國　立　交　通　大　學

應 用 數 學 系

碩 士 論 文

字典樹與基數樹之
二階保護點的數量

The Number of 2-Protected Nodes in Tries and PATRICIA Tries

研 究 生 ：余冠儒

指導教授 ：符麥克 教授

中 華 民 國 一 百 零 三 年 六 月

字典樹與基數樹之二階保護點的數量

The Number of 2-Protected Nodes in Tries and PATRICIA Tries

研 究 生 : 余冠儒　Student : Guan-Ru Yu

指導教授 : 符麥克　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 Science
National Chiao Tung University
in Partial Fulfillment of Requirements
for the Degree of Master
in Applied Mathematics

June 2014
Hsinchu, Taiwan

中 華 民 國 一 百 零 三 年 六 月

The Number of 2-Protected Nodes in Tries and PATRICIA Tries

Guan-Ru Yu

Department of Applied Mathematics,
National Chiao Tung University

This thesis was supervised by Dr. Michael Fuchs

摘 要

數位樹在電腦科學中的資料結構扮演極其重要的角色，而所謂的二階保護點在近期則受到相當大的矚目。舉例來說，J. Gaither、Y. Homma、M. Sellke 以及 M. D. Ward 曾探討隨機的字典樹之二階保護點的數量進而求出它的期望值之漸近展開式。不但如此，J. Gaither 與 M. D. Ward 更進一步求出它的變異數之漸近展開式，並猜測它數量的分布會滿足中央極限定理。

在這篇論文中，我們的主要目標是用符麥克、黃顯貴和 V. Zacharovas 三人所提出來一個有系統的方法來重新驗證（也可說是糾正）他們的結果。經過運算，即便我們得到一個與他們截然不同的展開式，但我們的在數值上與他們的結果卻是完全相同的。不但如此，我們更是證實了他們猜想的中央極限定理。事實上，我們證明了一個更一般化的結果就是字典樹之內點與二階保護點的二元中央極限定理。根據這個結果我們不但是證明了 J. Gaither 與 M. D. Ward 的猜想，更是直接得知基數樹也會滿足中央極限定理。最後，我們還算出基數樹之二階保護點數量的期望值及變異數之漸近展開式。

Preface

Digital trees are data structures which are of fundamental importance in Computer Science. Recently, so-called 2-protected nodes have attracted a lot of attention. For instance, J. Gaither, Y. Homma, M. Sellke, and M. D. Ward [6] derived an asymptotic expansion for the mean of the number of 2-protected nodes in random tries. Moreover, in [7], J. Gaither and M. D. Ward found an asymptotic expansion of the variance and conjectured a central limit theorem.

In this thesis, our main goal is to re-derive (and correct) their results by using a systematic method due to M. Fuchs, H.-K. Hwang, and V. Zacharovas [4]. The resulting expressions we obtain are quite different from the one in [6, 7], but numerically they of course coincide. Moreover, we prove the conjectured central limit theorem from [7]. In fact, we prove even a more general result, namely, a bivariate central limit theorem for the number of internal nodes and the number of 2-protected nodes in random tries. From this, not only the conjecture from [7] follows but we also obtain a central limit theorem for PATRICIA tries. Finally, we also derive asymptotic expansions of mean and variance for PATRICIA tries.

Next, we are going to give a short sketch of the structure of the thesis.

In Chapter 1, we give the definition of the three main classes of digital trees. Moreover, we introduce the random model and survey results on the number of internal nodes and number of 2-protected nodes in random tries. Finally, we explain the method from [4] which will be used in this thesis.

In Chapter 2, we will give a summary of the main tools of this method which are Mellin transform and analytic depoissonization. Moreover, we will recall the recent notion of JS-admissibility which can be used to systematically check assumptions for analytic depoissonization.

In Chapter 3, we will apply the method and tools from Chapter 1 and Chapter 2 to derive asymptotic expansions for mean and variance of the number of 2-protected nodes in random tries. An interesting aspect of our results is that they contain divergent series in the classical sense which however make sense if one appeals to the theory of Abel summability. We also show that our expressions (numerically) coincide with those from [6, 7]. In the final section of this chapter, we will prove that the number of 2-protected nodes satisfies the claimed central limit theorem.

Chapter 4 contains our results for PATRICIA tries. Here, the result for the mean follows from that of tries and the result for the variance follows from the same approach as used in Chapter 3. In order to keep the thesis short, we will however not repeat all the details, but only show the results.

In Chapter 5, we prove that the number of internal nodes and the number of 2-protected nodes satisfy a bivariate central limit theorem. As mentioned above, this result entails that also the number of 2-protected nodes of PATRICIA tries satisfies a central limit theorem.

Finally, we end the thesis with a conclusion in Chapter 6.

誌　　謝

寫著致謝的同時，也將整個碩士生涯快速的瀏覽一遍。首先我最要感謝指導教授 符麥克老師兩年來的栽培，亦師亦友的師生關係在我做研究的路上給了我充分的空間，引導我對自己的未來譜出不同以往的願景，讓我在學術上與生活上都留下許多深刻且美好的體驗。同時，我也要感謝我的學長李忠達，扮演老師的角色從旁輔導，給與我許多應時的協助。

除此還要感謝乙組所有的同學，每每與你們聊天課業壓力就拋諸腦後。我也要感謝新竹市召會的弟兄姊妹在我這兩年的求學階段時常給我鼓勵，讓我感覺教會就像我第二個家。

最後，我要感謝我的家人提供我一個溫暖的窩，讓我在忙碌的工作之餘可以沉澱情緒，謝謝爸爸、媽媽對我的疼愛和信任，讓我可以全心專注在自己喜歡的事物上，您們的鼓勵與支持是我向前的主要動力；謝謝我最親愛的妻子，在我忙於課業不在妳身邊的時候，常讓妳獨自去產檢，委屈妳了。謝謝妳的無怨無悔的付出與無條件的支持。

感謝神，為著祂難以形容的恩賜。

Contents

Chapter 1

Introduction

When it comes to data storage and data search, digital trees are one of the most effective data structures in computer science. Some famous applications of digital trees are in searching, sorting, dynamic hashing, coding, polynomial factorization, regular languages, contention tree algorithms, automatically correcting words in texts, retrieving IP address and satellite data, internet routing and so on. Due to their numerous applications, many recent studies have been concerned with the analysis of digital trees.

