle$ and has the singular expansion of the form (1.6). ∎

Remark. From the proof of Theorem 2, we can see that there is a principle: Let $g(x)$ be a truncated asymptotic expansion of a given function $f(x)$ at either 0 or ∞. Then the Mellin transform of $f(x) - g(x)$ does not change, but only the fundamental strip gets shifted. For example, $\mathcal{M}[e^x - 1; s] = \Gamma(s)$ with the fundamental strip $\langle -1, 0 \rangle$, and $\mathcal{M}[e^x - 1 + x; s] = \Gamma(s)$ with the fundamental strip $\langle -2, -1 \rangle$.

The following example appears in the Table 1.1.

Example 3. The function $f(x) = (1+x)^{-1}$ has fundament strip $\langle 0, 1\rangle$ and its Mellin transform is

$$f^*(s) = \int_0^\infty (1+x)^{-1} x^{s-1}\,\mathrm{d}x = \Gamma(1-s)\Gamma(s) = \frac{\pi}{\sin \pi s}.$$

Then the two expansions

$$\frac{1}{1+x} = \sum_{n=0}^{\infty} (-1)^n x^n \quad (x \to 0), \quad \text{and}$$
$$\frac{1}{1+x} = \sum_{n=1}^{\infty} (-1)^{n-1} x^{-n} \quad (x \to +\infty),$$

translate into

$$f^*(s) \asymp \sum_{n=0}^{\infty} \frac{(-1)^n}{s+n} \quad (s \in \langle -\infty, 1\rangle), \quad \text{and}$$
$$f^*(s) \asymp \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{s-n} \quad (s \in \langle 0, \infty\rangle).$$

This is consistent with the known form,

$$f^*(s) = \frac{\pi}{\sin \pi s} \asymp \sum_{n\in\mathbb{Z}} \frac{(-1)^n}{s+n} \quad (s \in \mathbb{C}). \tag{1.8}$$

The next question that arises is whether or not a converse of the direct mapping theorem still holds. Under some conditions the answer is yes as the following theorem demonstrates:

Theorem 3. *Let $f(x)$ be continuous with its Mellin transform $f^*(s)$ having nonempty fundamental strip $\langle \alpha, \beta\rangle$.*

(i) [Asymptotics for $x \to 0$] *Assume that $f^*(s)$ admits a meromorphic continuation to the strip $\langle \gamma, \beta\rangle$ for some $\gamma < \alpha$ with a finite number of poles there, and is analytic on $\Re(s) = \gamma$. Assume also that there exists a real number $\eta \in (\alpha, \beta)$ such that with $r > 1$,*

$$f^*(s) = \mathcal{O}(|s|^{-r}), \tag{1.9}$$

when $|s| \to \infty$ in $\gamma \le \Re(s) \le \eta$. If $f^*(s)$ admits the singular expansion for $s \in \langle \gamma, \alpha\rangle$,

$$f^*(s) \asymp \sum_{\xi,k} d_{\xi,k} \frac{1}{(s-\xi)^{k+1}}, \tag{1.10}$$

then an asymptotic expansion of $f(x)$ at 0 is

$$f(x) = \sum_{\xi,k} d_{\xi,k} \frac{(-1)^k}{k!} x^{-\xi} (\log x)^k + \mathcal{O}(x^{-\gamma}). \tag{1.11}$$

(ii) [Asymptotics for $x \to \infty$] *Similarly assume that $f^*(s)$ admits a meromorphic continuation to the strip $\langle \alpha, \gamma \rangle$ for some $\gamma > \beta$ and is analytic on $\Re(s) = \gamma$. Assume also that the growth condition (1.9) holds in $\langle \eta, \gamma \rangle$ for some $\eta \in (\alpha, \beta)$. If $f^*(s)$ admits the singular expansion (1.10) for $s \in \langle \beta, \gamma \rangle$, then an asymptotic expansion of $f(x)$ at ∞ is*

$$f(x) = -\sum_{\xi,k} d_{\xi,k} \frac{(-1)^k}{k!} x^{-\xi} (\log x)^k + \mathcal{O}(x^{-\gamma}). \tag{1.12}$$

Proof. As above it suffices to prove the case $x \to 0$. Let Ω be the set of poles in $\langle \gamma, \beta \rangle$, and set a large rectangle $R(T)$ with corners at the four points $\eta \pm iT$, $\gamma \pm iT$ in the direction of counter-clockwise. Assume that T is large enough such that $R(T)$ contains all poles in Ω. Consider the integral

$$J(T) = \frac{1}{2\pi i} \int_{R(T)} f^*(s) x^{-s} \, \mathrm{d}s,$$

we know $J(T)$ is equal to the sum of residues by Cauchy's theorem, which is

$$J(T) = \sum_{\xi,k} d_{\xi,k} \operatorname*{Res}_{s=\xi} \left(\frac{x^{-s}}{(s-\xi)^{k+1}} \right) = \sum_{\xi,k} d_{\xi,k} \frac{(-1)^k}{k!} x^{-\xi} (\log x)^k.$$

Now let T tend to $+\infty$. By assumption $J(T)$ along the top and bottom lines of $R(T)$ is bounded by $\mathcal{O}(T^{-r})$ which vanishes as $T \to \infty$. On the left we have the bound of the form

$$\left| \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} f^*(s) x^{-s} \, \mathrm{d}s \right| \le \mathcal{O}(1) \int_0^\infty \frac{x^{-\gamma}}{(1+t)^r} \, \mathrm{d}t = \mathcal{O}(x^{-\gamma}).$$

On the right the integral converges to $f(x)$ by the inverse theorem (1.4) since $f(x)$ is continuous. This proves the claim. ∎

From Theorem 2 and Theorem 3 we know that the poles of $f^*(s)$ are in a one-to-one correspondence with the terms in the asymptotic expansion of $f(x)$ at either 0 or ∞.

Example 4. The function

$$f^*(s) = \Gamma(1-s)\frac{\pi}{\sin \pi s}$$

is analytic in the strip $\langle 0, 1\rangle$. Note that $\pi/\sin \pi s = \mathcal{O}(e^{-\pi|\Im(s)|})$ as $|s| \to \infty$, and a similar exponential decay holds for $\Gamma(1-s)$ by the complex version of Stirling's formula:

$$\Gamma(\sigma + it) \sim \sqrt{2\pi}|t|^{\sigma-1/2}e^{-\pi|t|/2} \qquad (t \to \infty).$$

The singular expansion of $\pi/\sin \pi s$ was already considered in (1.8). Thus for $\Re(s) < 1$, we have the singular expansion

$$f^*(s) \asymp \sum_{n=0}^{\infty}(-1)^n \frac{n!}{s+n}.$$

Then the asymptotic expansion of the original function is

$$f(x) \sim \sum_{n=0}^{\infty}(-1)^n n! x^n \qquad (x \to 0).$$

Sometimes $f^*(s)$ has a vertical line of regularly spaced poles. In this case, we need the following weaker form of the growth condition (1.9).

Corollary 1. *The conclusions of Theorem 3 remain valid assuming only a weaker form of the growth condition* (1.9) *along a countable set of horizontal segments* $|\Im(s) = T_j|$ *where* $T_j \to +\infty$.

Proof. Restrict T to belong to the discrete set T_j which must avoid the poles of $f^*(s)$ in the proof of Theorem 3. ∎

Applications. Mellin transform is effective in the asymptotic analysis of harmonic sums.

Definition 3. *1. A harmonic sum $F(x)$ is a sum of the form*

$$F(x) = \sum_k \lambda_k g(\mu_k x), \tag{1.13}$$

where λ_k are called "amplitudes", μ_k are called "frequencies", and $g(x)$ is called the "base function".

2. The Dirichlet series of the harmonic sum is the sum

$$\Lambda(s) = \sum_k \lambda_k \mu_k^{-s}. \tag{1.14}$$

Remark. A Dirichlet series (1.14) has a half-plane of absolute convergence $\langle \sigma_a, \infty \rangle$ and a half-plane of simple convergence $\langle \sigma_c, \infty \rangle$ where $\sigma_a - \sigma_c \geq 0$.

Remark. The property of polynomial growth (1.2) in a closed strip holds for many Dirichlet series.

From F_3 in Table 1.2, we have

$$\mathcal{M}\Big[\sum_{k\in\mathcal{K}} \lambda_k g(\mu_k x); s\Big] = \Big(\sum_{k\in\mathcal{K}} \lambda_k \mu_k^{-s}\Big) \cdot g^*(s)$$

where $\mathcal{K}$ is a finite set. This formula can be extended to the harmonic sums (infinite sums) as defined above:

Lemma 3. *The Mellin transform of the harmonic sum (1.13) is defined in the intersection of the fundamental strip of the transform of the base function and the domain of absolute convergence of Dirichlet series, and it is given by*

$$F^*(s) = \Lambda(s) \cdot g^*(s). \tag{1.15}$$

Proof. Since both $g^*(s)$ and the Dirichlet series are analytic in the corresponding convergence regions, the interchange of summation and integration is valid by Fubini's theorem. ∎

To apply the converse mapping theorem for harmonic sums (1.13), we have to give another definition of controlled growth (we have already introduced polynomial growth in Definition 1).

Definition 4. *A function $\phi(s)$ is said to be of exponential decrease in a closed strip if for any $r > 0$,*

$$\phi(s) = \mathcal{O}(|s|^{-r}), \tag{1.16}$$

as $|s| \to \infty$ in the strip.

Now we suppose that the Mellin transform of the base function is of exponential decrease and the Dirichlet series of the harmonic sum is of polynomial growth in an extended region of the complex plane.

Theorem 4. *Consider the harmonic sum $F(x)$. Let the transform of the base function have the fundamental strip $\langle \alpha, \beta \rangle$, and the domain of simple convergence of Dirichlet series is $\langle \sigma_c, \infty \rangle$. Assume that*

(i) $\sigma_c < \beta$ and let $\alpha' = \max(\alpha, \sigma_c)$;

(ii) $g^*(s)$ *and* $\Lambda(s)$ *admit a meromorphic continuation in* $\langle \gamma, \beta \rangle$ *and are analytic on* $\Re(s) = \gamma$, *for some* $\gamma < \alpha$;

(iii) on the closed strip $\langle \gamma, (\alpha' + \beta)/2 \rangle$, $g^*(s)$ *is of exponential decrease and* $\Lambda(s)$ *is of polynomial growth.*

Then $F(x)$ *converges for all* $x > 0$ *on* $(0, \infty)$. *An asymptotic expansion of* $F(x)$ *as* $x \to 0$ *till an error term* $\mathcal{O}(x^{-\gamma})$ *is obtained by termwise translation of the singular expansion of* $F^*(s) = \Lambda(s) g^*(s)$ *according to the rule*

$$\frac{C}{(s-\xi)^{k+1}} \mapsto C \frac{(-1)^k}{k!} x^{-\xi} (\log x)^k.$$

Proof. By Theorem 3 it suffices to show that the fundamental relation $F^*(s) = \Lambda(s) g^*(s)$. First we select an arbitrary σ in (α', β) and take σ_0 such that $\alpha' < \sigma_0 < \sigma$. Then the inversion theorem provides

$$\sum_{n=1}^{N} \lambda_n g(\mu_n x) = \frac{1}{2\pi i} \int_{\sigma_0 - i\infty}^{\sigma_0 + i\infty} \sum_{n=1}^{N} \frac{\lambda_n}{\mu_n^s} g^*(s) x^{-s} \, \mathrm{d}s.$$

Since $|\Lambda(s)| \le C(|s| + 1)$ for some constant C (see [2]) we have

$$\Big| \sum_{n=1}^{N} \frac{\lambda_n}{\mu_n^s} g^*(s) x^{-s} \Big| \le C(|s|+1) \cdot |g^*(s)| \cdot x^{-\Re(s)} = \mathcal{O}(x^{-\Re(s)}),$$

which permits to apply the dominated convergence theorem and we obtain

$$G(x) = \frac{1}{2\pi i} \int_{\sigma_0 - i\infty}^{\sigma_0 + i\infty} \Lambda(s) g^*(s) x^{-s} \, \mathrm{d}s.$$

Thus, the strip $\langle \alpha', \beta \rangle$ is included in the fundamental strip of $G(x)$. On the other hand, since

$$\Big| \sum_{n=1}^{N} \lambda_n g(\mu_n x) \Big| \le \frac{1}{2\pi} \int_{\sigma_0 - i\infty}^{\sigma_0 + i\infty} \Big| \sum_{n=1}^{N} \frac{\lambda_n}{\mu_n^s} g^*(s) x^{-s} \Big| \, \mathrm{d}s = \mathcal{O}(x^{-\Re(s)}),$$

then the dominated convergence theorem applies once more

$$F^*(s) = \lim_{N \to \infty} \int_0^{\infty} \sum_{n=1}^{N} \lambda_n g(\mu_n x) x^{s-1} \, \mathrm{d}x = \Lambda(s) \cdot g^*(s).$$

This means that $F^*(s) = \Lambda(s) g^*(s)$. ∎

Remark. Similarly, a symmetric result holds near $x \to \infty$. Thus under the condition of Theorem 4,

$$\sum_k \lambda_n g(\mu_n x) \sim \pm \sum_p \operatorname*{Res}_{s=p}\big(g^*(s)\Lambda(s)x^{-s}\big),$$

As $x \to 0$ the sum is over the poles to the left of the fundamental strip and the sign is $+$; and as $x \to \infty$ the sum is over the poles to the right of the fundamental strip and the sign is $-$.

Example 5. The harmonic number H_n is

$$H_n = \sum_{k=1}^{n} \frac{1}{k} = \sum_{k=1}^{\infty} \Big[\frac{1}{k} - \frac{1}{k+n}\Big].$$

Thus the function

$$h(x) = \sum_{k=1}^{\infty} \Big[\frac{1}{k} - \frac{1}{k+x}\Big] = \sum_{k=1}^{\infty} \frac{1}{k} \frac{x/k}{1+x/k}$$

satisfies $h(n) = H_n$ and is a harmonic sum with $\lambda_k = \mu_k = 1/k$ and $g(x) = x/(1+x)$. Its Mellin transform is

$$\begin{aligned} h^*(s) =& \mathcal{M}\Big[x(\frac{\mathrm{d}}{\mathrm{d}x}\log(1+x)); s\Big] \cdot \sum_{k=1}^{\infty} k^{s-1} \\ =& -\frac{\pi}{\sin \pi s}\zeta(1-s), \end{aligned}$$

with fundamental strip $\langle -1, 0\rangle$. Note that for fixed $\sigma < 0$, one has

$$\zeta(\sigma + it) = \mathcal{O}(|t|^{1/2-\sigma}),$$

see [24, p. 95], and the exponential decay holds for $\pi/\sin \pi s$ (see Example 4). The singular expansion to the right of this fundamental strip is

$$h^*(s) \asymp \frac{1}{s^2} - \frac{\gamma}{s} - \sum_{k=1}^{\infty} (-1)^k \frac{\zeta(1-k)}{s-k}.$$

Thus we have the expansion at ∞:

$$H_n = \log n + \gamma + \mathcal{O}(n^{-1}).$$

1.3 Poissonization and De-poissonization

Poisson transform was introduced by Kac (1949). Sometimes a Poisson version of a problem (called Poisson model) is easier to solve than the original one (called the Bernoulli model). The purpose of this section is to introduce the basics of this important method.

Table 1.3: Some Poisson transforms and their properties

g_n	$\widetilde{G}(z)$
Constant	Constant
$(-1)^n$	e^{-2z}
α^n	$e^{(\alpha-1)z}$
$\frac{n!}{(n-k)!}$, $n \geq k$	z^k
$n!$	$\frac{e^{-z}}{1-z}$
$g_n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}(f_k + h_{n-k})$, $p+q=1$	$F(pz) + H(qz)$
$g_n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} f_k h_{n-k}$, $p+q=1$	$F(pz)H(qz)$

Poisson transform. Consider a sequence (g_n), we define the Poisson transform (or Poissonization) $\widetilde{G}(z)$ as follows:

Definition 5. *Let (g_n) be a sequence. Then the Poisson transform $\widetilde{G}(z)$ of (g_n) is defined as*

$$\widetilde{G}(z) = \sum_{n \geq 0} e^{-z} g_n \frac{z^n}{n!} \tag{1.17}$$

for arbitrary complex z.

Some Poisson transforms and their properties are presented in Table 1.3. Next, we give an example that is important in applications.

Example 6. Consider the recurrence

$$g_n = a_n + \beta \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}(g_k + g_{n-k}), \quad n \geqslant 1$$

with initial value g_0. Then, we find

$$\widetilde{G}(z) = \widetilde{A}(z) + \beta\big(\widetilde{G}(pz) + \widetilde{G}(qz)\big) - g_0 e^{-z},$$

where $\widetilde{G}$ and $\widetilde{A}$ are the Poisson transforms of g_n and a_n, respectively.

General de-poissonization theorems. Now we consider a sequence (g_n) and its Poisson transform $\widetilde{G}(z)$ (we also assume that $\widetilde{G}(z)$ is entire). If $\widetilde{G}(z)$ is well-known, then one

can extract the coefficient $g_n = n![z^n](\widetilde{G}(z)e^z)$ directly. Our aim is to extract asymptotically g_n from $\widetilde{G}(z)$. Our starting point for this will be Cauchy's formula:

$$g_n = \frac{n!}{2\pi i}\oint \frac{\widetilde{G}(z)e^z}{z^{n+1}}\,\mathrm{d}z = \frac{n!}{n^n 2\pi}\int_{-\pi}^{\pi} \widetilde{G}(ne^{it})\exp(ne^{it})e^{-nit}\,\mathrm{d}t. \tag{1.18}$$

Next, we give the definition of a linear cone:

Definition 6. *The region in the complex plane*

$$\mathcal{L}_\theta = \{z : |\arg z| \le \theta\},$$

where $|\theta| < \pi/2$ is called a linear cone.