1.1 Definition

Before we start with the main topic of this thesis, we will introduce the three main classes of digital trees, namely, digital search trees, tries and PATRICIA tries (in the binary case). What they have in common is that the data consists of n infinite $\{0, 1\}$-strings. From this data they are built recursively as described in the three subsections below.

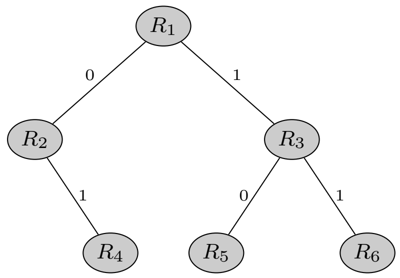
1.1.1 Digital search trees (DSTs)

Put the first string into the root. As for the other strings, direct them to the left or right subtree according to whether the first bit is 0 or 1 respectively. Next, the subtrees are built recursively using the same rules, but the direction is based on the second bit, etc. For an example; see Figure 1.1.

Remark 1. Digital search trees have the weak point that the cost of searching is very high since a comparison in every nodes is necessary.

Remark 2. From a practical point of view, digital search trees are not very important due to the above weak point. However, they are of theoretical interest, since they are closely related to the Lempel-Ziv compression scheme.

$$
\begin{aligned}
R_1 &= 000001\cdots\\
R_2 &= 000110\cdots\\
R_3 &= 110111\cdots\\
R_4 &= 011011\cdots\\
R_5 &= 100001\cdots\\
R_6 &= 111110\cdots
\end{aligned}
$$



$$
\begin{aligned}
R_6 &= 111110\cdots\\
R_5 &= 100001\cdots\\
R_4 &= 011011\cdots\\
R_3 &= 110111\cdots\\
R_2 &= 000110\cdots\\
R_1 &= 000001\cdots
\end{aligned}
$$

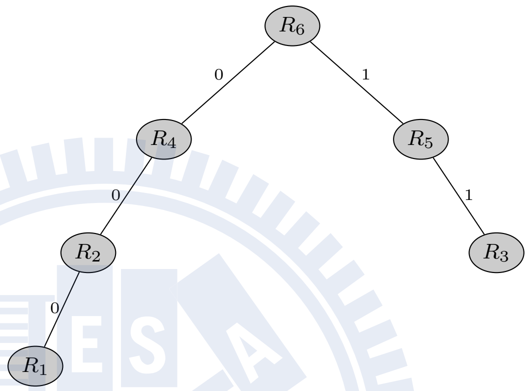


Figure 1.1: Two digital search trees built from the same keys $R_1, \ldots, R_6$ with different order. The first tree is built from R_1 to R_6 and the second one is built from R_6 to R_1, so different shapes arise from different orders.

1.1.2 Tries

If $n = 0$, then the trie is empty; if $n = 1$, then the trie is composed of a single node holding the input-string; if $n > 1$, then the trie contains three parts: a root (internal) node used to direct keys to the left or right (when the first bit of the string is 0 or 1, respectively), a left sub-trie of the root for keys whose first bits are 0 and a right sub-trie of the root for keys whose first bits are 1. The two subtrees are constructed recursively as tries (but using subsequent bits successively). For an example; see Figure 1.2.

Remark 3. Tries resolve the problem of high search cost of digital search trees, but the space requirement of tries is larger than the one of digital search trees.

1.1.3 PATRICIA tries

PATRICIA (Practical Algorithm To Retrieve Information Coded In Alphanumeric) is a compact representation of a trie, where any node which is an only child is

$R_1 = 000001\cdots$
$R_2 = 000110\cdots$
$R_3 = 110111\cdots$
$R_4 = 011011\cdots$
$R_5 = 100001\cdots$
$R_6 = 111110\cdots$

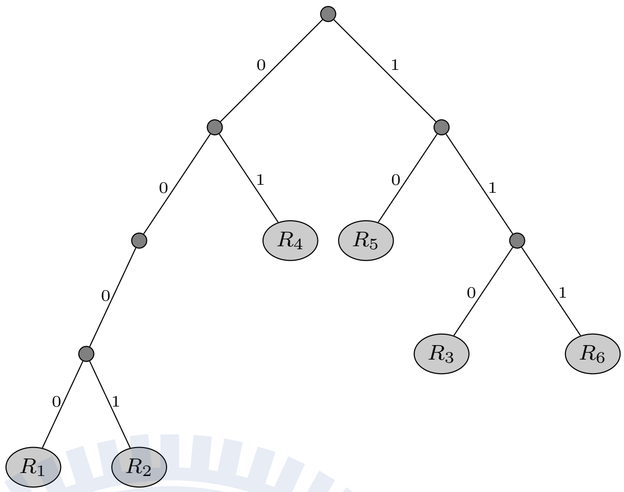


Figure 1.2: A trie built from the keys $R_1, \ldots, R_6$. The small circles represent the internal nodes while the big circles are the external nodes which store the keys. Note that the form of a trie does not depend on the order of keys.

merged with its parent. For an example; see Figure 1.3.

Remark 4. The large space requirement of tries is resolved by PATRICIA tries because of suppressing the creation of one-way branching. However, one disadvantage of PATRICIA tries is a more complex implementation.

1.2 Random model

In probability theory, a sequence of random variables is called independent and identically distributed (iid) if each random variable has the same probability distribution and all are mutually independent.

We call $X_1, X_2, \ldots$ a random string if $X_1, X_2, \ldots$ is an iid sequence of random variables with $P(X_n = 0) = p$ and $P(X_n = 1) = q := 1-p$. We also call a digital tree a random digital tree of size n if it is constructed from n infinite $\{0, 1\}$-random strings. A simple classification of random digital trees is as follows: symmetric $p = q = \frac{1}{2}$ and asymmetric $p \neq q$. Moreover, in the subsequent results, the later case will be further split according to whether $\log p / \log q$ is rational or not.

Remark 5. The above random model is too simplified for practical purposes. More realistic models have been proposed, but their analysis still remains complicated. Moreover, results for the above simple model normally hold for more general models as well. That is why most research has focused on the above simple model.

$$
\begin{aligned}
R_1 &= 000001\cdots\\
R_2 &= 000110\cdots\\
R_3 &= 110111\cdots\\
R_4 &= 011011\cdots\\
R_5 &= 100001\cdots\\
R_6 &= 111110\cdots
\end{aligned}
$$

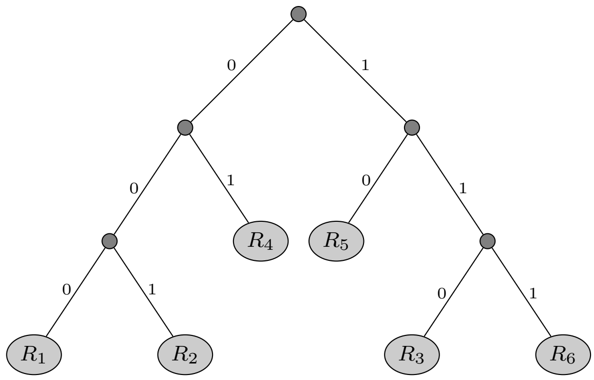


Figure 1.3: A compact represetation of the trie (from Figure 1.2) where any node which is an only child was merged with its parent.

1.3 Size and the number of 2-protected nodes in tries

1.3.1 Size of tries

The size of a digital tree is defined to be the number of internal nodes. For example, it is clear that the size of a random PATRICIA trie which contains n strings is $n - 1$. On the other hand, the size of a random trie is a random variable.

We next give a brief history of the probabilistic analysis of the size of tries, where we mainly focus on asymptotic expansions of moments.

First, the mean was derived by Knuth in 1973; see [12].

As for the variance, it took 15 years until the first analysis was obtained independently by two group of researchers.

The first group consisted of P. Kirschenhofer and H. Prodinger who gave in 1991 a complicated analysis of the symmetric case ($p = q = 1/2$) with explicit expressions for involved constants and periodic functions; see [11].

Independently, P. Jacquet and M. Regnier analysed the general case (symmetric and asymmetric case) but without explicit expressions for involved constants and periodic functions; for their paper which appeared in 1988 see [9]. One year later, they found some partial results on explicit expressions of involved constants and periodic functions; see [15]. Moreover, we should mention that apart from the variance, also the limit law was derived in these two papers.

Recently, M. Fuchs, H. K. Hwang and V. Zacharovas [4] proposed a general framework for obtaining asymptotic expansions of mean and variance of so-called additive shape parameter in random tries with explicit expressions for periodic functions in the general case. Their results apply to the size and we will introduce some of them. First, we need the following notation

$$\mathscr{F}[G](x) := \begin{cases} \dfrac{1}{h} \displaystyle\sum_{k \in \mathbb{Z} \setminus \{0\}} G(-1 + \chi_k) e^{2k\pi i x}, & \text{if } \frac{\log p}{\log q} \in \mathbb{Q}; \\ 0, & \text{if } \frac{\log p}{\log q} \notin \mathbb{Q}, \end{cases}$$

where $\chi_k = \frac{2rk\pi i}{\log p}$ when $\frac{\log p}{\log q} = \frac{r}{l}$ with $(r, l) = 1$.

Theorem 1. *The mean of the size in tries satisfies*

$$\frac{\mathbb{E}(N_n)}{n} = \frac{1}{h} + \mathscr{F}[G_1^{(N)}](r \log_{1/p} n) + o(1), \tag{1.1}$$

where $G_1^{(N)}(s) = -(s+1)\Gamma(s)$ *and* $h = -p \log p - q \log q$.

Theorem 2. *The variance of the size in tries satisfies*

$$\frac{\mathbb{V}(N_n)}{n} = \frac{G_2^{(N)}(-1)}{h} + \mathscr{F}[G_2^{(N)}](r \log_{1/p} n) + o(1),$$

where

$$\begin{aligned} G_2^{(N)}(-1) =& \frac{1}{2} - \frac{1}{h} + 2 \sum_{j \geq 2} \frac{(-1)^j (p^j + q^j)}{1 - p^j - q^j} \\ & - \begin{cases} \dfrac{1}{h \log p} \displaystyle\sum_{j \geq 1} \frac{4rj\pi^2}{\sinh \frac{2rj\pi^2}{\log p}}, & \text{if } \frac{\log p}{\log q} \in \mathbb{Q}; \\ 0, & \text{if } \frac{\log p}{\log q} \notin \mathbb{Q} \end{cases} \end{aligned}$$

and for $k \neq 0$ *(when* $\frac{\log p}{\log q} \in \mathbb{Q}$*)*

$$\begin{aligned} G_2^{(N)}(-1 + \chi_k) =& \chi_k \Gamma(-1 + \chi_k) \left(1 - \frac{\chi_k + 3}{2^{1 + \chi_k}}\right) - \frac{1}{h} \sum_{j \in \mathbb{Z}} \Gamma(\chi_j + 1) \Gamma(\chi_{k-j} + 1) \\ & - 2 \sum_{j \geq 1} \frac{(-1)^j (j + 1 + \chi_k) \Gamma(j + \chi_k)(p^{j+1} + q^{j+1})}{(j+1)!(j+1)(1 - p^{j+1} - q^{j+1})}. \end{aligned}$$

Remark 6. From the above results, we see that in the symmetric case ($p = q = 1/2$), the mean of the size is about $1.443n$ and the variance is about $0.846n$. Moreover, in both cases the periodic function has a very small amplitude.