Moreover, we need the following two lemmas. The first one is well-known, and the second one is a simple extension of the Cauchy estimate.

Lemma 4. *The following identities are true:*

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^k e^{-\alpha x^2}\,\mathrm{d}x = \begin{cases} 0, & k = 1, 3, 5, \ldots \\ \frac{\alpha^{-1/2-k/2}k!}{(k/2)!2^{k+1/2}}, & k = 0, 2, 4, \ldots \end{cases}$$

and

$$\int_{\theta}^{\infty} x^k e^{-\alpha x^2}\,\mathrm{d}x = \mathcal{O}\Big(e^{-(1/2)\alpha\theta^2}\Big)$$

where θ is a positive number.

Lemma 5. *Let $\theta_0 < \pi/2$ and $\xi > 0$. Moreover, let $\Psi(z)$ be a slowly varying function (that is, for fixed t, $\lim_{x\to\infty}\big(\Psi(tx)/\Psi(x)\big) = 1$) and assume that*

$$|z| > \xi \Rightarrow |G(z)| \le B|z|^\beta \Psi(|z|) \tag{1.19}$$

for all $z \in \mathcal{L}_{\theta_0}$, where β is a real constant. Then, for all $\theta < \theta_0$ there exist B' and $\xi' > \xi$ such that for all positive integers k the following holds in $\mathcal{L}_\theta$

$$|z| > \xi' \Rightarrow |G^{\langle k\rangle}(z)| \le k!(B')^k|z|^{\beta-k}\Psi(|z|). \tag{1.20}$$

Proof. See [11]. ∎

Now, we first give a basic de-poissonization result that holds for $\widetilde{G}(z)$ with a polynomial bound in a linear cone:

Theorem 5. *Let $\widetilde{G}(z)$ be the Poisson transform of a sequence (g_n) that is assumed to be entire. Suppose that in a linear cone $\mathcal{L}_\theta$ $(\theta < \pi/2)$ both of the following two conditions hold for some real numbers $A, B, R > 0$, β and $\alpha < 1$:*

(I) *For $z \in \mathcal{L}_\theta$*

$$|z| > R \Rightarrow |\widetilde{G}(z)| \le B|z|^\beta;$$

(O) *For $z \notin \mathcal{L}_\theta$*

$$|z| > R \Rightarrow |\widetilde{G}(z)e^z| \le Ae^{\alpha|z|}.$$

Then

$$g_n = \widetilde{G}(n) + \mathcal{O}(n^{\beta-1})$$

for large n.

Proof. The proof relies on the equation (1.18). By Stirling's approximation $n! = n^n e^{-n}\sqrt{2\pi n}\big(1+\mathcal{O}(n^{-1})\big)$, we have

$$\begin{aligned} g_n =& \big(1+\mathcal{O}(n^{-1})\big)\sqrt{\frac{n}{2\pi}}\int_{-\pi}^{\pi} \widetilde{G}(ne^{it})\exp(n(e^{it}-1-it))\,\mathrm{d}t \\ =& \big(1+\mathcal{O}(n^{-1})\big)(I_n+E_n), \end{aligned}$$

where

$$\begin{aligned} E_n =& \sqrt{\frac{n}{2\pi}}\int_{|t|\in[\theta,\pi]} \widetilde{G}(ne^{it})\exp\Big(n(e^{it}-1-it)\Big)\,\mathrm{d}t \\ =& \frac{n^n e^{-n}\sqrt{2\pi n}}{2\pi i}\int_{|t|\in[\theta,\pi]} \frac{\widetilde{G}(z)e^z}{z^{n+1}}\,\mathrm{d}t, \\ I_n =& \sqrt{\frac{n}{2\pi}}\int_{-\theta}^{\theta} \widetilde{G}(ne^{it})\exp\Big(n(e^{it}-1-it)\Big)\,\mathrm{d}t. \end{aligned}$$

By condition **(O)** we obtain that E_n decays exponentially to zero for $\alpha < 1$. Now, we turn to I_n. First we replace t by $t/\sqrt{n}$ and let $h_n(t) = \exp\Big(n(e^{it/\sqrt{n}}-1-it/\sqrt{n})\Big)$. Next, we split I_n into two parts, I'_n and I''_n (in order to find the Taylor expansion of $h_n(t)$) such that

$$\begin{aligned} I'_n =& \frac{1}{\sqrt{2\pi}}\int_{-\log n}^{\log n} \widetilde{G}\Big(ne^{it/\sqrt{n}}\Big)h_n(t)\,\mathrm{d}t, \\ I''_n =& \frac{1}{\sqrt{2\pi}}\int_{t\in[-\theta\sqrt{n},-\log n]} \widetilde{G}\Big(ne^{it/\sqrt{n}}\Big)h_n(t)\,\mathrm{d}t \\ &+ \frac{1}{\sqrt{2\pi}}\int_{t\in[\log n,\theta\sqrt{n}]} \widetilde{G}\Big(ne^{it/\sqrt{n}}\Big)h_n(t)\,\mathrm{d}t. \end{aligned}$$

Observe that $|h_n(t)| \le e^{-\mu t^2}$ for $t \in [-\theta\sqrt{n}, \theta\sqrt{n}]$, where μ is a constant. Then by condition **(I)** and Lemma 4 we obtain $I_n'' = \mathcal{O}\big(n^\beta e^{-\mu \log^2 n}\big)$. Now, we estimate I_n'. For $t \in [-\log n, \log n]$ we have the Taylor expansion of $h_n(t)$

$$h_n(t) = e^{-t^2/2}\Big(1 - \frac{it^3}{6\sqrt{n}} + \frac{t^4}{24n} + \mathcal{O}\Big(\frac{\log^5 n}{n\sqrt{n}}\Big)\Big).$$

Using condition **(I)** and Lemma 5 for $|z| > C\xi$ with constant C and $z \in \mathcal{L}_{\theta'}$ for $\theta' < \theta$, we have $|\widetilde{G}'(z)| \le C_1|z|^{\beta-1}$ and $|\widetilde{G}''(z)| \le C_2|z|^{\beta-2}$ for some constants C_1 and C_2. Thus we can expand $\widetilde{G}\Big(ne^{it/\sqrt{n}}\Big)$ around $t = 0$ as

$$\widetilde{G}\Big(ne^{it/\sqrt{n}}\Big) = \widetilde{G}(n) + it\sqrt{n}\widetilde{G}'(n) + \triangle_n(t)t^2,$$

where $|\triangle_n(t)| \le (C_1 + C_2)n^{\beta-1}$. Finally, the integral I_n' becomes

$$\begin{aligned} I_n' =& \frac{1}{\sqrt{2\pi}}\int_{-\log n}^{\log n} e^{-t^2}\Big(\widetilde{G}(n) + it\sqrt{n}\widetilde{G}'(n)\Big)\Big(1 - \frac{it^3}{6\sqrt{n}} + \frac{t^4}{24n}\Big)\,\mathrm{d}t \\ &+ \frac{1}{\sqrt{2\pi}}\int_{-\log n}^{\log n} e^{-t^2}\triangle_n(t)t^2 h_n(t)\,\mathrm{d}t \\ &+ \frac{1}{\sqrt{2\pi}}\int_{-\log n}^{\log n} e^{-t^2}\Big(\widetilde{G}(n) + it\sqrt{n}\widetilde{G}'(n)\Big)\mathcal{O}\Big(\frac{\log^5 n}{n\sqrt{n}}\Big)\,\mathrm{d}t. \end{aligned}$$

From Lemma 4 and Lemma 5 the first integral is equal to $\widetilde{G}(n) + \mathcal{O}(n^{\beta-1})$. The absolute value of second integral is smaller than $(C_1 + C_2)n^{\beta-1}$ by using the above estimate on $\triangle_n(t)$. Finally the third integral is $\mathcal{O}(n^{\beta-3/2}\log^5 n)$. Thus we have $I_n' = \widetilde{G}(n) + \mathcal{O}(n^{\beta-1})$ as desired. ∎

The next theorem extends the above one to a full asymptotic expansion of g_n:

Theorem 6. *Consider a linear cone $\mathcal{L}_\theta$ ($\theta < \pi/2$). Let the following two conditions hold for some numbers $A, B, R > 0$, and $\alpha > 0$, β, and γ:*

(I) *For $z \in \mathcal{L}_\theta$,*

$$|z| > R \Rightarrow |\widetilde{G}(z)| \le B|z|^\beta \Psi\big(|z|\big),$$

where $\Psi(x)$ is a slowly varying function.;

(O) *For all $z = \rho e^{i\theta}$ with $\theta \le \pi$ such that $z \notin \mathcal{L}_\theta$,*

$$\rho = |z| > R \Rightarrow |\widetilde{G}(z)e^z| \le A\rho^\gamma \exp\Big((1 - \alpha\theta^2)\rho\Big).$$

Then, for every non-negative integer m,

$$\begin{aligned} g_n =& \sum_{i=0}^{m}\sum_{j=0}^{i+m} b_{i,j} n^i \widetilde{G}^{\langle j\rangle}(n) + \mathcal{O}\Big(n^{\beta-(m+1)}\Psi(n)\Big) \\ =& \widetilde{G}(n) + \sum_{k=1}^{m}\sum_{i=1}^{k} b_{i,k+i} n^i \widetilde{G}^{\langle k+j\rangle}(n) + \mathcal{O}\Big(n^{\beta-(m+1)}\Psi(n)\Big), \end{aligned} \tag{1.21}$$

where $b_{i,j} = [x^i][y^j]\exp\big(x\log(1+y)-xy\big)$. Note that $b_{i,j}=0$ for $j<2i$.

Proof. The proof can be found in [11]. ∎

Remark. We present the expansion (1.21) above for $m=3$:

$$\begin{aligned} g_n =& \widetilde{G}(n) - \frac{1}{2}n\widetilde{G}^{\langle 2\rangle}(n) + \Big(\frac{1}{3}n\widetilde{G}^{\langle 3\rangle}(n) + \frac{1}{8}n^2\widetilde{G}^{\langle 4\rangle}(n)\Big) \\ & - \Big(\frac{1}{4}n\widetilde{G}^{\langle 4\rangle}(n) - \frac{1}{6}n^2\widetilde{G}^{\langle 5\rangle}(n) - \frac{1}{48}n^3\widetilde{G}^{\langle 6\rangle}(n)\Big) + \mathcal{O}(n^{\beta-4}\Psi(n)). \end{aligned}$$

Mean and variance. Let (X_n) be a sequence of integer random variables, and denote by $F_n(y)=\mathbb{E}[y^{X_n}]$ the probability generating function. Let

$$\widetilde{L}(z,y) = \sum_{n=0}^{\infty} F_n(y)\frac{z^n}{n!}e^{-z}$$

be the Poisson transform of the probability generating function. We introduce the Poisson mean $\widetilde{X}(z)$ and the Poisson variance $\widetilde{V}(z)$ as

$$\begin{aligned} \widetilde{X}(z) =& \widetilde{L}_y(z,1), \\ \widetilde{V}(z) =& \widetilde{L}_{yy}(z,1) + \widetilde{X}(z) - \widetilde{X}(z)^2, \end{aligned}$$

where $\widetilde{L}_y(z,1)$ and $\widetilde{L}_{yy}(z,1)$ denote respectively the first and the second derivative of $\widetilde{L}(z,u)$ with respect to y at $y=1$.

There is the following relationship between the Poisson mean $\widetilde{X}(z)$ and variance $\widetilde{V}(z)$ of X_n, and the Bernoulli mean $\mathbb{E}[X_n]$ and variance $\mathbb{V}[X_n]$.

Theorem 7. *Let $\widetilde{X}(z)$ and $\widetilde{V}(z)+\widetilde{X}(z)^2$ satisfy condition (O), and $\widetilde{X}(z)$ and $\widetilde{V}(z)$ satisfy condition (I) of Theorem 6 with $\beta\le 1$, e.g., $\widetilde{X}(z)=\mathcal{O}\big(|z|^\beta\Psi(|z|)\big)$, and $\widetilde{V}(z)=\mathcal{O}\big(|z|^\beta\Psi(|z|)\big)$ in a linear cone $\mathcal{L}_\theta$ and appropriate conditions (O) outside the cone, where $\Psi(z)$ is a slowly varying function. Then, the following holds*

$$\mathbb{E}[X_n] = \widetilde{X}(n) - \frac{n}{2}\widetilde{X}^{\langle 2\rangle}(n) + \mathcal{O}\Big(n^{\beta-2}\Psi(n)\Big), \tag{1.22}$$

$$\mathbb{V}[X_n] = \widetilde{V}(n) - n\widetilde{X}'(n)^2 + \mathcal{O}\Big(\max\Big(n^{\beta-1}\Psi(n); n^{2\beta-2}\Psi^2(n)\Big)\Big), \tag{1.23}$$

for large n.

Proof. From Theorem 6, we have directly (1.22) for $m = 1$. Since $\mathbb{V}[X_n] = \mathbb{E}[X_n^2] - \mathbb{E}[X_n]^2$, we observe that the Poisson transform of $\mathbb{E}[X_n^2]$ is $\widetilde{V}(z) + \widetilde{X}(z)^2$. Thus by Theorem 6 again

$$\begin{aligned}\mathbb{E}[X_n^2] =& \widetilde{V}(n) + \widetilde{X}(n)^2 - \frac{n}{2}\Big(\widetilde{V}^{\langle 2\rangle}(n) + 2n\widetilde{X}'(n)^2 + 2n\widetilde{X}(n)\widetilde{X}^{\langle 2\rangle}(n)\Big) + \mathcal{O}\big(n^{2\beta-2}\Psi^2(n)\big)\\ =& \widetilde{V}(n) + \widetilde{X}(n)^2 - n\widetilde{X}'(n)^2 - n\widetilde{X}(n)\widetilde{X}^{\langle 2\rangle}(n) + \mathcal{O}\big(n^{\beta-1}\Psi(n)\big) + \mathcal{O}\big(n^{2\beta-2}\Psi^2(n)\big),\end{aligned}$$

where the last error term is a consequence of $n\widetilde{V}^{\langle 2\rangle}(n) = \mathcal{O}(n^{\beta-1}\Psi(n))$ (see Lemma 5). Thus the result follows from $\mathbb{V}[X_n] = \mathbb{E}[X_n^2] - [\mathbb{E}X_n]^2$. ∎

1.4 Singularity Analysis

In this section, we restrict our attention to functions with a unique dominant singularity. By the scaling rule $g(z) = f(z\xi)$ if $f(z)$ has singular at $z = \xi$, we may always assume that the sole singularity occurs at $z = 1$, and we consider functions $f(z)$ of the form

$$f(z) = (1-z)^{-\alpha}\Big(\log\frac{1}{1-z}\Big)^{\gamma}, \tag{1.24}$$

with non-negative real numbers α and γ. Our general objective is to translate an approximation of a function near a singularity into an asymptotic approximation of its coefficients. More precisely, when all $h_0(z), \cdots, h_k(z), g(z)$ are as (1.24), then

$$f(z) = h_0(z) + h_1(z) + \cdots + h_k(z) + \mathcal{O}\big(g(z)\big) \tag{1.25}$$

with $h_0(z) \gg \cdots \gg h_k(z) \gg g(z)$ for $z \to 1$, will imply

$$[z^n]f(z) = h_{0,n} + h_{1,n} + \cdots + h_{k,n} + \mathcal{O}\big(g_n\big)$$

with $h_{0,n} \gg \cdots \gg h_{k,n} \gg g_n$ for $n \to \infty$. We omit all the proofs in this section since they can be found in [3].

From the binomial expansion, we have, with $\alpha \neq 0$,

$$[z^n](1-z)^{-\alpha} = \binom{n+\alpha-1}{n} = \frac{\Gamma(n+\alpha)}{\Gamma(\alpha)\Gamma(n+1)}.$$

Then from Stirling's formula $[z^n](1-z)^{-\alpha}$ has the asymptotic expansion, as $n \to \infty$,

$$[z^n](1-z)^{-\alpha} \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)}\Big(1 + \sum_{k\geq 1}\frac{e_k}{n^k}\Big), \tag{1.26}$$

where e_k is a polynomial in α of degree $2k$.

Table 1.4: Some commonly functions and the asymptotic forms of their coefficients.