1.3.2 Number of 2-protected nodes in tries

We start with the definition of 2-protected nodes in trees: they are nodes that have distance at least 2 from each leaf. In 2012, J. Gaither, Y. Homma, M. Sellke and M. D. Ward [9] derived an asymptotic expansion of the mean of the number of 2-protected nodes in random tries. Soon after, J. Gaither and M. D. Ward [15] derived an asymptotic expansion of the variance as well.

Theorem 3. *The mean of the number of 2-protected nodes in tries satisfies*

$$\frac{\mathbb{E}(X_n^{(T)})}{n} = \frac{pq+1}{h} - 1 + \mathscr{F}[G_1^{(T)}](r\log_{1/p} n) + o(1),$$

where $G_1^{(T)}(z)$ is a 1-periodic function.

Remark 7. In [9], the above result was stated as

$$\frac{\mathbb{E}(X_n^{(T)})}{n} = \frac{pq+1}{h} - 1 + \delta(\log n) + O(1),$$

where $\delta(\log n)$ is a small periodic function (possibly constant). Note that the error term is wrong and was corrected above.

Theorem 4. *The variance of the number of 2-protected nodes in tries satisfies*

$$\frac{\mathbb{V}(X_n^{(T)})}{n} = c_1 + c_2 - c_3^2 + c_4 + \mathscr{F}[G_2^{(T)}](r\log_{1/p} n) + o(1),$$

where $G_2^{(T)}(z)$ is a 1-periodic function, c_4 is a constant (which is 0 when $\frac{\log p}{\log q} \notin \mathbb{Q}$) and c_1, c_2 and c_3 are given by

$$\begin{aligned}
c_1 =& \frac{1}{h}\left(\frac{2p^3q(2p^2-2pq+5p+3)}{(p+1)^3} + \frac{2pq^3(2q^2-2pq+5q+3)}{(q+1)^3} + \frac{pq}{2} - \frac{p^2q^2}{4}\right. \\
&\left. - \frac{2p}{p+1} - \frac{2q}{q+1} + \frac{1}{2} + h - 2pq\left(1 - \frac{p}{(p+1)^2} - \frac{q}{(q+1)^2}\right)\right), \\
c_2 =& \frac{2}{h}\sum_{j\geq 2}(-1)^j \frac{(p^j+q^j)^2}{1-p^j-q^j}\left(pq+1-p^2q^2(j-1)j\right), \\
c_3 =& \frac{pq+1}{h} - 1.
\end{aligned}$$

Remark 8. Again the above result was wrongly stated in [15] as

$$\frac{\mathbb{V}(X_n^{(T)})}{n} = c_1 + c_2 - c_3^2 + \delta_1(\log n) - 2c_2\delta_2(\log n) - (\delta_2(\log n))^2 + O(n^{-\varepsilon}).$$

where $\delta_1(\log n)$ and $\delta_2(\log n)$ are periodic functions with average value zero. Apart from the wrong error term, the authors also forgot to pull out the (non-zero) average value from $-(\delta_2(\log n))^2$.

Remark 9. From the above results, we obtain that in symmetric case ($p = q = 1/2$), the mean of the number of 2-protected nodes in tries is about $0.803n$ and the variance is about $0.934n$.

1.4 Additive shape parameters and main idea of the analysis

In this thesis, our main goal is the analysis of the number of 2-protected nodes in tries and PATRICIA tries. In fact, 2-protected nodes are just a special case of so-called additive shape parameters. In [4], the authors proposed a general method for the analysis of these shape parameters. We will explain the main ideas of this method in this section. We start with a precise definition of additive shape parameters.

Additive shape parameters of tries are parameters which can be computed recursively by computing the shape parameter for the subtrees, adding them up and adding a cost. More precisely, consider a random trie built from n infinite $\{0, 1\}$-strings and denote by X_n the additive shape parameter. Then, by splitting the trie (see Figure 1.4), X_n can be described by

$$X_n \stackrel{d}{=} X_{I_n} + X^*_{n-I_n} + T_n, \qquad (n \geq 2), \tag{1.2}$$

where X_n has the same distribution as X^*_n, T_n is some fixed sequence of random variables representing the cost, (X_n), (X^*_n) and (I_n, T_n) are independent and I_n is the size of the left subtree. Note that by the definition of the trie

$$\pi_{n,k} = P(I_n = k) = \binom{n}{k} p^k q^{n-k},$$

or in other words, I_n has a binomial distribution with parameter n and p.

The method in [4] was devised for deriving asymptotic expansions of the moments of X_n. The method proceeds in five steps. The first step is to take moments on both sides of the distributional recurrence for X_n. This gives a recurrence of the following type

$$a_n = \sum_{0 \leq k \leq n} \pi_{n,k}(a_k + a_{n-k}) + b_n$$

for all moments, where b_n is a function of moments of lower order.

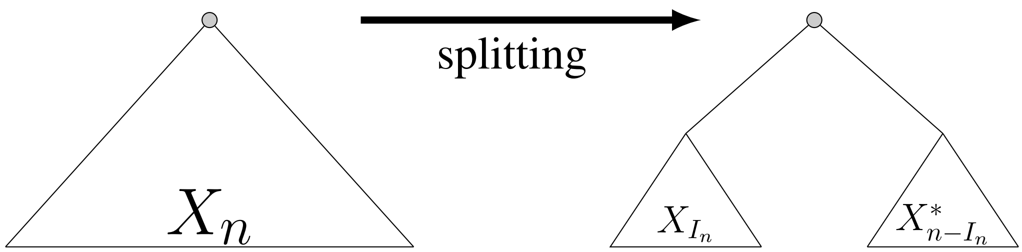


Figure 1.4: Splitting the tree into three parts: root, left sub-tree and right sub-tree.