$f(z)$	$[z^n]f(z)$
1	0
$\log(1-z)^{-1}$	$\frac{1}{n}$
$(1-z)^{-1}$	1
$(1-z)^{-1}\log\frac{1}{1-z}$	$\log n+\gamma+\frac{1}{2n}-\frac{1}{12n^2}+\frac{1}{120n^4}+\mathcal{O}(n^6)$
$(1-z)^{-1}\big(\log\frac{1}{1-z}\big)^2$	$\log^2 n+2\gamma\log n+\gamma^2-\frac{\pi^2}{6}+\mathcal{O}\Big(\frac{\log n}{n}\Big)$
$(1-z)^{-2}$	$n+1$

Remark. In particular:

$$[z^n](1-z)^{-\alpha}\sim\frac{n^{\alpha-1}}{\Gamma(\alpha)}\Big(1+\frac{\alpha(\alpha-1)}{2n}+\frac{\alpha(\alpha-1)(\alpha-2)(3\alpha-1)}{24n^2}+\frac{\alpha^2(\alpha-1)^2(\alpha-2)(\alpha-3)}{48n^3}+\mathcal{O}\Big(\frac{1}{n^4}\Big)\Big).$$

Next, we consider logarithmic factors, that is, $f(z)=(1-z)^{-\alpha}\big(\log(1-z)^{-1}\big)^{\gamma}$ with $\alpha\neq 0$. Similarly, we have the asymptotic expansion

$$[z^n]f(z)\sim\frac{n^{\alpha-1}}{\Gamma(\alpha)}(\log n)^{\gamma}\Big(1+\sum_{k\geq 1}\frac{C_k}{\log^k n}\Big),$$

where $C_k=\binom{\gamma}{k}\Gamma(\alpha)\frac{\mathrm{d}^k}{\mathrm{d}s^k}\frac{1}{\Gamma(-s)}\Big|_{s=\alpha}$.

Next, we want to establish our claim in (1.25). Therefore, we have to give conditions under which the following holds:

$$f(z)=\mathcal{O}(g(z))\quad\Rightarrow\quad[z^n]f(z)=\mathcal{O}([z^n]g(z)).$$

We first need a definition.

Definition 7. *Let $\triangle:=\triangle(\phi,\eta)$ denote the closed domain*

$$\triangle(\phi,\eta)=\{z\;\big|\;|z|<\eta,\ z\neq 1,\ |\arg(z-1)|\geq\phi\},$$

where $\eta>1$ and $0<\phi<\pi/2$.

Then, we have the following theorem:

Theorem 8. *Assume that $f(z)$ is analytic in $\triangle = \triangle(\phi, \eta)$, where $\eta > 1$ and $0 < \phi < \pi/2$, and that as $z \to 1$ in $\triangle$,*

$$f(z) = \mathcal{O}\Big((1-z)^{-\alpha}\Big(\log\frac{1}{1-z}\Big)^{\gamma}\Big),$$

for some non-negative integers α, γ with $\alpha \neq 0$. Then one has

$$[z^n]f(z) = \mathcal{O}\Big(n^{\alpha-1}(\log n)^{\gamma}\Big).$$

Finally, by the linearity

$$f(z) = f_1(z) + f_2(z) \quad \Rightarrow \quad [z^n]f(z) = [z^n]f_1(z) + [z^n]f_2(z).$$

We have the following theorem:

Theorem 9. *Assume that $f(z)$ is analytic in $\triangle = \triangle(\phi, \eta)$, where $\eta > 1$ and $0 < \phi < \pi/2$, and that as $z \to 1$ in $\triangle$,*

$$f(z) = (1-z)^{-\alpha}\Big(\log\frac{1}{1-z}\Big)^{\gamma}\Big(\sum_{j=0}^{m-1} c_j\Big(\log\frac{1}{1-z}\Big)^{-j} + \mathcal{O}\Big(\Big(\log\frac{1}{1-z}\Big)^{-m}\Big)\Big),$$

for non-negative real numbers α, γ with $\alpha \neq 0$ and $\gamma \geq m$. Then as $n \to \infty$,

$$[z^n]f(z) = \frac{n^{\alpha-1}}{\Gamma(\alpha)}\log^{\gamma} n\Big(\sum_{j=0}^{m-1} c'_j \log^{-j} n + \mathcal{O}(\log^{-m} n)\Big)$$

with some suitable constants c'_j.

Chapter 2

Results for Digital Search Trees

In this chapter, we first introduce digital search trees and their generalizations such as bucket digital search trees. Next, we present the results concerning the internal path length and explain how the results are proved. We also present results concerning other parameters of DSTs in Section 2.5.

2.1 Digital Search Trees

Digital trees are a general data structure to manipulate sequences which are built over a binary alphabet $\{0, 1\}$. There are three kinds of digital trees: "tries", "Patricia tries" and "digital search trees". In this thesis we only consider digital search trees and omit the others.

Suppose now we have an ordered set of records, say n of them, and each record has a key being an infinite sequence over $\{0, 1\}$. Then these records are stored in a digital search tree in the following way: Set k to 1. If $n = 1$, then the only record is put in a node and we are finished. If $n > 1$, then

- The first record is saved in a node (which becomes the root of the tree).
- According to the kth bit of the records in the remaining set:

  0: It goes to the left subtree where it is linked as a left child of the root.

  1: It goes to the right subtree where it becomes a right child of the root.

  We can split the remaining set into two subtrees.

- Finally, the subtrees are constructed by the same process recursively and set k to $k + 1$.

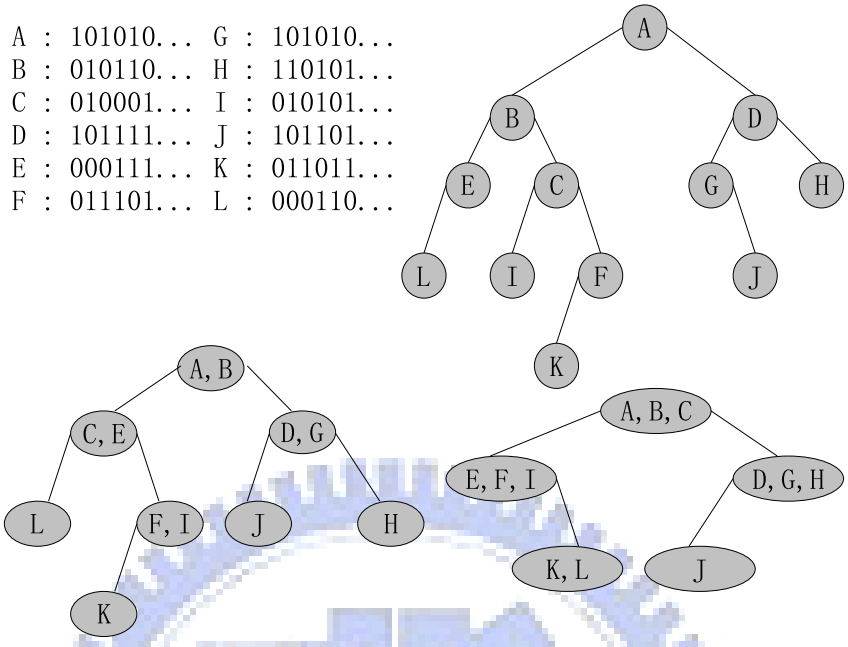


Figure 2.1: Examples of generalized digital search trees for $b = 1, 2, 3$ built from 12 records.

Thus we can see that digital search trees are build up of nodes, each node has a record containing a key and 2 links which point to subtrees. Obviously, the order in which the keys are inserted is relevant.

Next we equip the set of all digital search trees with a random model. Therefore we assume that each bit $\{0, 1\}$ is generated independently with probability p and $q = 1 - p$. For $p \neq q$ this leads to the asymmetric (biased) DST, where if $p = q = 1/2$, we obtain the symmetric (unbiased) DST.

Many generalizations of digital search trees have been considered. One of them are so called bucket digital search trees, where every node can hold up to b records.

The internal path length of a tree is the sum of the lengths of the paths to every node. More precisely, it is the sum of the number of edges on the path from the root to each node. In this work we denote by L_n the internal path length of a DST built from n (sufficiently long) records comprised of random digits.

Digital search trees have been quite thoroughly investigated in recent decades. Knuth [15] and Flajolet and Sedgewick [5] introduced analytical methods for the analysis of digital search trees. Their research was continued by Flajolet and Richmond [4], Louchard

[17], Szpankowski [22], Jacquet [10], Kirschenhofer and Prodinger [12] and others.

2.2 Internal Path Length for Symmetric DSTs

Now we are discussing the internal path length of a symmetric DST. Let $\pi(n,k)$ be the splitting probability which is the probability that the left subtree holds k records (and the right subtree holds $n-1-k$ records). Clearly $\pi(n+1,k) = \binom{n}{k}/2^k$. Under the condition of $\{\pi(n+1,k)\}$ we have the recurrence $L_{n+1} \stackrel{\mathrm{d}}{=} L_k + L_{n-k} + n$, which implies that the corresponding probability generating functions $F_n(z) = \mathbb{E}[z^{L_n}]$ satisfy for $n \geq 0$

$$F_{n+1}(z) = z^n 2^{-n} \sum_{k=0}^{n} \binom{n}{k} F_k(z) F_{n-k}(z), \qquad F_0(z) = 1. \tag{2.1}$$

Mean. Knuth [15] first used an approach suggested by Koheim and Newman [16] to derive the mean, but his approach is not useful for the analysis of other parameters. Flajolet and Sedgewick [5] gave another approach to analyze the mean which we will discuss here.

The expectation $f_n = \mathbb{E}[L_n]$ can be obtained from the probability generating functions (2.1) by $f_n = F_n'(1)$. Consequently,

$$f_{n+1} = n + 2^{1-n} \sum_{k=0}^{n} \binom{n}{k} f_k \quad (n \geqslant 0), \qquad f_0 = 0.$$

The above recurrence falls into the general type discussed in the following lemma:

Lemma 6. *Let (x_n) be a sequence of numbers satisfying $x_0 = x_1 = 0$,*

$$x_{n+1} = a_{n+1} + 2^{1-n} \sum_{k=0}^{n} \binom{n}{k} x_k \quad (n \geqslant 1),$$

where (a_n) is any sequence of numbers with $a_0 = a_1 = 0$;. We define the binomial inverse relations

$$\hat{a}_n = \sum_{k=0}^{n} (-1)^k \binom{n}{k} a_k \quad \text{and} \quad a_n = \sum_{k=0}^{n} (-1)^k \binom{n}{k} \hat{a}_k. \tag{2.2}$$

Then the solution is given by

$$x_n = -\sum_{k=2}^{n} (-1)^k \binom{n}{k} \hat{x}_{k-2},$$

where $Q_n = \prod_{1\le j\le n}(1-2^{-j})$ *and*

$$\hat{x}_n = Q_n \sum_{i=1}^{n+1} \frac{\hat{a}_i - \hat{a}_{i+1}}{Q_{i-1}}.$$

Proof. See [14]. ∎

Thus we obtain an explicit formula for f_n:

$$f_n = \sum_{j=2}^{n} (-1)^k \binom{n}{k} Q_{k-2}.$$

This is exactly Example 2 discussed in Section 1.1. Thus, we have the following theorem.

Theorem 10 (Flajolet and Sedgewick)**.** *The average internal path length of a symmetric digital search tree built from n records is*

$$\begin{aligned}\mathbb{E}[L_n] =& n\log_2 n + n\left(\frac{\gamma-1}{\log 2} + \frac{1}{2} - \alpha + \delta_1(\log_2 n)\right) + \log_2 n \\ &+ \frac{2\gamma-1}{2\log 2} + \frac{5}{2} - \alpha + \delta_2(\log_2 n) + \mathcal{O}(\log n/n),\end{aligned}$$

where $\gamma = 0.577216\cdots$ *is Euler's constant,* $\alpha = 1+\frac{1}{3}+\frac{1}{7}+\cdots = 1.606695\cdots$*, and* $\delta_1(x)$ *and* $\delta_2(x)$ *are continuous periodic functions of period 1, mean 0, and very small amplitude* $(< 10^{-6})$*. The approximate value of the coefficient of the linear term is* $-1.7155\cdots$*.*

Proof. Collecting all contributions as in Section 1.1. gives the expansion. The pole at $z=0$ yield a contribution of $\log_2 n + \frac{\gamma}{\log 2} + \frac{5}{2} - \alpha$, and the poles $z = \frac{2k\pi i}{\log 2}$ yield a periodic contribution of order n^0 and so on. ∎

Variance. By applying the same technique, Kirschenhofer, Prodinger and Szpankowski [14] derived the variance of the internal path length. More precisely, they used that the variance satisfies $\mathbb{V}[L_n] = s_n + f_n - f_n^2$ with $s_n = F_n''(1)$. From (2.1) we get the following recurrence for $n \ge 0$,

$$s_{n+1} = n2^{2-n}\sum_{k=0}^{n}\binom{n}{k} f_k + n(n-1) + 2^{1-n}\sum_{k=0}^{n}\binom{n}{k} f_k f_{n-k} + 2^{1-n}\sum_{k=0}^{n}\binom{n}{k} s_k$$

and $s_0 = 0$. We split it into three parts. Let $s_n = u_n + v_n + w_n$, where

$$u_{n+1} = 2n(f_{n+1} - n) + 2^{1-n}\sum_{k=0}^{n}\binom{n}{k}u_k \quad (n \geqslant 0), \qquad u_0 = 0, \tag{2.3a}$$

$$v_{n+1} = n(n-1) + 2^{1-n}\sum_{k=0}^{n}\binom{n}{k}v_k \quad (n \geqslant 0), \qquad v_0 = 0, \tag{2.3b}$$

$$w_{n+1} = 2^{1-n}\sum_{k=0}^{n}\binom{n}{k}f_k f_{n-k} + 2^{1-n}\sum_{k=0}^{n}\binom{n}{k}w_k \quad (n \geqslant 0), \qquad w_0 = 0, \tag{2.3c}$$

All of the above three recurrences are of the type as discussed in Lemma 6. Thus, the solutions of (2.3a)–(2.3c) follow from the binomial relations (2.2), where

$$\hat{u}_k = 2Q_{k-2}\Big(4 + \sum_{j=1}^{k-2}\frac{1}{2^j-1} - \sum_{j=1}^{k-2}\frac{j}{2^j-1} - \frac{2k}{2^{k-2}-1}\Big) \quad (k \geqslant 3), \qquad \hat{u}_0 = \hat{u}_1 = \hat{u}_2 = 0; \tag{2.4a}$$

$$\hat{v}_k = -4Q_{k-2} \quad (k \geqslant 3), \qquad \hat{v}_0 = \hat{v}_1 = \hat{v}_2 = 0; \tag{2.4b}$$

$$\hat{w}_k = -Q_{k-2}\sum_{j=4}^{k-1}\frac{2^{1-j}}{Q_{j-1}}\sum_{i=2}^{j-2}\binom{j}{i}Q_{i-2}Q_{j-i-2} \quad (k \geqslant 5), \qquad \hat{w}_0 = \cdots = \hat{w}_4 = 0. \tag{2.4c}$$

Next we focus on the asymptotics of u_n. In order to find an appropriate analytic continuation of $\hat{u}_k$, we can rewriting the sums appearing in (2.4a) as follows:

$$\sum_{j=1}^{k-2}\frac{1}{2^j-1} = \alpha - \sum_{j\geq 1}\frac{1}{2^{k-2+j}-1};$$
$$\sum_{j=1}^{k-2}\frac{j}{2^j-1} = \sum_{j\geq 1}\frac{j}{2^j-1} - \sum_{j\geq 1}\frac{k-2+j}{2^{k-2+j}-1},$$

where α is as defined in Theorem 10. Thus we may continue $\hat{u}_k$ via the function

$$\hat{u}(z) = \frac{2Q_\infty}{Q(2^{2-z})}\Big(4 + \alpha - \sum_{j\geq 1}\frac{1}{2^{z-2+j}-1} - \sum_{j\geq 1}\frac{j}{2^j-1} + \sum_{j\geq 1}\frac{z-2+j}{2^{z-2+j}-1} - \frac{2z}{2^{z-2}-1}\Big),$$

where $Q_\infty = 0.28878809$ and $Q(z) = \prod_{j\geq 1}(1-t/2^j)$. Now, we can apply the Rice method to obtain the asymptotics of u_n.

Next, the recurrence for v_n is easier. After simple algebra one proves

$$v_n = 4\binom{n}{2} - 4f_n,$$

and it is easy to get the asymptotics of v_n.

The appropriate extension of $\hat{w}_n$ is intricate. From (2.4c) we have

$$\hat{w}_{k+1} = -Q_{k-1}\sum_{j=4}^{k}\frac{\xi(j+1)}{2^{j-1}Q_{j-1}} \quad \text{with} \quad \xi(j+1) = \sum_{i=2}^{j-2}\binom{j}{i}Q_{i-2}Q_{j-2-i}.$$

Since $\xi(j+1) \sim 2^j Q_\infty^2$, let $\eta(j+1) = \xi(j+1) - 2^j Q_\infty^2$. Then

$$\begin{aligned}\hat{w}_{k+1} &= -Q_{k-1}\sum_{j=4}^{k}\frac{\eta(j+1)+2^jQ_\infty^2}{2^{j-1}Q_{j-1}} \\ &= Q_{k-1}\Big(-2Q_\infty(k-3) - \sum_{j\ge3}\frac{\eta(j+2)}{2^jQ_j} + \sum_{j\ge0}\frac{\eta(k+j+2)}{2^{k+j}Q_{k+j}} \\ &\qquad + 2Q_\infty^2\Big(\sum_{j\ge0}\Big(\frac{1}{Q_{k+j}} - \frac{1}{Q_\infty}\Big) - \sum_{j\ge3}\Big(\frac{1}{Q_j} - \frac{1}{Q_\infty}\Big)\Big)\Big).\end{aligned}$$

All series are absolutely convergent, we may sum them up term-by-term and get

$$\hat{w}_{k+1} = Q_{k-1}\Big(-2Q_\infty k + \frac{\xi(k+2)}{2^kQ_k} + \frac{\xi(k+3)}{2^{k+1}Q_{k+1}} + \sum_{j\ge2}\Big(\frac{\xi(k+j+2)}{2^{k+j}Q_{k+j}} - \frac{\xi(j+2)}{2^jQ_j}\Big)\Big).$$

From an appropriate interpretation for $\xi(z+1)$ (see [14])

$$\begin{aligned}\xi(z+1) = &\sum_{r\ge0}\frac{(-1)^r2^{-\binom{r+1}{2}}}{Q_r}\cdot\frac{Q_\infty}{Q(2^{3-z-r})}\cdot \\ &\Big(2^z - \frac{2}{1-2^{1-z-r}} - \frac{2z}{1-2^{2-z-r}} + 2\sum_{k\ge2}\binom{z}{k}\frac{1}{2^{r+k-1}-1}\Big),\end{aligned}$$

we immediately obtain the representation for $\hat{w}(z)$:

$$\hat{w}(z+1) = Q_{z-1}\Big(-2Q_\infty z + \frac{\xi(z+2)}{2^zQ_z} + \frac{\xi(z+3)}{2^{z+1}Q_{z+1}} + \sum_{j\ge2}\Big(\frac{\xi(z+j+2)}{2^{z+j}Q_{z+j}} - \frac{\xi(j+2)}{2^jQ_j}\Big)\Big)$$

with $Q_z = Q_\infty/Q(2^{-z})$, where $Q(z)$, Q_∞ are defined as above. Then, we again can obtain the asymptotics of w_n by Rice method.