The second step is poissonization. For this step, define

$$\tilde{f}(z) := e^{-z}\sum_n a_n \frac{z^n}{n!}, \qquad \tilde{g}(z) := e^{-z}\sum_n b_n \frac{z^n}{n!}.$$

Note that $\tilde{f}(z)$ can be interpreted as a_n with n replaced by a Poisson random variable N with parameter z, or more precisely,

$$\tilde{f}(z) = \mathbb{E}(a_N).$$

Now, from the following computation

$$\begin{aligned}
\tilde{f}(z) :&= e^{-z}\sum_n \frac{a_n}{n!} z^n \\
&= e^{-z}\sum_n \frac{\sum_{0\le k\le n} \pi_{n,k}(a_k + a_{n-k})}{n!} z^n + e^{-z}\sum_n \frac{b_n}{n!} z^n \\
&= e^{-z}\sum_n \sum_{0\le k\le n} \frac{\binom{n}{k} p^k q^{n-k} a_k}{n!} z^n + e^{-z}\sum_n \sum_{0\le k\le n} \frac{\binom{n}{k} p^k q^{n-k} a_{n-k}}{n!} z^n + \tilde{g}(z) \\
&= e^{-z}\sum_{k\ge 0} \sum_{(n-k)\ge 0} \frac{p^k a_k}{k!} z^k \frac{q^{n-k}}{(n-k)!} z^{n-k} \\
&\quad + e^{-z}\sum_{(n-k)\ge 0} \sum_{k\ge 0} \frac{q^{n-k} a_{n-k}}{(n-k)!} z^{n-k} \frac{p^k}{k!} z^k + \tilde{g}(z) \\
&= e^{-z}\sum_{k\ge 0} \frac{p^k a_k}{k!} z^k \sum_{(n-k)\ge 0} \frac{q^{n-k}}{(n-k)!} z^{n-k} \\
&\quad + e^{-z}\sum_{(n-k)\ge 0} \frac{q^{n-k} a_{n-k}}{(n-k)!} z^{n-k} \sum_{k\ge 0} \frac{p^k}{k!} z^k + \tilde{g}(z)
\end{aligned}$$

$$
\begin{aligned}
&= e^{-z}e^{qz}\sum_{k\geq 0}\frac{p^k a_k}{k!}z^k + e^{-z}e^{pz}\sum_{(n-k)\geq 0}\frac{q^{n-k}a_{n-k}}{(n-k)!}z^{n-k} + \tilde{g}(z)\\
&= e^{-pz}\sum_{k\geq 0}\frac{a_k}{k!}(pz)^k + e^{-qz}\sum_{(n-k)\geq 0}\frac{a_{n-k}}{(n-k)!}(qz)^{n-k} + \tilde{g}(z)\\
&= \tilde{f}(pz) + \tilde{f}(qz) + \tilde{g}(z),
\end{aligned}
$$

we get a functional equation

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

This functional equations describes the moments when n is replaced by N (this is the so-called Poisson model, whereas the original model is called the Bernoulli model).

The third step is doing Mellin transform on both sides of the above functional equation; for an introduction into Mellin transform see Section 2.1. First, recall

$$\mathscr{M}[\tilde{f}(z);s] := \int_0^\infty \tilde{f}(z)z^{s-1}\mathrm{d}s.$$

Then, we obtain

$$
\begin{aligned}
\mathscr{M}[\tilde{f}(z);s] &= \mathscr{M}[\tilde{f}(pz);s] + \mathscr{M}[\tilde{f}(qz);s] + \mathscr{M}[\tilde{g}(z);s]\\
&= p^{-s}\mathscr{M}[\tilde{f}(z);s] + q^{-s}\mathscr{M}[\tilde{f}(z);s] + \mathscr{M}[\tilde{g}(z);s].
\end{aligned}
$$

Solving for $\mathscr{M}[\tilde{f}(z)]$ yields

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

The fourth step is using inverse Mellin transform

$$
\begin{aligned}
\tilde{f}(z) &= \frac{1}{2\pi i}\int_{\uparrow}\mathscr{M}[\widetilde{f}(z);s]z^{-s}ds\\
&= \frac{1}{2\pi i}\int_{\uparrow}\frac{\mathscr{M}[\tilde{g}(z);s]z^{-s}}{1-p^{-s}-q^{-s}}ds,
\end{aligned}
$$

and applying the converse mapping theorem from Section 2.1.2 to get an asymptotic expansion of $\tilde{f}(z)$.

Finally, the fifth and last step is depoissonization which is based on the Poisson heuristic

$$\tilde{f}(n) = \mathbb{E}(a_N) \sim a_n.$$

To make this step precise, we will use Cauchy's integral formula

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

and the saddle-point method; for details see Section 2.2.

Finally, we mention that all the above steps can be merged via

$$a_n = \frac{n!}{2\pi i} \int_{\uparrow} \frac{\mathscr{M}[\tilde{g}(z); s]}{(1 - p^{-s} - q^{-s})\Gamma(n + 1 - s)} ds.$$

This integral is related to Newton sums and therefore Flajolet coined the term Poisson-Mellin-Newton cycle for the whole cycle we just described; see Figure 1.5.

The above five-step procedure is well-suited for obtaining asymptotics expansions of moments. So, for the variance, once could use it in order to first derive asymptotic expansions for mean and second moment and then use the definition of the variance

$$\mathbb{V}(X_n) = \mathbb{E}(X_n^2) - (\mathbb{E}(X_n))^2.$$

Note, however, that for the examples from Section 1.3, this leads to cancellations since mean and variance are both of linear order and thus $(\mathbb{E}(X_n))^2$ is of quadratic order. The cancellations are not easy treatable in many examples as was, e.g., shown by P. Kirschenhofer and H. Prodinger in [11].