Finally, from the relation $\mathbb{V}[L_n] = (u_n + v_n + w_n) + f_n - f_n^2$ we obtain the theorem:

Theorem 11 (Kirschenhofer, Prodinger and Szpankowski)**.** *The variance of the internal path length of symmetric digital search trees built from n records is*

$$\mathbb{V}[L_n] = n \cdot \Big(C + \delta(\log_2 n)\Big) + \mathcal{O}(\log^2 n/n),$$

where C is a constant with $C = 0.2660\ldots$ and all four digits after the decimal point are significant. The explicit form of C is

$$\begin{aligned} C = &-\frac{28}{3L} - \frac{39}{4} - 2\sum_{n\geq 1}\frac{n2^n}{(2^n-1)^2} + \frac{2\alpha}{L} + \frac{\pi^2}{2L^2} + \frac{2}{L^2} - \frac{2}{L}\sum_{k\geq 3}\frac{(-1)^{k+1}(k-5)}{(k+1)k(k-1)(2^k-1)} \\ &+ \frac{2}{L}\sum_{r\geq 1}(-1)^r 2^{-\binom{r+1}{2}}\left(\frac{L(1-2^{-r+1})/2-1}{1-2^{-r}} - \sum_{k\geq 2}\frac{(-1)^{k+1}}{k(k-1)(2^{r+k}-1)}\right) \\ &+ \frac{2}{L}\hat{w}'(3) - 2\delta_0 - \delta_1 \end{aligned} \tag{2.5}$$

with $L = \log 2$, the fluctuating function $\delta(x)$ is a continuous with period 1, mean zero, and $|\delta(x)| \leq 10^{-6}$, δ_0, δ_1 are two non-zero numbers with $|\delta_0| \leq 10^{-10}$ and $|\delta_1| \leq 10^{-10}$, and $\hat{w}(z)$ is defined above.

2.3 Internal Path Length for Asymmetric DSTs

From the last section, we know that Kirschenhofer et al. [14] obtained an asymptotic expression for the variance of the internal path length in the symmetric DST model. However, they did not extend their results to the asymmetric model. Jacquet and Szpankowski devised another approach to give the mean and variance of the internal path length of the asymmetric model in a DST [10]. We will introduce this method in this subsection.

Therefore, we suppose the binary digital search tree model is asymmetric with the probabilities p, q $(p+q=1)$. Similar as in the last section, we have $\mathbb{P}(\pi(n+1)=k) = \binom{n}{k}p^k q^{m-k}$ and the probability generating functions $F_n(y) = \mathbb{E}[y^{L_n}]$ of L_n satisfy for $n \geq 0$,

$$F_{n+1}(y) = z^n \sum_{k=0}^{n}\binom{n}{k}p^k q^{n-k}F_k(y)F_{n-k}(y), \qquad F_0(y) = 1.$$

Now, define $L(z,y) = \sum_{n\geq 0} F_n(y)z^n/n!$. Then one has

$$\frac{\partial}{\partial z}L(z,y) = L(pzy,y)L(qzy,y), \qquad L(z,0) = 1.$$

Finally, we consider the Poisson generating function $\widetilde{L}(z,y) = L(z,y)e^{-z}$ and obtain

$$\widetilde{L}(z,y) + \frac{\partial}{\partial z}\widetilde{L}(z,y) = e^{(y-1)z}\widetilde{L}(pzy,y)\widetilde{L}(qzy,y). \tag{2.6}$$

Next, denote by $\widetilde{X}(z) = \widetilde{L}_y(z,1)$ and $\widetilde{V}(z) = \widetilde{L}_{yy}(z,1) + \widetilde{X}(z) - \widetilde{X}(z)^2$ the Poisson mean and Poisson variance as already defined in Section1.3.

Poisson model. Consider first $\widetilde{X}(z)$. From (2.6) we obtain the following recurrence

$$\widetilde{X}(z) + \widetilde{X}'(z) = \widetilde{X}(pz) + \widetilde{X}(qz) + z, \qquad \widetilde{X}(0) = 0. \tag{2.7}$$

Let $X^*(s)$ denote the Mellin transform of $\widetilde{X}(z)$. Note that $\widetilde{X}(z) = \mathcal{O}(z^2)$ as $z \to 0$ and $\widetilde{X}(z) = \mathcal{O}(|z| \log |z|)$ as $z \to \infty$ in a linear cone (see the appendix in [10]). Thus the fundamental strip of $X^*(s)$ is $\langle -2, -1\rangle$ and the Mellin transform of $\widetilde{X}'(z) - z$ is also defined in the same strip. Then (2.7) translates into

$$X^*(s) - (s-1)X^*(s-1) = (p^{-s} + q^{-s})X^*(s) \tag{2.8}$$

in terms of the Mellin transform. Next, we set $X^*(s) = \xi(s)\Gamma(s)$ where $\Gamma(s)$ is the gamma function, and $\xi(s)$ satisfies the following recurrence:

$$\xi(s) - \xi(s-1) = (p^{-s} + q^{-s})\xi(s).$$

After some algebra one obtains

$$\xi(s) = \prod_{k=0}^{\infty} \frac{1 - p^{k+2} - q^{k+2}}{1 - p^{-s+k} - q^{-s+k}} = \frac{Q(-2)}{Q(s)}$$

for $s \in \langle -2, -1\rangle$, where $Q(s) = \prod_{k\geq 0}(1 - p^{-s+k} - q^{-s+k})$. We need a lemma to find the singularities:

Lemma 7. *Let s_k for $k \in \mathbb{Z}$ be solutions of*

$$p^{-s+r} + q^{-s+r} = 1,$$

where $p + q = 1$ and s is complex.

(i) For all $k \in \mathbb{Z}$

$$-1 + r \leq \Re(s_k) \leq \sigma_0 + r,$$

where σ_0 is a positive solution of $1 + q^{-s} = p^{-s}$. Furthermore,

$$\frac{(2k-1)\pi}{\log p} \leq \Im(s_k) \leq \frac{(2k+1)\pi}{\log p}.$$

(ii) If $\Re(s_k) = -1 + r$ *and* $\Im(s_k) \neq 0$, *then* $\log p/\log q$ *must be rational. More precisely, if* $\frac{\log p}{\log q} = \frac{w}{t}$, *where* $\gcd(w,t) = 1$ *for* $w,\, t \in \mathbb{Z}$, *then*

$$-1 + r + \frac{2mw\pi i}{\log p}, \quad m \in \mathbb{Z},$$

are all zeros with $\Re(s_k) = -1 + r$.

Inverting the Mellin transform then yields the the following asymptotic expansion of the Poisson mean:

$$\widetilde{X}(z) = \frac{z}{h}\Big(\log z + \gamma - 1 + \frac{h_2}{2h} - \alpha - \delta_1(\log z)\Big) + o(z) \qquad (z \to \infty), \tag{2.9}$$

where $h = -p\log p - q\log q$ is the entropy of the alphabet, $\gamma = 0.577\ldots$ is the Euler constant, $h_2 = -p\log^2 p - q\log^2 q$,

$$\alpha = -\sum_{k=1}^{\infty} \frac{p^{k+1}\log p + q^{k+1}\log q}{1 - p^{k+1} - q^{k+1}}, \tag{2.10}$$

and $\delta_1(\log z)$ is a fluctuating function for $\log p/\log q$ rational with small amplitude, and zero otherwise.

The variance is more intricate. Let $\widetilde{W}(z) = \widetilde{V}(z) - \widetilde{X}(z)$. From (2.6) we observe that $\widetilde{W}(z)$ satisfies the recurrence

$$\widetilde{W}(z) + \widetilde{W}'(z) = \widetilde{W}(pz) + \widetilde{W}(qz) + 2pz\widetilde{X}'(pz) + 2qz\widetilde{X}'(qz) + \widetilde{X}'(z)^2, \quad \widetilde{W}(0) = 0.$$

This functional equation is harder to solve due to the last term for which there is no closed-form expression for the Mellin transform, but it can be proved that the last term only contributes $\mathcal{O}(z)$. Let $\widetilde{W}(z) = \widetilde{W}_1(z) + \widetilde{W}_2(z)$ where

$$\begin{aligned}
\widetilde{W}_1(z) + \widetilde{W}_1'(z) &= \widetilde{W}_1(pz) + \widetilde{W}_1(qz) + 2pz\widetilde{X}'(pz) + 2qz\widetilde{X}'(qz), \quad \widetilde{W}_1(0) = 0,\\
\widetilde{W}_2(z) + \widetilde{W}_2'(z) &= \widetilde{W}_2(pz) + \widetilde{W}_2(qz) + \widetilde{X}'(z)^2, \quad \widetilde{W}_2(0) = 0.
\end{aligned}$$

Then, it was shown that in [10] that $\widetilde{W}_2(z)$ satisfies $\widetilde{W}_2(z) = \mathcal{O}(z)$ for z tends to infinity. Note that $\widetilde{W}_1(z) = \mathcal{O}(z^3)$ as $z \to 0$ and $\widetilde{W}_1(z) = \mathcal{O}(|z|\log|z|)$ as $z \to \infty$ in a linear cone. Hence the fundamental strip of $W_1^*(s)$ is $\langle -3, -1\rangle$ and the Mellin transform of $\widetilde{W}_1'(z)$ is defined in $\langle -2, 0\rangle$. For $s \in \langle -2, -1\rangle$, the Mellin transform $W_1^*(s)$ becomes

$$W_1^*(s) + g^*(s) = (p^{-s} + q^{-s})W_1^*(s) - 2(p^{-s} + q^{-s})sX^*(s),$$

where $g^*(s) = \mathcal{M}[\widetilde{W}_1'(z); s]$. Solving it, we obtain

$$W_1^*(s) = \frac{-g^*(s)}{1 - p^{-s} - q^{-s}} - \frac{2(p^{-s} + q^{-s})sX^*(s)}{1 - p^{-s} - q^{-s}}.$$

Since $g^*(s)$ is analytic on $\langle -2, 0\rangle$, $g^*(s)/(1-p^{-s}-q^{-s})$ only contributes terms up to $\mathcal{O}(z)$. Next, we can manipulate $\widetilde{W}_1(z)$ similar as the Poisson mean and get the asymptotic expansion of the Poisson variance

$$\widetilde{V}(z) = \frac{z\log^2 z}{h^2} + \frac{2z\log z}{h^3}\Big(\gamma h + h_2 - \frac{h^2}{2} - \alpha h - h\delta_1(\log z) - h\delta_1'(\log z)\Big) + \mathcal{O}(z). \tag{2.11}$$

Bernoulli model. From the two asymptotic expansions (2.9) and (2.11), we can observe that they satisfy the condition (I) of Theorem 7. To verify condition (O), we consider $Y(z) = \widetilde{X}(z)e^z$ and get

$$Y'(z) = Y(pz)e^{qz} + Y(qz)e^{pz} + ze^z, \quad Y(0) = 0.$$

Observe that the above equation can be represented as

$$Y(z) = \int_0^z \Big(Y(pw)e^{qw} + Y(qw)e^{pw} + we^w\Big)\,\mathrm{d}w.$$

We can apply mathematical induction over increasing domains and get a bound for $Y(z) = \widetilde{X}(z)e^z$ (see [11] for more details), as needed to verify condition (O) of Theorem 7. In a similar manner we can handle $\widetilde{V}(z) + \widetilde{X}(z)^2$. Thus we have the following theorem of the mean and the variance of the internal path length (see [10]):

Theorem 12 (Jacquet and Szpankowski)**.** *Consider a digital search tree built from n records under the asymmetric DST Bernoulli model. Then asymptotically the average value $\mathbb{E}[L_n]$ and the variance $\mathbb{V}[L_n]$ of the internal path length of the digital search tree become*

$$\begin{aligned}\mathbb{E}[L_n] =& \frac{n}{h}\left(\log n + \frac{h_2}{2h} + \gamma - 1 - \alpha + \delta_0(\log n)\right) + o(n),\\ \mathbb{V}[L_n] \sim& c_2 n\log n,\end{aligned} \tag{2.12}$$

where $h = -p\log p - q\log q$ is the entropy of the alphabet, $\gamma = 0.577\ldots$ is the Euler constant, $h_2 = p\log^2 p + q\log^2 q$, and $c_2 = (h_2 - h^2)/h^3$, α is defined in (2.10) and $\delta_0(\log n)$ is a fluctuating function for $\log p/\log q$ rational with small amplitude, and zero otherwise.

2.4 B-DSTs

Now we consider a b-DST, which is similar to the DST but now up to b records are stored in the nodes (the bucket capacity is b). The random model is as before. Flajolet and

Richmond [4] devised a method to give the average size of a digital search tree under the symmetric model. Hubalek [8] further developed the approach by Flajolet and Richmond to give the mean and variance of the internal path length of a symmetric b-DST.

From now on we fix the capacity b as an integer, and consider a b-DST built from n records ($n \geq 0$). Let L_n be the internal path length of a symmetric b-DST built from n records. Since we know that the first b records are stored in the root, thus the corresponding probability generating functions $F_n(z) = \mathbb{E}[z^{L_n}]$ satisfy for $n \geq 0$

$$F_{n+b}(z) = z^n \sum_{k=0}^{n} 2^{-n} \binom{n}{k} F_k(z) F_{n-k}(z), \qquad F_0(z) = \cdots = F_{b-1}(z) = 1.$$

Mean. As before, the expectation is $f_n = \mathbb{E}[L_n] = F_n'(1)$. Hence,

$$f_{n+b} = n + 2^{1-n} \sum_{k=0}^{n} \binom{n}{k} f_k, \qquad f_0 = f_1 = \cdots = f_{b-1}(z) = 0. \tag{2.13}$$

Again similar as before, we first investigate the general recurrence:

$$x_{n+b} = a_{n+b} + 2^{1-n} \sum_{k=0}^{n} \binom{n}{k} x_k, \quad x_0 = a_0,\, x_1 = a_1, \ldots, x_{b-1} = a_{b-1}.$$

One of the innovations in [4] is to consider the ordinary generating function. If we set the ordinary generating function $X(z) = \sum_{n\geq 0} x_n z^n$ and $A(z) = \sum_{n\geq 0} a_n z^n$ with respect to the sequences (x_n) and (a_n), we derive the following lemma.

Lemma 8. *The generating function $X(z)$ is given by $X(z) = \frac{1}{1-z}\widetilde{X}(\frac{z}{1-z})$, where $\widetilde{X}(z)$ satisfies*

$$(1+z)^b \widetilde{X}(z) = (1+z)^b \widetilde{A}(z) + 2z^b \widetilde{X}(\frac{z}{2}) \tag{2.14}$$

and $\widetilde{A}(z) = \frac{1}{1+z} A(\frac{z}{1+z})$.