A better approach to the variance is carefully defining a *poissonized variance* in the Poisson model which already incorporates the above cancellations. P. Jacquet and M. Regnier used the following

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

as poissonized variance. However, for the parameters from Section 1.3, this still leads to cancellations in Step 5 of the above procedure.

In a recent work by M. Fuchs, H.-K. Hwang and V. Zacharovas [8] it was pointed out that a better choice of the poissonized variance for the examples from Section 1.3 (or more general, for all shape parameters with linear mean and variance) is the following function

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

We will make use of this poissonized variance in our analysis of 2-protected nodes in tries and PATRICIA tries; see Chapter 3 and Chapter 4.

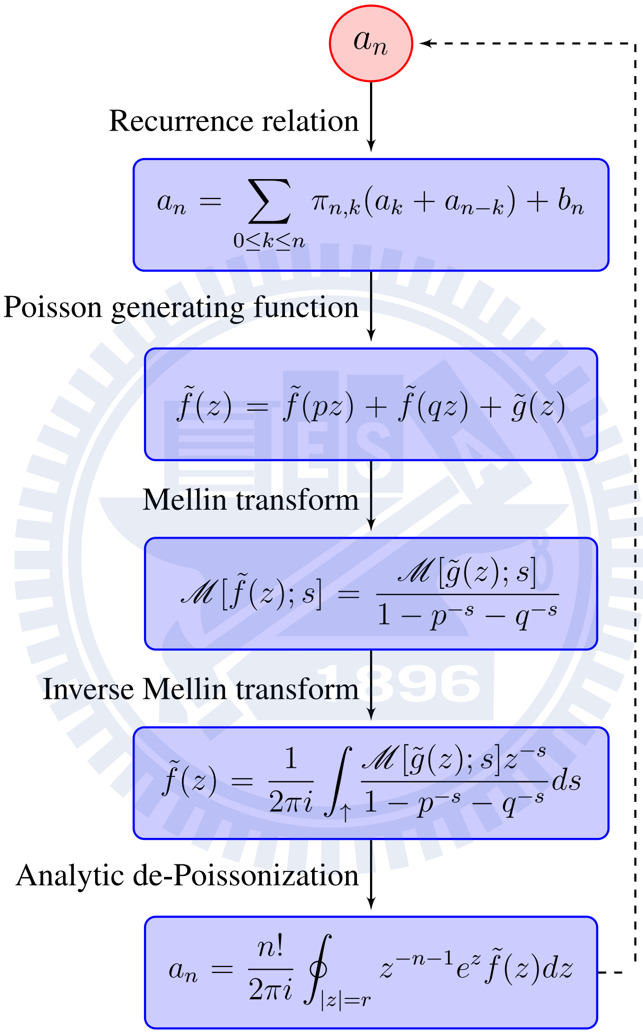


Figure 1.5: Poisson-Mellin-Newton cycle.

Chapter 2

Tools

In the previous chapter, we have presented the methodology which we are going to apply in this thesis. Our methodology consisted of five steps, where we used Mellin transform in the third and fourth step and analytic de-Poissonization in the final step. The next two sections will be concerned with an introduction into these two tools, for more about Mellin transform see [2] and for more about analytic de-Poissonization see [4, 10].

2.1 Mellin transform

Hjalmar Mellin (1854-1933) was a Finnish mathematician and functional theorist, he gave his name to the *Mellin transform* which is an integral transform that may be regarded as the multiplicative version of the two-sided Laplace transform.

Mellin transform is often used in number theory and the theory of asymptotic expansions, it is closely related to the Laplace transform and the Fourier transform, and the theory of the gamma function and other applied special functions.

In applications, the Mellin transform is widely used in computer science, especially, in the analysis of algorithms because of its scale invariance property and inversion theorem which will be introduced below.

2.1.1 Mellin transform and its inverse transform

Informally, the Mellin transform of a function $f(x)$ is

$$\mathscr{M}[f(x);s] = f^*(s) := \int_0^{+\infty} f(x)x^{s-1}dx.$$

The domain of the Mellin transform turns out to be a vertical strip in the complex plane. Moreover, the Mellin transform has an inverse

$$f(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} f^*(s)x^{-s}ds.$$

In the sequel, we will use the notation $\langle\alpha,\beta\rangle$ for the *open strip* of complex numbers $s=\sigma+it$ such that $\alpha<\sigma<\beta$.

We now make the above precise.

Definition 1. *Let $f(x)$ be locally Lebesgue integrable over $(0,+\infty)$. The Mellin transform of $f(x)$ is defined by*

$$\mathscr{M}[f(x);s] = f^*(s) := \int_0^{+\infty} f(x)x^{s-1}dx.$$

The largest open strip $\langle\alpha,\beta\rangle$ in which the integral converges is called the fundamental strip.

The domain of the Mellin transform can be found with the following lemma.

Lemma 1. *If*

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

with $u>v$, then $f^*(s)$ exists in the strip $\langle -u,-v\rangle$.

Proof. Because of the decomposition $\int_0^\infty = \int_0^1 + \int_1^\infty$ and from the assumptions $f(x)x^{s-1} \overset{x\to 0^+}{=} O(x^{u+\sigma-1})$ and $f(x)x^{s-1} \overset{x\to +\infty}{=} O(x^{v+\sigma-1})$, so, in order that both integrals $\int_0^1$ and $\int_1^\infty$ converge, $u+\sigma-1$ must be greater than -1 and $u+\sigma-1$ must be less than -1. Thus, $-u<\sigma<-v$. ∎

Example 1. *The function $f(x)=e^{-x}$ has Mellin transform*

$$\mathscr{M}[f(x);s] = \int_0^{+\infty} e^{-x}x^{s-1}dx = \Gamma(s). \tag{2.1}$$

Moreover, we know that $e^{-x} \overset{x\to 0^+}{=} 1$ and $e^{-x} \overset{x\to +\infty}{=} O(x^{-b})$ for any positive number b. So, the fundamental strip of the Mellin transform is $\langle 0,+\infty\rangle$.