Proof. Consider the Poisson transform $\widetilde{x}(z)$ and $\widetilde{a}(z)$ of the sequences (x_n) and (a_n), respectively. Then, we obtain for the coefficients $\widetilde{x}_n = n![z^n]\widetilde{x}(z)$ and $\widetilde{a}_n = n![z^n]\widetilde{a}(z)$

$$\sum_{j=0}^{b} \binom{b}{j} \widetilde{x}_{n+j} = \sum_{j=0}^{b} \binom{b}{j} \widetilde{a}_{n+j} + 2^{1-n} \widetilde{x}_n, \qquad \widetilde{x}_0 = \widetilde{x}_1 = \cdots = \widetilde{x}_{b-1} = 0. \tag{2.15}$$

From the equivalent relations (similar to the sequence (a_n) and $(\widetilde{a}_n)$)

$$x_n = \sum_{k=0}^{n} \binom{n}{k} \widetilde{x}_k \quad \Longleftrightarrow \quad \widetilde{x}_n = \sum_{k=0}^{n} \binom{n}{k} (-1)^{n-k} x_k,$$

we have

$$F(z) = \frac{1}{1-z}\widetilde{F}\Big(\frac{z}{1-z}\Big) \quad \text{and} \quad \widetilde{A}(z) = \frac{1}{1+z}A\Big(\frac{z}{1+z}\Big), \tag{2.16}$$

where $\widetilde{X}(z) = \sum_{n\geq 0}\widetilde{x}_n z^n$ and $\widetilde{A}(z) = \sum_{n\geq 0}\widetilde{a}_n z^n$. Finally, multiplying z^{n+b} to (2.15) and summing over n we obtain the relation

$$(1+z)^b\widetilde{X}(z) = (1+z)^b\widetilde{A}(z) + 2z^b\widetilde{X}(z/2). \quad \blacksquare$$

Remark 1. In Lemma 8, let $\hat{X}(t) = \widetilde{X}(t^{-1})$, $\hat{A}(t) = \widetilde{A}(t^{-1})$ and

$$\phi(t) = \prod_{j\geq 0}(1+2^{-j}t). \tag{2.17}$$

Then, by iterating:

$$\begin{aligned}\hat{X}(t) =& \hat{A}(t) + \frac{2}{(1+t)^b}\hat{X}(2t) = \sum_{j\geq 0}\frac{2^j\hat{A}(2^jt)}{(1+t)^b\cdots(1+2^{j-1}t)^b}\\ =& \phi\Big(\frac{t}{2}\Big)^b\sum_{j\geq 0}2^j\frac{(1+2^jt)^b\hat{A}(2^jt)}{\phi(2^jt)^b}.\end{aligned} \tag{2.18}$$

Thus, we obtain the harmonic sum $\Phi(t) = \sum_{j\geq 0}2^j\hat{P}(2^jt)/\phi(2^jt)^b$, where $\hat{P}(t) = (1+t)^b\hat{A}(t)$. Since $\phi(\frac{t}{2})^b = 1 + bt + \mathcal{O}(t^2)$ (the Taylor expansion at 0), it suffices to know the asymptotic behavior of $\Phi(t)$ whose Mellin transform is given by

$$\Phi^*(s) = \frac{1}{1-2^{1-s}}\cdot\Big(\frac{\hat{P}(t)}{\phi(t)^b}\Big)^*(s). \tag{2.19}$$

Now, we will turn to the mean. From (2.13) and Lemma 8:

$$(1+z)^b\widetilde{F}(z) = z^{b+1} + 2z^b\widetilde{F}(z/2).$$

Using Remark 1 one has

$$\begin{aligned}\hat{F}(t) =& \phi\Big(\frac{t}{2}\Big)^b\Big(\frac{1}{t}\sum_{j\geq 0}\frac{2^j}{2^j\phi(2^jt)^b}\Big)\\ =& \phi\Big(\frac{t}{2}\Big)^b H(t).\end{aligned} \tag{2.20}$$

From the integral relation $\int_0^\infty \log(1+z)z^{s-1}\,\mathrm{d}z = \frac{\pi}{s\sin\pi s}$ for $\Re(s)\in\langle -1,0\rangle$, we have

$$\begin{aligned}\log\phi(t) =& \sum_{j\geq 0}\log(1+2^{-j}t)\\ =& \frac{1}{2\pi i}\int_{1/2-i\infty}^{1/2+i\infty}\frac{\pi}{(1-2^s)s\sin\pi s}t^{-s}\,\mathrm{d}s\\ \sim& \frac{\log^2 t}{2\log 2}+\frac{\log t}{2},\end{aligned} \tag{2.21}$$

uniformly for $|t|\to\infty$ in the linear cone $\mathcal{L}_\theta$ for any fixed $\theta\in(0,\pi)$. Thus,

$$\phi(t)^{-b} = \begin{cases} 1-2bt+\mathcal{O}(t^2), & t\to 0,\\ \mathcal{O}\big(\exp(-(b/2\log 2)\log^2 t)\big), & t\to\infty,\end{cases} \tag{2.22}$$

in the cone. This guarantees the existence of the Mellin transform of $H(t)$ which is

$$H^*(s) = \frac{1}{1-2^{1-s}}I^*(s-1)\qquad(\Re(s)>1), \tag{2.23}$$

where

$$I^*(s) = \int_0^\infty \phi(t)^{-b}t^{s-1}\,\mathrm{d}t \tag{2.24}$$

converges in the strip $\langle 0,\infty\rangle$.

Remark. $I^*(s)$ is exponentially small as $\Im(s)\to\pm\infty$ for $\Re(s)>0$ [4]. Moreover, one can prove

$$I^*(s) = \frac{\pi}{\sin\pi s}J(s),\quad \text{with } J(s) = \frac{1}{2\pi i}\int_{\mathcal{H}}\frac{1}{\phi(t)^b}(-t)^{s-1}\,\mathrm{d}t, \tag{2.25}$$

where $\mathcal{H}$ is a Hankel-type contour starting at $+\infty-0\cdot i$, turning around 0 clockwise before returning to $+\infty+0\cdot i$. Flajolet and Richmond [4] also give the representation

$$J(s) = A_0(2^s)+(s-1)A_1(2^s)+\cdots+(s-1)(s-2)\cdots(s-b+1)A_{b-1}(2^s), \tag{2.26}$$

where $A_k(x)$'s are entire functions, thus $J(k)=0$, for all $k\geq 1$. Furthermore, (2.22) implies that $I^*(s)\sim s^{-1}$ as $s\to 0$ and $I^*(s)\sim -2b(s+1)^{-1}$ as $s\to -1$. Thus we can obtain the singular expansion of $I^*(s)$.

From the above remark and (2.23), we know that $H(s)$ has a double pole at $s=1$ and simple poles at $s=1+\chi_k$, where $\chi_k = 2k\pi i/L$ $(k\in\mathbb{Z})$ with $L=\log 2$. Applying the inversion formula

$$H(t) = \frac{1}{2\pi i}\int_{3/2-i\infty}^{3/2+i\infty}H^*(s)t^{-s}\,\mathrm{d}s,$$

we have the asymptotic expansion of $H(t)$ as $t \to 0$ (the remainder term is due to a simple pole at $s = -1$) [8]

$$H(t) = -\frac{1}{L} t^{-1} \log t + \Big(\frac{1}{L} J'(0) + \frac{1}{2}\Big) t^{-1} + \frac{1}{L} \sum_{k \neq 0} I^*(\chi_k) t^{-1-\chi_k} + 2b + \mathcal{O}(t), \tag{2.27}$$

where

$$J'(0) = \int_0^1 \Big(\frac{1}{\phi(t)^b} - 1\Big) t^{-1} \, \mathrm{d}t + \int_1^\infty \frac{t^{-1}}{\phi(t)^b} \, \mathrm{d}t. \tag{2.28}$$

Remark. First we rewrite

$$\begin{aligned} J'(0) &= \int_0^1 \Big(\frac{1}{\phi(t)^b} - 1\Big) t^{-1} \, \mathrm{d}t + \int_1^\infty \frac{t^{-1}}{\phi(t)^b} \, \mathrm{d}t \\ &= -b \int_0^\infty \phi(t)^{-b} \frac{\phi'(t)}{\phi(t)} \log t \, \mathrm{d}t, \end{aligned}$$

then

$$J'(0) \sim -2b \int_0^\infty e^{-2bt} \log t \, \mathrm{d}t = 2b \frac{\mathrm{d}}{\mathrm{d}s} (2b)^{-s} \Gamma(s)|_{s=1} = -\log b - \gamma - L$$

as $b \to \infty$.

Equations (2.20) and (2.27) give

$$\begin{aligned} \hat{F}(t) = &-\frac{1}{L} t^{-1} \log t + \Big(\frac{1}{L} J'(0) + \frac{1}{2}\Big) t^{-1} + \frac{1}{L} \sum_{k \neq 0} I^*(\chi_k) t^{-1-\chi_k} \\ &-\frac{b}{L} \log t + \Big(\frac{b}{L} J'(0) + \frac{5b}{2}\Big) + \frac{b}{L} \sum_{k \neq 0} I^*(\chi_k) t^{-\chi_k} + \mathcal{O}(t \log t), \end{aligned} \tag{2.29}$$

and by the elementary substitution (2.16) we obtain the asymptotics of $F(z)$. Finally, using Theorem 9 we obtain the following theorem for the mean of symmetric b-DSTs.

Theorem 13 (Hubalek)**.** *The expected generalized internal path length of a b-digital search tree built from n records satisfies as $n \to \infty$*

$$\begin{aligned} \mathbb{E}[L_n] = &n \log_2 n + \Big(\frac{1}{L} J'(0) + \frac{1}{2} + \frac{\gamma}{L} - \frac{1}{L} + \delta_1(\log_2 n)\Big) n + b \log_2 n \\ &+ \Big(\frac{b}{L} J'(0) + \frac{5b}{2} + \frac{b\gamma}{L} - \frac{1}{2L} + \delta_2(\log_2 n)\Big) + \mathcal{O}\Big(\frac{\log n}{n}\Big), \end{aligned}$$

where $L = \log 2$, γ denotes Euler's constant, $J'(0)$ is defined in (2.28), $\delta_1(x)$ and $\delta_2(x)$ are analytic, periodic functions with mean 0 and period 1.

Variance. To compute the variance, we use the formula $\mathbb{V}[L_n] = s_n - f_n^2 + f_n$ where $s_n = F_n''(1)$ as for the classical symmetric DST. Then,

$$s_{n+b} = n2^{2-n}\sum_{k=0}^{n}\binom{n}{k}f_k + n(n-1) + 2^{1-n}\sum_{k=0}^{n}\binom{n}{k}f_k f_{n-k} + 2^{1-n}\sum_{k=0}^{n}\binom{n}{k}s_k$$

with $s_0 = s_1 = \cdots = s_{b-1} = 0$. We again split the above recurrence into three components, $s_n = u_n + v_n + w_n$, where

$$u_{n+b} = n2^{2-n}\sum_{k=0}^{n}\binom{n}{k}f_k + 2^{1-n}\sum_{k=0}^{n}\binom{n}{k}u_k, \quad u_0 = \cdots = u_{b-1} = 0; \tag{2.30a}$$

$$v_{n+b} = n(n-1) + 2^{1-n}\sum_{k=0}^{n}\binom{n}{k}v_k, \quad v_0 = \cdots = v_{b-1} = 0; \tag{2.30b}$$

$$w_{n+b} = 2^{1-n}\sum_{k=0}^{n}\binom{n}{k}f_k f_{n-k} + 2^{1-n}\sum_{k=0}^{n}\binom{n}{k}w_k, \quad w_0 = \cdots = w_{b-1} = 0. \tag{2.30c}$$

Applying Lemma 8 to (2.30a)–(2.30c) yields

$$(1+z)^b\widetilde{U}(z) = 4z^{b+1}\widetilde{F}\Big(\frac{z}{2}\Big) + 2z^{b+1}\widetilde{F}'\Big(\frac{z}{2}\Big) + 2z^{b+2}\widetilde{F}'\Big(\frac{z}{2}\Big) + 2z^b\widetilde{U}\Big(\frac{z}{2}\Big); \tag{2.31a}$$

$$(1+z)^b\widetilde{V}(z) = 2z^{b+2} + 2z^b\widetilde{V}\Big(\frac{z}{2}\Big); \tag{2.31b}$$

$$(1+z)^b\widetilde{W}(z) = 2z^b\widetilde{M}(z) + 2z^b\widetilde{W}\Big(\frac{z}{2}\Big); \tag{2.31c}$$

where

$$\widetilde{m}_n = [z^n]\widetilde{M}(z) = 2^{-n}\sum_{k=0}^{n}\binom{n}{k}\widetilde{f}_k\widetilde{f}_{n-k} \quad (n \geqslant 0). \tag{2.31d}$$

Now we again apply Remark 1 to (2.31a) to obtain the expression for $\hat{U}(t)$ with

$$\hat{P}(t) = 4t^{-1}\hat{F}(2t) - 8\hat{F}'(2t) - 8t\hat{F}'(2t)$$

as (2.18). Next, let $\Upsilon(t) = \hat{P}(t)/\phi(t)^b$. From the derivative of $\hat{F}(2t) = \phi(t)^b H(2t)$, we get

$$\begin{aligned}\Upsilon(t) =& 4t^{-1}H(2t) - 4b\big(T(t)-2\big)H(2t) - 8bH(2t)\\ &- 8H'(2t) - 4btT(t)H(2t) - 8tH'(2t),\end{aligned}$$

where $T(x) = \phi'(x)/\phi(x) = \sum_{j\geq 0} 2^{-j}/(1+2^j x)$ and $\Phi_U^*(s) = \Upsilon^*(s)/(1-2^{1-s})$. Since,

$$T(x) = \begin{cases} 2 + \mathcal{O}(x), & x \to 0,\\ \mathcal{O}\big(x^{-1}\big), & x \to \infty,\end{cases}$$

then $T(x)$ is a harmonic sum with Mellin transform

$$T^*(s) = \frac{1}{1-2^{s-1}} \frac{\pi}{\sin \pi s} \quad (s \in \langle 0, 1\rangle),$$

and $\mathcal{M}[T(x) - 2; s] = T^*(s)$ for $s \in \langle -1, 0\rangle$. The Mellin transform of $\Upsilon(t)$ is

$$\Upsilon^*(s) = s2^{3-s}H^*(s-1) - 4b\Upsilon_0^*(s) - b2^{3-s}H^*(s) - 4b\Upsilon_1^*(s) + s2^{2-s}H^*(s)$$

for $s \in \langle 2, \infty\rangle$, where $\Upsilon_0^*(s) = \mathcal{M}[(T(t)-2)H(2t); s]$ and $\Upsilon_1^*(s) = \mathcal{M}[T(t)H(2t); s]$ exist for $s \in \langle 0, \infty\rangle$. For asymptotic analysis of $\Phi^*(s)$, we have to take $\Upsilon_0^*(1)$ and $\Upsilon_1^*(1)$ into account. One of the innovations in [8] is the use of the Mellin convolution formula.

Remark. The Mellin's convolution formula is

$$\mathcal{M}[F(t) \cdot G(t); s] = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} F^*(\tau) \cdot G^*(s-\tau)\, \mathrm{d}\tau, \tag{2.32}$$

valid for c and $s-c$ in the fundamental strip of F^* and G^*, respectively.

From (2.32), we obtain for $j = 0, 1$ respectively,

$$\Upsilon_j^*(s) = \frac{1}{2\pi i} \int_{1/2-i\infty}^{1/2+i\infty} T^*(\tau + j) \cdot 2^{-(s-\tau)} H^*(s-\tau)\, \mathrm{d}\tau.$$

First we compute $\Upsilon_0^*(1)$ by splitting

$$\begin{aligned} T^*(\tau)2^{-(1-\tau)}H^*(1-\tau) =& \frac{\pi}{\sin \pi\tau} \frac{2^{\tau-1}}{(1-2^{\tau-1})(1-2^{\tau})} I^*(0-\tau) \\ =& -T^*(\tau+1)I^*(0-\tau) - T^*(\tau)I^*(0-\tau). \end{aligned}$$

Then the first part is

$$\begin{aligned} \frac{1}{2\pi i} \int_{1/2-i\infty}^{1/2+i\infty} T^*(\tau+1)I^*(0-\tau)\, \mathrm{d}\tau =& \mathcal{M}[tT(t)I(t); s=0] \\ =& -\frac{1}{b}\mathcal{M}[I'(t); s=0] = \frac{1}{b}, \end{aligned}$$

and the second part yields

$$\begin{aligned} \frac{1}{2\pi i} \int_{1/2-i\infty}^{1/2+i\infty} T^*(\tau)I^*(0-\tau)\, \mathrm{d}\tau =& \mathcal{M}[(T(t)-2)I(t); s=0] \\ =& \lim_{s\to 0}\Big\{ -\frac{1}{b}\mathcal{M}[T(t)I(t); s] - 2I^*(s)\Big\} \\ =& \frac{1}{b}J'(-1) - 2J'(0) - 2. \end{aligned}$$

Thus $\Upsilon_0^*(1) = -\frac{1}{b} - \frac{1}{b}J'(-1) + 2J'(0) + 2$. It is more difficult to compute $\Upsilon_1^*(1)$, for which it can be proved that

$$\Upsilon_1^*(1) = -\frac{1}{4b}\int_0^\infty t^{-1}\frac{\Lambda(t)}{\phi(t)^b}\,\mathrm{d}t \sim -2 \ (b \to \infty)$$

with $\Lambda(t) = 2\sum_{j\geq 0} j2^{-j}t/(1+2^{-j}t)$. Thus, we can manipulate the expansion for $U(z)$ as $z \to 1$ similar as for $F(z)$ and get the asymptotics of u_n as $n \to \infty$.

The asymptotic of v_n is simple. Again applying Remark 1 to (2.31b), we obtain $\hat{V}(t)$ with $\hat{P}(t) = 2t^{-2}$ and $\Phi_V^*(s) = 2I^*(s-2)/(1-2^{1-s})$. We immediately get the asymptotics of v_n as $n \to \infty$ from the properties of $I^*(s)$.

Because of the appearance of the "binomial convolution" (2.31d), it is non-trivial to apply the same method to (2.31c). But, since the exponential generating function $\widetilde{m}(z) = \sum_{N\geq 0}\widetilde{m}_N z^N/N!$ satisfies $\widetilde{m}(z) = \widetilde{f}(z/2)^2$, it can be proved that

$$\hat{M}^*(s) = 2^{-s}\cdot\frac{1}{2\pi i}\int_{3/2-i\infty}^{3/2+i\infty}\binom{s}{\tau}\hat{F}^*(\tau)\hat{F}^*(s-\tau)\,\mathrm{d}\tau, \tag{2.33}$$

where $\binom{s}{\tau} = \Gamma(1+s)/\Gamma(1+\tau)\Gamma(1+s-\tau)$ is the complex binomial coefficient. Next, from the singular expansions of $\hat{F}$ and the Taylor series of complex binomial coefficients, we obtain the asymptotics of $\hat{M}^*(s)$ as $s \to 2$. Similarly, one treats the case $s \to 1$.

From (2.31c) we have $\hat{W}(t) = \phi(t/2)^b\Phi_W(t)$, where $\Phi_W(t) = 2\sum_{j\geq 0} 2^j\hat{P}(2^j t)$ with $\hat{P}(t) = \hat{M}(t)I(t)$. Presupposing some properties of $\hat{P}$, then $\Phi^*(s) = 2\hat{P}^*(s)/(1-2^{1-s})$ where

$$\hat{P}^*(s) = \frac{1}{2\pi i}\int_{1/2-i\infty}^{1/2+i\infty} I^*(\tau)\hat{M}^*(s-\tau)\,\mathrm{d}\tau,$$

for $s \in \langle\frac{5}{2}, 2b+\frac{5}{2}\rangle$. Now shifting the contour to the left yields the analytic continuation

$$\hat{P}^*(s) = \hat{M}^*(s) - 2b\hat{M}^*(s+1) + \frac{1}{2\pi i}\int_{-3/2-i\infty}^{-3/2+i\infty} I^*(\tau)\hat{M}^*(s-\tau)\,\mathrm{d}\tau.$$

in $s \in \langle\frac{1}{2}, 2b+\frac{1}{2}\rangle$. Thus we get the Laurent series of $\hat{P}^*(s)$ as $s \to 2$ and $s \to 1$. After hard calculating, we obtain the asymptotics of w_n. Overall, the following theorem for the variance of the internal path length over a b-DST holds:

Theorem 14 (Hubalek)**.** *The variance of the generalized internal path length of a b-digital search tree built from n records satisfies as* $n \to \infty$*,*

$$\mathbb{V}[L_n] = \Big(C + \delta(\log n)\Big)n + \mathcal{O}(\log^2 n), \tag{2.34}$$

where

$$\begin{aligned}C =&\frac{2b}{L^2}J'(0)+\frac{1}{L}[\delta_1\varphi_0]_0+\frac{23b}{6}+\frac{1}{2L}\bar{F}_0^*(0)-\frac{3}{2L}J'(-1)-\frac{\pi^2 b}{3L^2}-\frac{9b}{L}J'(0)\\&+\frac{3b}{2L}\bar{F}_1^*(1)+\frac{1}{12}+\frac{6b}{L}J'(-1)+2b[\delta_1^2]_0+\frac{3}{L}-\frac{b}{L^2}J''(0)+\frac{1}{L}J''(0)^2\\&-\frac{1}{L^2}J'(0)^2+\frac{\pi^2}{6L^2}-\frac{2b}{L}[\delta_1\varphi_2]_0+\frac{1}{L}K'(0)-\frac{2b}{L^2}-[\delta_1^2]_0-\frac{b}{L^2}J'(0)\cdot\bar{F}_1^*(1)\\&+\frac{1}{L^2}J'(0)\cdot J'(-1)+\frac{b}{L^2}\bar{F}_1^*(1)+\frac{1}{L}[\delta_1\varphi_1]_0+\frac{b}{L^2}\bar{F}_1^{*\prime}(1)+\frac{1}{L^2}J'(0)\cdot\bar{F}_0^*(0)\\&-\frac{4b}{L}\bar{M}_2^*(2)-\frac{1}{L^2}\bar{F}_0^*(0)-\frac{1}{L^2}J'(-1)-\frac{1}{2L^2}J''(-1)-\frac{1}{L^2}\bar{F}_0^{*\prime}(0)\\&+\frac{2}{L}\Phi_1^*(1)-\frac{3b}{L}-2[\delta_1\delta_2]_0.\end{aligned}\tag{2.35}$$

The constants and functions are defined in [8].

Remark. Hubalek gives the following values of C for $b=1,\ldots,5$.

b	1	2	3	4	5
C	0.26600	0.13285	0.08883	0.07032	0.06109

Later on we will see that most of the digits are incorrect.

2.5 Other Parameters

Now we are going to introduce results on other parameters which have been studied for digital search trees of size n. Our main emphasis will again be on mean and variance. First let us fix some notation: set $L=\log 2$, α is given in Theorem 10, $\beta=\sum_{k\geq 1}(2^k-1)^{-2}=0.788343\cdots$, γ is the Euler number, the constants h, h_2 and c_2 are given in Theorem 12, and $\hat{h}_2=p^2\log p+q^2\log q$.

Depth. The depth of a node is the number of nodes on the path from root to the selected node. Let D_n be the depth of a randomly selected node in a digital search tree. Knuth [15] first gave an approach to the mean value of the symmetric DSTs which later was improved by Flajolet and Sedgewick [5]. Kirschenhofer and Prodinger used Flajolet and Sedgewick's approach to give the variance in the symmetric DSTs [12]. Szpankowski used a method which is also similar to Flajolet and Sedgewick to give all moments in the asymmetric DSTs [22].

Theorem 15. *1. The asymptotics of the mean $\mathbb{E}[D_n]$ and the variance $\mathbb{V}[D_n]$ of the depth of the symmetric digital search tree built over n records are*

$$\begin{aligned}\mathbb{E}[D_n] =& \log_2 n + \frac{\gamma - 1}{L} + \frac{1}{2} - \alpha + \sigma_1(n) + \mathcal{O}\big(n^{-1/2}\big),\\ \mathbb{V}[D_n] =& \frac{1}{12} + \frac{\pi^2}{6L^2} + \frac{1}{L^2} - \alpha - \beta + \sigma_2(n),\end{aligned}$$

where $\sigma_1(x)$ and $\sigma_2(x)$ are small fluctuating functions in [12].

2. Under the asymmetric Bernoulli model, the mean and the variance become

$$\begin{aligned}\mathbb{E}[D_n] =& \frac{1}{h}\Big\{\log n + \gamma - 1 + \frac{h_2}{2h} - \theta + \sigma_3(n)\Big\} + \mathcal{O}\big(n^{-1/2}\big),\\ \mathbb{V}[D_n] =& c_2 \log n + C + \sigma_4(z) + \mathcal{O}(n^{-1}\log^2 n),\end{aligned}$$

where C is a constant, $\sigma_3(x)$ and $\sigma_4(x)$ are small fluctuating functions in [22]

As for limit results, Louchard used a probabilistic technique to give the asymptotic distributions of the depth in a symmetric DST [17]. Moreover, Louchard and Szpankowski proved the normal limiting distribution of the depth in the asymmetric DSTs [19]. Finally, in the generalized b-DSTs, Louchard, Szpankowski and Tang derived the mean, the variance, and the limiting distribution for the symmetric and asymmetric b-DSTs [20].

Distance. The distance between two nodes is the number of nodes on the path connecting the selected two nodes. Let d_n be the distance between two randomly selected nodes in a digital search tree. Aguech, Lasmar and Mahmoud used the methods developed for the depth to determine the moments and to obtain the limit law of the distance in a DST [1]. We only give their results concerning mean and variance.

Theorem 16. *1. Consider an asymmetric digital search tree built from n records. Then asymptotically the average value $\mathbb{E}[d_n]$ and the variance $\mathbb{V}[d_n]$ of the distance between two random nodes in the digital search tree become*

$$\begin{aligned}\mathbb{E}[d_n] =& \frac{2}{h}\log n + \frac{1}{h}\Big(\frac{\hat{h}_2}{pq} + \frac{h_2}{h} - 2(1-\gamma) + \log(pq) - 2L\alpha\Big)\\ &+ 2 - 2\delta_q(n) + \mathcal{O}\big(n^{-0.49999}\big),\\ \mathbb{V}[d_n] =& 2c_2 \log n + \mathcal{O}(1),\end{aligned}$$

where $\delta_q(n)$ is a small fluctuating function in [1].

2. Now, consider the digital search tree under the symmetric Bernoulli model. Then

$$\begin{aligned}\mathbb{E}[d_n] =& 2\log_2 n - 1 + \frac{2(\gamma-1)}{L} - \alpha - 2\delta_{\frac{1}{2}}(n) + \mathcal{O}\big(n^{-0.49999}\big),\\ \mathbb{V}[d_n] =& \frac{6+\pi^2}{3L^3} + \frac{22}{3} - 2(\alpha+\beta) + \frac{4(\gamma-1)}{L}\delta_{\frac{1}{2}}(n) - 2\delta^2_{\frac{1}{2}}(n)\\ &+ \frac{4}{L}\hat{\delta}(n) + \mathcal{O}\big(n^{-0.49999}\big),\end{aligned}$$

where $\hat{\delta}(x)$ is another small fluctuating function in [1].

External-internal nodes. A node with both links null is called an external-internal node. Knuth gave the open question in [15] to analyze the number of such nodes in random DSTs (Prodinger showed that Knuth could have solved it himself in [21]). This question was solved by Flajolet and Sedgewick who gave the mean value in a symmetric digital search tree. Moreover, the variance in the symmetric DSTs was derived by Kirschenhofer and Prodinger [13]. Since the latter result is very messy we just give the result for the mean.

Theorem 17. *The average number of external-internal nodes in a symmetric digital search tree built from n records is*

$$n\Big(\beta + 1 - \frac{1}{Q_\infty}\big(\frac{1}{L} + \alpha^2 - \alpha\big) + \delta(n)\Big) + \mathcal{O}(n^{1/2}),$$

where $Q_\infty = \prod_{k\geq 1}(1-2^{-j}) = 0.288788\ldots$,

$$\beta = \sum_{k=1}^{\infty} \frac{k\cdot 2^{k(k-1)/2}}{1\cdot 3\cdot 5\cdots(2^k-1)}\cdot\Big(\sum_{j=1}^{k}\frac{1}{2^j-1}\Big) = 7.74313\cdots,$$

and $\delta(x)$ is a small fluctuating function in [5]

As for b-DSTs, mean, variance and limit laws in the symmetric b-DSTs were derived in Hubalek, Hwang, et al. [9].

The Size. The size of a tree is the number of nonempty nodes. For a classical DST, the size is equal to the number of nodes, but this does not hold for the b-DSTs. Flajolet and Richmond gave the expected value of the size of a symmetric b-DST [4]. Moreover, variance and limit distributions were derived in Hubalek, Hwang, et al. [9]. Again we just give the result for the mean.

Theorem 18. *The expected number of nonempty nodes in a symmetric b-DST built from n records satisfies*

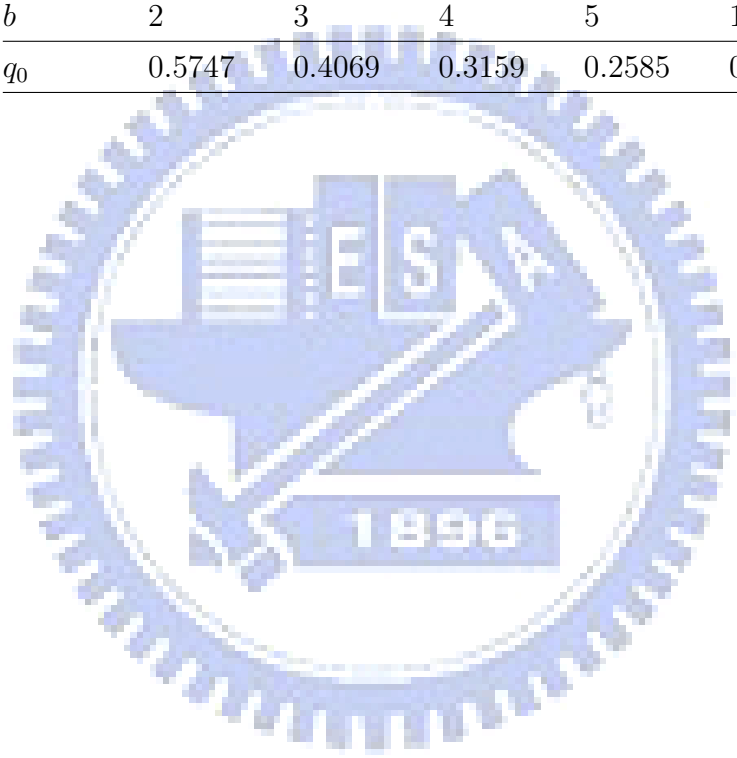
$$n(q_0 + S(n)) + \mathcal{O}(n^{1/2})$$

where

$$q_0 = \frac{1}{L}\int_0^\infty \frac{(1+t)^{b-1}}{\phi(t)^b}\,\mathrm{d}t,$$

where $\phi(t)$ is defined in (2.17), $S(x)$ is a periodic function with mean 0 and the following are few values of the leading constant q_0:

b	2	3	4	5	10
q_0	0.5747	0.4069	0.3159	0.2585	0.1360



Chapter 3

New Method for Internal Path Length

In this chapter, we explain a new method which will appear in a forthcoming paper of Fuchs, Hwang, and Zacharovas to improve the analysis of the internal path length of symmetric b-DSTs. Moreover, we will use the method to derive some exact and asymptotic results.

3.1 Introduction

Let L_n be the internal path length of the b-DSTs built from n records. Let $P_n(y) = \mathbb{E}[e^{L_n y}]$ be the moment generating function of L_n. Then $P_j(y) = 1$ with $j \le b$ and

$$P_{n+b}(y) = e^{ny} 2^{-n} \sum_{j=0}^{n} \binom{n}{j} P_j(y) P_{n-j}(y) \qquad (n \geqslant 1).$$

Next, we define $P(z, y) = \sum_{n \ge 0} \frac{P_n(y)}{n!} z^n$. This gives

$$\frac{\partial^b}{\partial z^b} P(z, y) = P\Big(\frac{e^y z}{2}, y\Big)^2.$$

Now we consider the Poisson generating function $\widetilde{P}(z, y) = e^{-z} P(z, y)$. Then

$$\sum_{j=0}^{b} \frac{\partial^j}{\partial z^j} \widetilde{P}(z, y) = e^{(e^y - 1)z} \widetilde{P}\Big(\frac{e^y z}{2}, y\Big)^2.$$

Thus, if we set $\widetilde{P}(z,y) = \sum_{m\geq 0} \frac{\widetilde{f}_m(z)}{m!} y^m$, then we obtain the following relations for the Poisson transforms of the first two moments

$$\sum_{j=0}^{b} \binom{b}{j} \widetilde{f}_1^{(j)}(z) = 2\widetilde{f}_1(z/2) + z, \tag{3.1}$$

$$\sum_{j=0}^{b} \binom{b}{j} \widetilde{f}_2^{(j)}(z) = 2\widetilde{f}_2(z/2) + 2\widetilde{f}_1(z/2)^2 + 4z\widetilde{f}_1(z/2) + 2z\widetilde{f}_1'(z/2) + z + z^2 \tag{3.2}$$

with initials $\widetilde{f}_k^{(j)}(0) = 0$ for $0 \leq j \leq b$ and $k = 1, 2$.

New method. First, we consider the recurrence of the general type:

$$\sum_{j=0}^{b} \binom{b}{j} \widetilde{f}^{(j)}(z) = 2\widetilde{f}(z/2) + g(z), \tag{3.3}$$

with initials $\widetilde{f}^{(j)}(0) = 0$ for $0 \leq j \leq b$. By using the Laplace transform, we can deduce a more simpler recurrence. More precisely, we denote the Laplace transform of $f(z)$ by $F(s)$ and obtain

$$\sum_{j=0}^{b} \binom{b}{j} s^j F(s) = 4F(2s) + G(s), \tag{3.4}$$

where $G(s)$ is the Laplace transform of $g(z)$. Define

$$\varphi(s) = \prod_{j\geq 1} (1 + 2^{-j}s)^b \tag{3.5}$$

and write

$$\hat{F}(s) = \frac{F(s)}{\varphi(s)}.$$

Then, we have

$$\hat{F}(s) = 4\hat{F}(2s) + \frac{G(s)}{\varphi(2s)},$$

and by iteration

$$\hat{F}(s) = \sum_{j\geq 0} 4^j \frac{G(2^j s)}{\varphi(2^{j+1}s)}. \tag{3.6}$$

Obviously, this is a harmonic sum. Therefore, we use the Mellin transform and get

$$F^*(w) = \frac{G^*(w)}{1-2^{2-w}},$$

where $F^*(w) = \mathcal{M}[\hat{F}; w]$ and

$$G^*(w) = \int_0^\infty \frac{s^{w-1}}{\varphi(2s)} \int_0^\infty e^{-sz} g(z)\,\mathrm{d}z\,\mathrm{d}s.$$

If $g(z) = \mathcal{O}(z^\beta)$ for large z, where $\beta < 1$, then $G(s) = \mathcal{O}(|s|^{-\max\{\beta+1,0\}})$ as $|s| \to 0$. From (2.22) ($\phi(s)^b = \varphi(2s)$), we know $1/\varphi(s)$ is very small at infinity. Thus, the Mellin transform $G^*(w)$ is well-defined in $\Re(w) > \max\{\beta+1, 0\}$ and from the inverse Mellin transform we have

$$\hat{F}(s) = \frac{1}{2\pi i} \int_{3-i\infty}^{3+i\infty} \frac{G^*(w)}{1-2^{2-w}} s^{-w}\,\mathrm{d}w.$$

Thus, by moving the line of integration to $\Re(w) = \max\{\beta+1+\epsilon, \epsilon\}$, and adding all residues at the poles at $w = \chi_j = 2j\pi i/\log 2$, we obtain

$$\hat{F}(s) = \frac{1}{\log 2} \sum_{j\in\mathbb{Z}} G^*(2+\chi_j) s^{-2-\chi_j} + \mathcal{O}\Big(|s|^{-\beta-1-\epsilon} + |s|^{-\epsilon}\Big)$$

as $|s| \to 0$. From (2.22) we know

$$\varphi(s) = 1 + bs + \mathcal{O}(|s|^2)$$

as $|s| \to 0$. Then we have

$$F(s) = \frac{1}{\log 2} \sum_{j\in\mathbb{Z}} G^*(2+\chi_j) s^{-2-\chi_j} + \frac{b}{\log 2} \sum_{j\in\mathbb{Z}} G^*(2+\chi_j) s^{-1-\chi_j} + \mathcal{O}\Big(|s|^{-\beta-1-\epsilon} + |s|^{-\epsilon}\Big)$$

as $|s| \to 0$. Finally, by the inverse Laplace transform, we get

$$\widetilde{f}(z) = \frac{1}{\log 2} \sum_{j\in\mathbb{Z}} \frac{G^*(2+\chi_j)}{\Gamma(2+\chi_j)} z^{1+\chi_j} + \frac{b}{\log 2} \sum_{j\in\mathbb{Z}} \frac{G^*(2+\chi_j)}{\Gamma(1+\chi_j)} z^{\chi_j} + \mathcal{O}\Big(|z|^{\beta+\epsilon} + |z|^{\epsilon-1}\Big). \tag{3.7}$$

From this we obtain the asymptotics of f_n via de-poissonization.