Next, we recall some basic properties of the Mellin transform.

Theorem 5. *(Functional properties). Let $f(x)$ be a function whose transform admits the fundamental strip $\langle \alpha, \beta \rangle$. Let ρ be a nonzero real number and μ, ν be positive real numbers. Then we have the following relations:*

$$
\begin{aligned}
&F_1 : \mathscr{M}[f(\mu x); s] = \mu^{-s} f^*(s), \qquad s \in \langle \alpha, \beta \rangle; \\
&F_2 : \mathscr{M}[x^{\nu} f(x); s] = f^*(s+\nu), \qquad s \in \langle \alpha - \nu, \beta - \nu \rangle; \\
&F_3 : \mathscr{M}[f(x^{\rho}); s] = \frac{1}{\rho} f^*(\frac{s}{\rho}), \qquad s \in \langle \rho\alpha, \rho\beta \rangle; \\
&F_4 : \mathscr{M}[f(x)\log x; s] = \frac{d}{ds} f^*(s), \qquad s \in \langle \alpha, \beta \rangle; \\
&F_5 : \mathscr{M}[\frac{d}{dx} f(x); s] = -(s-1) f^*(s-1), \qquad s \in \langle \alpha^* + 1, \beta^* + 1 \rangle.
\end{aligned}
$$

Remark 10. By linearity of the transform and F_1, we also have

$$\mathscr{M}[\sum_k \lambda_k g(\mu_k x); s] = \left(\sum_k \frac{\lambda_k}{\mu_k^s} \right) g^*(s),$$

whenever k ranges over a finite set of indices. Moreover, this formula can usually be extended to infinite sums with the dominating convergence theorem.

Example 2. *Consider the function*

$$f(x) = \frac{e^{-x}}{1 - e^{-x}} = e^{-x} + e^{-2x} + e^{-3x} + \cdots$$

which is of the above form with $g(x) := e^{-x}$. Thus,

$$
\begin{aligned}
f^*(s) &= \int_0^{+\infty} \sum_{k=1}^{\infty} g(kx) x^{s-1} dx \\
&= \left(\sum_{k=1}^{\infty} k^{-s} \right) \left(\int_0^{+\infty} g(x) x^{s-1} dx \right) \\
&= \zeta(s)\Gamma(s).
\end{aligned}
$$

Finally, we give the inversion theorem for the Mellin transform.

Theorem 6. *(Inversion).*
(i) Let $f(x)$ be integrable with fundamental strip $\langle \alpha, \beta \rangle$. If c is between α and β and $f^*(c+it)$ is integrable, then the equality

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} f^*(s) x^{-s} ds = f(x)$$

hold almost everywhere. Moreover, if $f(x)$ is continuous, then the equality holds everywhere on $(0, +\infty)$.
(ii) Let $f(x)$ be locally integrable with fundamental strip $\langle \alpha, \beta \rangle$ and be of bounded variation in a neighborhood of x_0. Then, for any c between α and β,

$$\lim_{T\to\infty} \frac{1}{2\pi i} \int_{c-iT}^{c+iT} f^*(s) x^{-s} ds = \frac{f(x_0^+) + f(x_0^-)}{2}.$$

2.1.2 Direct mapping and converse mapping

There is a very precise correspondence between the asymptotic expansion of a function at 0 (and ∞) and poles of its Mellin transform in a left (resp. right) half-plane.

Before we state this correspondence, we recall some notation. Let $\phi(s)$ be meromorphic with pole at $s = s_0$. Recall that $\phi(s)$ admits near $s = s_0$ a Laurent expansion

$$\phi(s) = \sum_{k\geq -r} c_k (s - s_0)^k, \tag{2.2}$$

where $r > 0$ is the order of the pole. Also recall that $\sum_{-r\leq k\leq -1} c_k(s - s_0)^k$ is called the principal part of the Laurent series.

Definition 2. *(Singular expansion). Let $\phi(s)$ be meromorphic in Ω. A singular expansion E of $\phi(s)$ in Ω is a formal sum of all principal parts of all poles of $\phi(s)$ in Ω. When E is a singular expansion of $\phi(s)$ in Ω, we write*

$$\phi(s) \asymp E \quad (s \in \Omega).$$

Example 3. *For instance, we have*

$$\frac{1}{(s-1)(s-2)^2} \asymp \left[\frac{1}{(s-1)} + 2 + 3(s-1) + 4(s-1)^2 + \cdots\right]_{s=1} + \left[\frac{1}{(s-2)^2} - \frac{1}{(s-2)} + 1 - (s-2) + \cdots\right]_{s=2}, \qquad (s \in \mathbb{C}),$$

because

$$\frac{1}{(s-1)(s-2)^2} \overset{s\to 1}{=} \frac{1}{(s-1)} + 2 + 3(s-1) + 4(s-1)^2 + 5(s-1)^3 + \cdots$$

and

$$\frac{1}{(s-1)(s-2)^2} \overset{s\to 2}{=} \frac{1}{(s-2)^2} - \frac{1}{(s-2)} + 1 - (s-2) + (s-2)^2 + \cdots .$$

Example 4. *(The Gamma function). Recall that the Mellin transform of the function* e^{-x} *is the Gamma function*

$$\Gamma(s) = \int_0^{+\infty} e^{-x} x^{s-1} dx,$$

for $\Re s > 0$*. We know that* $\Gamma(s)$ *has poles at the points* $s = -m$ *with* $m \in \mathbb{N} \cup \{0\}$ *and we have*

$$\Gamma(s) \overset{s \to -m}{\sim} \frac{(-1)^m}{m!} \frac{1}{s+m},$$

so that the Gamma function admits

$$\Gamma(s) \asymp \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \frac{1}{s+k} \qquad (s \in \mathbb{C}) \tag{2.3}$$

as singular expansion in $\mathbb{C}$*.*

The function e^x has a Taylor expansion at $x = 0$:

$$e^{-x} = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} x^k. \tag{2.4}$$

There is a striking coincidence of coefficients in the Taylor expansion (2.4) of the original function e^{-x} and in the singular expansion (2.3) of the transform $\Gamma(s)$ expressed by the rule

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

This is in fact a completely general phenomenon.