Mean. From (3.1), we know that for the expected internal path length we have $g(z) = z$, and by the process above we get

$$G^*(w) = \int_0^\infty \frac{s^{w-3}}{\varphi(2s)}\,\mathrm{d}s,$$

which has a simple pole at $w = 2$ (see (2.24)). Then, we can write

$$G^*(w) = \frac{1}{w-2} + G_1^*(w),$$

where

$$G_1^*(w) = \int_0^1 \Big(\frac{1}{\varphi(2s)} - 1\Big)s^{w-3}\,\mathrm{d}s + \int_1^\infty \frac{s^{w-3}}{\varphi(2s)}\,\mathrm{d}s,$$

and deduce that

$$\begin{aligned}\hat{F}(s)s^2 =& \frac{1}{2\pi i}\int_{3-i\infty}^{3+i\infty} \frac{G^*(w)}{1-2^{2-w}} s^{2-w}\,\mathrm{d}w \\ =& \log_2\frac{1}{s} + \frac{1}{2} + \frac{G_1^*(2)}{\log 2} + \frac{1}{\log 2}\sum_{j\in\mathbb{Z}\setminus\{0\}} G^*(2+\chi_j)s^{-\chi_j} + \mathcal{O}(|s|)\end{aligned} \tag{3.8}$$

as $|s| \to 0$ (in fact, $G_1^*(2)$ is equal to $J'(0)$ as (2.28)). Then we obtain

$$\begin{aligned}\widetilde{f}_1(z) =& z\log_2 z + z\Big(\frac{\gamma-1}{\log 2} + \frac{1}{2} + \frac{G_1^*(2)}{\log 2} + \frac{1}{\log 2}\sum_{j\in\mathbb{Z}\setminus\{0\}} \frac{G^*(2+\chi_j)}{\Gamma(2+\chi_j)} z^{\chi_j}\Big) \\ &+ b\log_2 z + \mathcal{O}(1).\end{aligned} \tag{3.9}$$

Thus, using Theorem 7 we obtain

$$\begin{aligned}\mathbb{E}[L_n] =& \widetilde{f}_1(n) + \mathcal{O}(1) \\ =& n\log_2 n + \Big(\frac{1}{\log 2}G_1^*(2) + \frac{1}{2} + \frac{\gamma}{\log 2} - \frac{1}{\log 2} + \delta_1(\log_2 n)\Big)n + \mathcal{O}(1).\end{aligned}$$

This coincides with the expansion obtained in Theorem 13.

Variance. To compute the variance, we consider the function

$$\widetilde{V}(z) := \widetilde{f}_2(z) - \widetilde{f}_1(z)^2 - z\widetilde{f}_1'(z)^2. \tag{3.10}$$

The advantage of using this function is

$$\widetilde{V}(n) = \widetilde{f}_2(n) - \widetilde{f}_1(n)^2 - n\widetilde{f}_1'(n)^2,$$

whose right hand side corresponds to the right hand of (1.23) in Theorem 7. From (3.2), $\widetilde{V}(z)$ satisfies the general type (3.3)

$$\sum_{j=0}^{b}\binom{b}{j}\widetilde{V}^{(j)}(z)=2\widetilde{V}(z/2)+g(z), \tag{3.11}$$

where

$$\begin{aligned} g(z)=&\Big(\sum_{j=0}^{b}\binom{b}{j}\widetilde{f}_1^{(j)}(z)\Big)^2+z\Big(\sum_{j=0}^{b}\binom{b}{j}\widetilde{f}_1^{(j+1)}(z)\Big)^2\\ &-\sum_{j=0}^{b}\binom{b}{j}\Big(\big(\widetilde{f}_1^2(z)\big)^{(j)}+\big(z\widetilde{f}_1'(z)^2\big)^{(j)}\Big). \end{aligned} \tag{3.12}$$

Remark. (1) Note that from (3.9), we have

$$g(z)=\frac{b}{z}\Big(\frac{1}{\log 2}+\sum_{j\in\mathbb{Z}}c_jz^{\chi_j}\Big)^2+\mathcal{O}\big(|z|^{-2}\big) \tag{3.13}$$

as $|z|\to\infty$.

(2) $g(z)=g_b(z)$ is given by (for simplicity, we set $\phi_j=\widetilde{f}_1^{(j)}(z)$)

$$\begin{aligned} g_1(z)=&z\phi_2^2,\\ g_2(z)=&z(2\phi_2^2+4\phi_2\phi_3+\phi_3^2)+\phi_2^2,\\ g_3(z)=&z(3\phi_2^2+12\phi_2\phi_3+6\phi_2\phi_4+9\phi_3^2+6\phi_3\phi_4+\phi_4^2)\\ &+3\phi_2^2+6\phi_2\phi_3+\phi_3^2,\\ g_4(z)=&z(4\phi_2^2+24\phi_2\phi_3+24\phi_2\phi_4+8\phi_2\phi_5+30\phi_3^2\\ &+48\phi_3\phi_4+12\phi_3\phi_5+16\phi_4^2+8\phi_4\phi_5+\phi_5^2)\\ &+6\phi_2^2+24\phi_2\phi_3+12\phi_2\phi_4+16\phi_3^2+8\phi_3\phi_4+\phi_4^2. \end{aligned}$$

Now, by the same process as for the mean, we get

$$\begin{aligned} \mathbb{V}[L_n]=&\widetilde{V}(n)+\mathcal{O}(1)\\ =&\frac{G^*(2)}{\log 2}n+\frac{1}{\log 2}\sum_{j\in\mathbb{Z}\setminus\{0\}}G^*(2+\chi_j)n^{1+\chi_j}+\mathcal{O}(1), \end{aligned}$$

which is as same as (2.34), where

$$\frac{G^*(2)}{\log 2}=\frac{1}{\log 2}\int_0^\infty\frac{s}{\varphi(2s)}\int_0^\infty e^{-sz}g(z)\,\mathrm{d}z\,\mathrm{d}s.$$

Note that the last expression is much easier than the corresponding expression in Theorem 14. In particular, we can use Maple to obtain the values for small b (see Table 3.1). We will do this in Section 3.3.

Table 3.1: Some values of the leading constant $G^*(2)/\log 2$.

b	1	2	3	4	5
$G^*(2)/\log 2$	0.26600	0.13260	0.09004	0.06958	0.05781

3.2 Exact Results

b=1. First, we consider the case $b = 1$. By (3.4) and iteration, we obtain that the Laplace transform $\widetilde{F}_1(s)$ of $\widetilde{f}_1(z)$ satisfies

$$\widetilde{F}_1(s) = \sum_{j\geq 0} \frac{1}{s^2} \frac{1}{(s+1)\cdots(2^j s+1)}.$$

Now, by partial fraction expansion, we have

$$\frac{1}{(s+1)\cdots(2^j s+1)} = \sum_{0\leq h\leq j} \frac{(-1)^{j-h} 2^{-\binom{j-h}{2}-j}}{(s+2^{-h})Q_h Q_{j-h}},$$

where $Q_n = \prod_{1\leq j\leq n}(1-2^{-j})$, $Q_0 = 1$. Thus,

$$\begin{aligned}\widetilde{F}_1(s)s^2 &= \sum_{j\geq 0}\sum_{0\leq h\leq j} \frac{(-1)^{j-h} 2^{-\binom{j-h}{2}-j}}{(s+2^{-h})Q_h Q_{j-h}} \\ &= \sum_{h\geq 0} \frac{1}{Q_h(2^h s+1)} \sum_{j\geq 0} \frac{(-1)^j 2^{-\binom{j+1}{2}}}{Q_j}.\end{aligned}$$

By the Euler identity

$$1 + \sum_{j\geq 1} \frac{q^{\binom{j}{2}} z^j}{(1-q)\cdots(1-q^j)} = \prod_{k\geq 0}\Big(1+q^k z\Big),$$

we can see that

$$\sum_{j\geq 0} \frac{(-1)^j 2^{-\binom{j+1}{2}}}{Q_j} = Q_\infty = \lim_{n\to\infty} Q_n \approx 0.288788\ldots$$

This gives

$$\widetilde{F}_1(s) = \frac{Q_\infty}{s^2} \sum_{h\geq 0} \frac{1}{Q_h(2^h s+1)},$$

and then by inverse Laplace transform

$$\widetilde{f}_1(z) = Q_\infty \sum_{h\geq 0} \frac{2^h}{Q_h}\Big(e^{-z/2^h} - 1 + \frac{z}{2^h}\Big). \tag{3.14}$$

Next, from (3.12) we have $g_1(z) = z\widetilde{f}_1''(z)^2$ for $b = 1$. By (3.14), we have

$$\widetilde{f}_1''(z) = \sum_{h\geq 0} \frac{Q_\infty}{Q_h 2^h} e^{-z/2^h},$$

and thus, letting $G_1(s)$ be the Laplace transform of $g_1(z)$, we get

$$G_1(s) = \sum_{h,k\geq 0} \frac{Q_\infty^2}{Q_h Q_k 2^{h+k}} \cdot \frac{1}{(s + 2^{-h} + 2^{-k})^2}.$$

Then, we have

$$\begin{aligned} G_1^*(w) =& \mathcal{M}[G_1(s)/\varphi(2s); w] \\ =& \sum_{h,k\geq 0} \frac{Q_\infty^2}{Q_h Q_k 2^{h+k}} \int_0^\infty \frac{s^{w-1}}{\varphi(2s)(s + 2^{-h} + 2^{-k})^2}\, \mathrm{d}s. \end{aligned}$$

Hence, By the identity

$$\frac{1}{\varphi(2s)} = \sum_{j\geq 0} \frac{(-1)^j 2^{-\binom{j}{2}}}{Q_j Q_\infty (s + 2^j)},$$

we have the explicit form for $G_1^*(w)$:

$$G_1^*(w) = Q_\infty \sum_{j,h,k\geq 0} \frac{(-1)^j 2^{-\binom{j}{2}}}{Q_j Q_h Q_k 2^{h+k}} \int_0^\infty \frac{s^{w-1}}{(s + 2^j)(s + 2^{-h} + 2^{-k})^2}\, \mathrm{d}s.$$

This explicit form is more simpler than (2.5) and (2.35).

b=2. For $b = 2$, we have

$$\widetilde{F}_1(s) = \sum_{j\geq 0} \frac{1}{s^2} \frac{1}{(s+1)^2 \cdots (2^j s + 1)^2},$$

which, by partial fraction expansion,

$$\frac{1}{(s+1)^2 \cdots (2^j s + 1)^2} = \sum_{0\leq l\leq j} \left(\frac{\triangle_1^j(l)}{2^l s + 1} + \frac{\triangle_2^j(l)}{(2^l s + 1)^2} \right),$$

where

$$\begin{aligned}\triangle_1^j(l) =& \frac{2^{-(j-l+1)(j-l)}}{Q_l^2 Q_{j-l}^2} \sum_{q=0,q\neq l}^{j} \frac{-2}{(2^{l-q}-1)},\\ \triangle_2^j(l) =& \frac{2^{-(j-l+1)(j-l)}}{Q_l^2 Q_{j-l}^2},\end{aligned}$$

has the form

$$\begin{aligned}\widetilde{F}_1(s) =& \frac{1}{s^2}\sum_{j\geq 0}\sum_{l=0}^{j}\left(\frac{\triangle_1^j(l)}{2^l s+1}+\frac{\triangle_2^j(l)}{(2^l s+1)^2}\right)\\ =& \frac{1}{s^2}\sum_{l\geq 0}\frac{A(l)}{Q_l^2(2^l s+1)}+\frac{1}{s^2}\sum_{l\geq 0}\frac{A}{Q_l^2(2^l s+1)^2},\end{aligned}$$

where

$$\begin{aligned}A(l) =& \sum_{j\geq 0}\frac{2^{-j(j+1)}}{Q_j^2}\sum_{q=0,q\neq l}^{j+l}\frac{-2}{(2^{l-q}-1)},\\ A =& \sum_{j\geq 0}\frac{2^{-j(j+1)}}{Q_j^2}\approx 2.113388773\ldots\end{aligned}$$

Thus, we have

$$\widetilde{f}_1(z)=\sum_{l\geq 0}\frac{A(l)2^l}{Q_l^2}\left(e^{-z/2^l}-1+\frac{z}{2^l}\right)+\sum_{l\geq 0}\frac{A}{Q_l^2}\left(ze^{-z/2^l}+2^{l+1}e^{-z/2^l}-2^{l+1}+z\right). \quad (3.15)$$

Similarly, from (3.12) and (3.15), we get a more complicated form

$$G_2(s)=\sum_{h,l\geq 0}\frac{1}{Q_h^2 Q_l^2 2^{2h+2l}}\sum_{k=1}^{4}\frac{\alpha_k(h,l)}{(s+2^{-h}+2^{-l})^k},$$

where

$$\begin{aligned}\alpha_1(h,l) =& 2^{h+l}A(h)A(l),\\ \alpha_2(h,l) =& A^2+(2^{h+1}3-2)AA(h)+(2^{1+h+l}-2^{2+h}+1)A(h)A(l),\\ \alpha_3(h,l) =& (10-2^{1-h}-2^{1-l})A^2+(2^{3+h}+2^{2-l}-2^{3+h-l}-2^3)AA(h),\\ \alpha_4(h,l) =& 6(1+2^{-h-l}-2^{2-h})A^2.\end{aligned}$$

By the identity

$$\frac{1}{\varphi(2s)} = \sum_{p,q=0}^{\infty} \frac{\Lambda(p,q)}{(s+2^p)(s+2^q)},$$

where

$$\Lambda(p,q) = \frac{(-1)^{p+q} 2^{-\binom{p}{2}-\binom{q}{2}}}{Q_p Q_q Q_\infty^2},$$

we have

$$\begin{aligned} G_2^*(w) =& \mathcal{M}[G_2(s)/\varphi(2s); w] \\ =& \sum_{k=1}^{4} \left(\sum_{h,l,p,q\geq 0} \frac{\alpha_k(h,l)\Lambda(p,q)}{Q_h^2 Q_l^2 2^{2h+2l}} \int_0^\infty \frac{s^{w-1}}{(s+2^p)(s+2^q)(s+2^{-h}+2^{-l})^k} ds \right). \end{aligned}$$

Again this is easier than (2.5) and (2.35). Moreover, by similar computations, larger values of b can be treated as well.

3.3 Numerical Results

In this section, we discuss the computation of the numerical value of $G^*(2)$. First, we separate the integral into three parts

$$\begin{aligned} \int_0^\infty \frac{s}{\varphi(2s)} \int_0^\infty e^{-sz} g(z)\,\mathrm{d}z\,\mathrm{d}s =& \int_N^\infty \frac{s}{\varphi(2s)} \int_0^T e^{-sz} g(z)\,\mathrm{d}z\,\mathrm{d}s \\ &+ \int_0^N \frac{s}{\varphi(2s)} \int_0^T e^{-sz} g(z)\,\mathrm{d}z\,\mathrm{d}s \\ &+ \int_0^\infty \frac{s}{\varphi(2s)} \int_T^\infty e^{-sz} g(z)\,\mathrm{d}z\,\mathrm{d}s. \end{aligned}$$

Now we consider the first part of the integral. Since $g(z)$ is given by (3.12) and $\widetilde{f}_1^{(j)}(0) = 0$ for $j \leq b$ and $\widetilde{f}_1^{(b+1)}(0) = 1$, we have

$$g(z) = \mathcal{O}(|z|),$$

as $|z| \to 0$. Then,

$$\int_0^\infty g(z) e^{-sz}\,\mathrm{d}z = \mathcal{O}(s^{-2}),$$

for large s. Thus, from this and (2.21) we obtain for N tends to infinity,

$$\begin{aligned}\int_N^\infty \frac{s}{\varphi(2s)} \int_0^T g(z)e^{-sz}\,\mathrm{d}z\,\mathrm{d}s &= \mathcal{O}\bigg(\int_N^\infty s^{-1-b/2}e^{-b\log^2 s/(2\log 2)}\,\mathrm{d}s\bigg)\\ &= \mathcal{O}\Big(N^{-b/2}e^{-b\log^2 N/(2\log 2)}\Big).\end{aligned}$$

This means that if we choose N large enough, then the first part can be safely neglected.

Next, we need a good way of computing $g(z)$ for the second integral. We first consider $\widetilde{f}_1(z)$. Since $\widetilde{f}_1(z) = e^{-z}\sum_{n>b}\mu_n z^n/n!$ is an entire function, we have

$$\widetilde{f}_1(z) \approx e^{-z}\sum_{b<n\le N}\frac{\mu_n}{n!}z^n \qquad (0\le z\le T), \tag{3.16}$$

and the error term introduced by this approximation is

$$e^{-z}(\log N)\frac{z^{N+1}}{N!} \approx \frac{z\log N}{\sqrt{2\pi N}}e^{-z}\Big(\frac{ez}{N}\Big)^N,$$

which is very small if we choose T small enough, say $N \gg eT$. So, in order to compute $\widetilde{f}_1(z)$, we just have to generate $\mu_{b+1},\dots,\mu_N$ (this can be done via the recurrence satisfied by μ_n) and then use (3.16). Since a similar approach also works for the derivatives of $\widetilde{f}_1(z)$, this can be used to compute $g(z)$ as well.

Finally, we consider the last integral

$$\int_0^\infty \frac{s}{\varphi(2s)}\int_T^\infty e^{-sz}g(z)\,\mathrm{d}z\,\mathrm{d}s = \int_T^\infty g(z)\int_0^\infty \frac{s}{\varphi(2s)}e^{-sz}\,\mathrm{d}s\,\mathrm{d}z. \tag{3.17}$$

We use Watson's lemma to get an asymptotic expansion of the Laplace transform of $s/\varphi(2s)$:

Lemma 9. *Consider a Laplace integral $\int_0^\infty f(s)e^{-zs}\,\mathrm{d}s$ and assume*

(i) $f(s)$ has a power series expansion which converges for $|s|<R$.

(ii) There exists an $\alpha>0$ such that $f(s)=\mathcal{O}(e^{\alpha s})$ as $s\to\infty$.

Then,

$$\int_0^\infty f(s)e^{-zs}\,\mathrm{d}s \sim \sum_{h=0}^\infty \frac{f^{(h)}(0)}{z^{h+1}}.$$

We still need an asymptotic expansion of $g(z)$ for z large. Again first consider $\widetilde{f}_1(z)$. Here, we put

$$\frac{1}{\varphi(2s)} \sim \sum_{h\geq 0} \varphi_h s^h.$$

This implies that

$$\hat{G}^*(w) = \int_0^\infty \frac{s^{w-3}}{\varphi(2s)}\,\mathrm{d}s = \int_0^\infty \sum_{h\geq 0} \varphi_h s^{w+h-3}\,\mathrm{d}s$$

has simple poles at $w = 2-h$. Then, similar as in (3.8), we obtain

$$\hat{F}(s) = \frac{1}{s^2}\log_2\frac{1}{s} + \frac{1}{2s^2} + \frac{\hat{G}_1^*(2)}{\log 2 s^2} + \frac{1}{\log 2}\sum_{j\in\mathbb{Z}\setminus\{0\}} \hat{G}^*(2+\chi_j)s^{-2-\chi_j} - \sum_{h\geq 1}\frac{\varphi_h s^{h-2}}{2^h-1},$$

as $|s| \to 0$. Next, put

$$\varphi(s) \sim \sum_{h\geq 0} \widetilde{\varphi}_h s^h.$$

From this and the expression above, we get

$$\begin{aligned} F(s) \approx & \frac{\widetilde{\varphi}_0}{s^2}\log_2\frac{1}{s} + \frac{\widetilde{\varphi}_0}{2s^2} + \frac{\widetilde{\varphi}_0\hat{G}_1^*(2)}{\log 2 s^2} + \frac{\widetilde{\varphi}_0}{\log 2}\sum_{j\in\mathbb{Z}\setminus\{0\}} \hat{G}^*(2+\chi_j)s^{-2-\chi_j} - \widetilde{\varphi}_0\varphi_1 s^{-1} \\ & + \frac{\widetilde{\varphi}_1}{s}\log_2\frac{1}{s} + \frac{\widetilde{\varphi}_1}{2s} + \frac{\widetilde{\varphi}_1\hat{G}_1^*(2)}{\log 2 s} + \frac{\widetilde{\varphi}_1}{\log 2}\sum_{j\in\mathbb{Z}\setminus\{0\}} \hat{G}^*(2+\chi_j)s^{-1-\chi_j} \\ & + \sum_{h\geq 2}\frac{\widetilde{\varphi}_h}{\log 2}\sum_{j\in\mathbb{Z}\setminus\{0\}} \hat{G}^*(2+\chi_j)s^{-2+h-\chi_j}. \end{aligned}$$

Here, we have dropped all terms with s^i, $i \geq 0$ (their contributions to $\widetilde{f}_1(z)$ is negligible). Then, by inverse Laplace transform,

$$\begin{aligned} \widetilde{f}_1(z) \sim & \frac{\widetilde{\varphi}_0}{\log 2} z(-1+\gamma+\log z) + \frac{\widetilde{\varphi}_0}{2}z + \frac{\widetilde{\varphi}_0\hat{G}_1^*(2)}{\log 2}z + \frac{\widetilde{\varphi}_0}{\log 2}\sum_{j\in\mathbb{Z}\setminus\{0\}} \frac{\hat{G}^*(2+\chi_j)}{\Gamma(2+\chi_j)} z^{1+\chi_j} \\ & - \widetilde{\varphi}_0\varphi_1 + \frac{\widetilde{\varphi}_1}{\log 2}(\gamma+\log z) + \frac{\widetilde{\varphi}_1}{2} + \frac{\widetilde{\varphi}_1\hat{G}_1^*(2)}{\log 2} + \frac{\widetilde{\varphi}_1}{\log 2}\sum_{j\in\mathbb{Z}\setminus\{0\}} \frac{\hat{G}^*(2+\chi_j)}{\Gamma(1+\chi_j)} z^{\chi_j} \\ & + \sum_{h\geq 2}\frac{\widetilde{\varphi}_h}{\log 2}\sum_{j\in\mathbb{Z}\setminus\{0\}} \frac{\hat{G}^*(2+\chi_j)}{\Gamma(2-h+\chi_j)} z^{1-h+\chi_j}. \end{aligned} \tag{3.18}$$

By differentiation we obtain similar expansions for the derivatives of $\widetilde{f}_1(z)$ and hence for $g(z)$.

Plugging this expansion together with the expansion obtained from Watson's lemma for the Laplace transform of $s/\varphi(2s)$ into (3.17) and integration then yields an approximation of (3.17) up to arbitrary large order. Choosing sufficiently many terms will then give a good approximation for very small T.

Finally, we explain how to compute $\hat{G}^*(2+\chi_j)$ involved in (3.18). First notice that due to the very fast decay of $\hat{G}^*(2+\chi_j)$ only $j=0,1,-1$ are needed. For the computation we use the following result which is due to Flajolet and Richmond.

Lemma 10. *The function $J(s)$ defined in (2.26) admits the representation*

$$J(s) = A_0(2^s) + (s-1)A_1(2^s) + \cdots + (s-1)(s-2)\cdots(s-b+1)A_{b-1}(2^s),$$

where $A_k(x)$'s are entire functions, and

$$A_k(x) = \frac{(-1)^k}{k!}\cdot\frac{1}{Q_\infty^b}\sum_{j=0}^{\infty}(-1)^{jb}\frac{2^{-bj(j+1)/2}}{Q_j^b}Y_{b-1-k}(j)(x2^{b-1-k})^j,$$

with $Q_m=\prod_{l=1}^m(1-2^{-l})$, $Q_\infty=\prod_{l=1}^\infty(1-2^{-l})$ and the $Y_\beta(j)$ are defined by

$$\sum_{\beta\geq 0}Y_\beta(j)w^\beta = \exp\left(b\sum_{\alpha\geq 1}\frac{(-1)^\alpha w^\alpha}{\alpha}\Big(\sum_{l\neq j}\frac{1}{(2^l-2^j)^\alpha}\Big)\right).$$

Overall, by incorporating all the above ideas, we obtain the following program for computing the required values.

```
> b := 2; T := 30; N := 4*T;
> μ := vector(120);
  for y from 1 to b do
  μ[y] := 0
  end do;
> for y from 1 to 120-b do
  μ[y+b] := y+2^(1-y)*(sum(binomial(y, i)*μ[i], i=1..y))
  end do;
> fnew := exp(-z)*(sum(μ[i]*z^i/i!, i=1..120));
> f := vector(b);
  for y from 1 to b do
  f[y] := diff(fnew, z$y+1)
  end do;
> evalf(evalf(Int(s*exp(-z*s)*((2*z+1)*f[1]^2+z*f[2]^2+4*z*f[1]*f[2])/
  evalf(Product(1+s/2^i, i=0..60))^b, [z=0..T, s=0..N]))/ln(2));
```

0.1300797679

```
> with(PolynomialTools);
  φ := sum(sum(b*(-1)^(n-1)*s^n/(n*(2^n-1)), n=1..12)^i/i!, i=0..12);
  coeff_phi := CoefficientVector(expand(φ), s);
> q := evalf(add((-1)^j*2^(-binomial(j+1, 2)) /evalf(Product(1-2^(-i), i=1..j)), j=0..60));
  e := sum((-1)^i*w^i*(sum(1/(2^d-2^j)^i, d=0..j-1)+sum(1/(2^d-2^j)^i, d=j+1..60
  ))/i, i=1..b);
> coeff_y := CoefficientVector(expand(sum((b*e)^i/i!, i=0..b-1)), w);
> a := vector(b);
  for k from 0 to b-1 do
  a[k+1] := evalf((-1)^k*add((-1)^(b*j)*2^(-(1/2)*b*j*(j+1))*coeff_y[b-k]*2^((b-1-k)
  *j)/(Product(1-2^(-i), i=1..j))^b, j=0..60)/(k!*q^b))
  end do;
> G := (s) → evalf(Pi*(sum((product(s-j, j=1..i-1))*a[i], i=1..b))/sin(Pi*s));
> χ := (2*Pi*I)/log(2);
  func_vec := vector(13);
  func_vec[1] := z*(-1+γ+ln(z))+G(χ)*z^(1+χ)/GAMMA(2+χ)
  +G(-χ)*z^(1-χ)/GAMMA(2-χ);
  func_vec[2] := γ+ln(z)+G(χ)*z^χ/GAMMA(1+χ)+G(-χ)*z^(-χ) /GAMMA(1-χ);
  for k from 3 to 13 do
  func_vec[k] := z^(-k+2)*(-1)^(k-3)*(k-3)!+G(χ)*z^(-k+2+χ)/GAMMA(-k+3+χ)
  +G(-χ)*z^(-k+2-χ)/GAMMA(-k+3-χ)
  end do;
```

```
> gnew := 0;
  for k from 3 to 13 do
  gnew := gnew+coeff_phi[k]*func_vec[k]
  end do;
  gnew := gnew/ln(2);
> g := vector(b);
  for k from 1 to b do
  g[k] := diff(gnew, z$k+1))
  end do;
> ϕ2 := s*(sum((sum(-b*(-1)^(n-1)*2^n*s^n/(n*(2^n-1)), n=1..12))^i/i!, i=0..12));
  coeff_phi_2 := CoefficientVector(expand(ϕ2), s);
> watson := 0;
  for k from 1 to 13 do
  watson := watson + coeff_phi_2[k+1]*k!/z^(k+1)
  end do;
> evalf(Int(((2*z+1)*g[1]^2+z*g[2]^2+4*z*g[1]*g[2])*watson/log(2), z=T..∞,
  method = _CCquad));
```

$$0.002522975851 + 1.220918538 \cdot 10^{-19} I$$

```
> 0.002522975851 + 0.1300797679;
```

$$0.1326027438$$

Chapter 4

Conclusion

To conclude this thesis, we briefly summarize the main contributions.

Our first goal was to give a self-contained survey of recent results in the analysis of DSTs. These results have been widely spread across the research literature before and this is up to our knowledge the first time that they appear in collected form. For clearance of presentation, we started by explaining the main techniques used in the analysis of DSTs in Chapter 1. Then, we showed applications of these techniques in Chapter 2. Therefore, we used the internal path length as guiding example and concentrated on mean value and variance. It should be stressed that all the results from Chapter 1 and Chapter 2 are not original. However, we improved and shortened several proofs, in particular, those concerned with the analysis of asymmetric DSTs.

Our second goal was to introduce a new approach of Fuchs, Hwang, and Zacherovas (which has not appeared yet) and to compare it with previous approaches from Chapter 2. This new approach has two major improvements: one is the use of the Laplace transform which simplifies the overall analysis and the other is the consideration of $\widetilde{V}(z)$ which makes the computation of the variance more easier. Then, we applied the new approach to the variance of the internal path length of b-DSTs and obtained more simpler expressions for the leading constant in the asymptotic expansion for $b = 1, 2$ (larger values of b can be treated as well). Finally, we explained how to obtain numerical values of the leading constant for small values of b via Maple. Our numerical computations suggest that previous computations contain several imprecisions.

Bibliography

[1] R. Aguech, N. Lasmar and H. Mahmoud, Distances in random digital search trees, *Acta Informatica*, **43** (2006) 243–264.

[2] P. Flajolet, X. Gourdon and P. Dumas, Mellin transforms and asymptotics: Harmonic sums, *Theoret. Comput. Sci.*, **144** (1995) 3–58.

[3] P. Flajolet and A. Odlyzko, Singularity analysis of generating functions, *SIAM J. Discrete Math.*, **3** (1990) 216–240.

[4] P. Flajolet and B. Richmond, Generalized digital trees and their difference-differential equations, *Random Structures and Algorithms*, **3** (1992) 305–320.

[5] P. Flajolet and R. Sedgewick, Digital search trees revisited, *SIAM J. Comput.*, **15** (1986) 748–767.

[6] P. Flajolet and R. Sedgewick, Mellin transforms and asymptotics: Finite differences and Rice's integrals, *Theoret. Comput. Sci.*, **144** (1995) 101–124.

[7] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*, Cambridge University Press, 2009.

[8] F. Hubalek, On the variance of the internal path length of generalized digital trees - The Mellin convolution approach, *Theoret. Comput. Sci.*, **242** (2000) 143–168.

[9] F. Hubalek, H. Hwang, W. Lew, H. Mahmoud and H. Prodinger, A multivariate view of random bucket digital search trees, *Journal of Algorithms*, **44** (2002) 121–158.

[10] P. Jacquet and W. Szpankowski, Asymptotic behavior of the Lempel-Ziv parsing scheme and digital search trees, *Theoret. Comput. Sci.*, **144** (1995) 161–197.

[11] P. Jacquet and W. Szpankowski, Analytical depoissonization and its applications, *Theoret. Comput. Sci.*, **201** (1998) 1–62.

[12] P. Kirschenhofer and H. Prodinger, Further results on digital search trees, *Theoret. Comput. Sci.*, **58** (1988) 143–154.

[13] P. Kirschenhofer and H. Prodinger, Eine Anwendung der Theorie der Modulfunktionen in der Informatik. Sitzungsber., Abt. II, Österr. Akad. Wiss., Math.-Naturwiss. Kl. 197, No.4-7, (1988) 339-366.

[14] P. Kirschenhofer, H. Prodinger and W. Szpankowski, Digital search trees again revisited: The internal path length perspective, *SIAM J. Comput.*, **23** (1994) 598–616.

[15] D.E. Knuth, *The Art of Computer Programming, Vol. 3: Sorting and Searching*, Addison-Wesley, Reading, MA, 1973.

[16] A.G. Konheim and D.J. Newman, A note on growing binary trees, *Discrete Math.*, **4** (1973) 57–63.

[17] G. Louchard, Exact and asymptotic distributions in digital and binary search trees, *RAIRO Theoret. Inform. Appl.*, **21** (1987) 479–495.

[18] G. Louchard, Digital search trees revisited, *Cahiers Centre Études Rech. Oper.*, **36** (1995) 259–278.

[19] G. Louchard and W. Szpankowski, Average profile and limiting distribution for a phrase size in the Lempel-Ziv parsing algorithm, *IEEE Trans. Inform. Theory*, **41** (1995) 478–488.

[20] G. Louchard, W. Szpankowski and J. Tang, Average profile of the generalized digital search tree and the generalized Lempel-Ziv algorithm, *SIAM J. Comput.*, **28** (1999) 904–934.

[21] H. Prodinger, External internal nodes in digital search tree via Mellin transforms, *SIAM J. Comput.*, **21** (1992) 1180–1183.

[22] W. Szpankowski, A characterization of digital search trees from the successful search viewpoint, *Theoret. Comput. Sci.*, **85** (1991) 117–134.

[23] W. Szpankowski, *Average Case Analysis of Algorithms on Sequences*, John Wiley & Sons, Inc., New York, NY, 2001.

[24] E.C. Titchmarsh and D.R. Heath-Brown, *The Theory of the Riemann Zeta-function* (Oxford Science Publications, Oxford, 2nd ed., 1986).