Theorem 7. *(Direct mapping). Let* $f(x)$ *have a transform* $f^*(s)$ *with nonempty fundamental strip* $\langle \alpha, \beta \rangle$*.*
(i) Assume that $f(x)$ *admits as* $x \to 0^+$ *an asymptotic expansion of the form*

$$f(x) = \sum_{(\zeta,k) \in A} c_{\zeta,k} x^{\zeta} (\log x)^k + O(x^{\gamma}), \tag{2.5}$$

where the ζ*'s satisfy* $-\gamma < -\zeta \leq \alpha$ *and the* k*'s are nonnegative. Then* $f^*(s)$ *is continuable to a meromorphic function in the strip* $\langle -\gamma, \beta \rangle$ *where it admits the singular expansion*

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

(ii) Similarly, assume that $f(x)$ admits as $x \to +\infty$ an asymptotic expansion of the form (2.5) where now $\beta \leq -\zeta < -\gamma$. Then $f^*(s)$ is continuable to a meromorphic to a meromorphic function in the strip $\langle \alpha, -\gamma \rangle$ where it admits the singular expansion

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

Example 5. *The function $f(x) = (1+x)^{-1}$ has $\langle 0, 1 \rangle$ as its fundamental strip. The following two expansion*

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

translate into

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

which is consistent with the known form,

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

Under a set of mild conditions, a converse to the direct mapping theorem also holds: The singularities of a Mellin transform which is small enough towards $\pm i\infty$ encode the asymptotic properties of the original function.

Theorem 8. *(Converse mapping). Let $f(x)$ have a Mellin transform $f^*(s)$ with nonempty fundamental strip $\langle \alpha, \beta \rangle$.*
(i) 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

$$f^*(s) = O(|s|^{-r}) \quad \text{with} \quad r > 1, \tag{2.6}$$

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

$$f^*(s) \asymp \sum_{(\zeta,k)\in A} d_{\zeta,k} \frac{1}{(s-\zeta)^k}, \tag{2.7}$$

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

$$f(x) = \sum_{(\zeta,k)\in A} d_{\zeta,k} \left(\frac{(-1)^{k-1}}{(k-1)!} x^{-\zeta} (\log x)^{k-1} \right) + O(x^{-\gamma}).$$

(ii) 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 (2.6) holds in $\langle \eta, \gamma \rangle$ for some $\eta \in \langle \alpha, \beta \rangle$. If $f^*(s)$ admits the singular expansion (2.7) for $\Re(s) \in \langle \eta, \gamma \rangle$, then an asymptotic expansion of $f(x)$ at ∞ is

$$f(x) = -\sum_{(\zeta,k)\in A} d_{\zeta,k} \left(\frac{(-1)^{k-1}}{(k-1)!} x^{-\zeta} (\log x)^{k-1} \right) + O(x^{-\gamma}).$$

2.1.3 Some other properties

With the converse mapping theorem (Theorem 8), we obtain asymptotic expansions of $f(x)$ at 0 and ∞ under the condition (2.6). Now, we explain that the condition can be modified such that the asymptotic expansions hold even in a suitable region in the complex plane.

Theorem 9. *If we change the condition (2.6) in Theorem 8 into*

$$f^*(\sigma + it) = O(e^{-\theta|t|}) \qquad \textit{with} \qquad -\pi < \theta < \pi,$$

then the asymptotic expansions in Theorem 8 holds in the cone $|\arg(z)| \leq \theta$.

Such a result has the advantage that the asymptotic expansion for derivatives of $f(z)$ can be obtained from that of $f(z)$ by differentiation (see [14]).

Theorem 10 (Ritt)**.** *Let $f(z)$ be an analytic function in the cone $|\arg(z)| \leq \theta$, where $-\pi < \theta < \phi$. Assume that*

$$f(z) = O(z^{\alpha})$$

for z in the cone with $\alpha \in \mathbb{R}$. Then, in the slightly smaller cone $|\arg(z)| \leq \theta - \epsilon$ with $\epsilon > 0$, we have

$$f'(z) = O(z^{\alpha-1}).$$

2.2 Analytic de-Poissonization and JS-admissible functions

In this section, we are going to describe the fifth step from the five-step procedure from Section 1.4. Recall that for this step, we have to find a justification of the

Poisson heuristic

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

For this we need the following definition.

Definition 3. *Let $\tilde{f}(z)$ be an entire function. Then we say that $\tilde{f}(z)$ is JS-admissible and write $\tilde{f} \in \mathscr{J}\mathscr{S}$ (or more precisely, $\tilde{f} \in \mathscr{J}\mathscr{S}_{\alpha,\beta}$, $\alpha, \beta \in \mathbb{R}$) if for some $0 < \theta < \pi/2$ and $|z| \geq 1$ the following two conditions hold.*

(**I**) *(Polynomial growth inside a sector) Uniformly for $|\arg(z)| \leq \theta$*

$$\tilde{f}(z) = O(|z|^{\alpha} \left(\log_{+} |z|^{\beta}\right)), \tag{2.8}$$

where $\log_{+} x := \log(1 + x)$.

(**O**) *(Exponential bound) Uniformly for $\theta \leq |\arg(z)| \leq \pi$*

$$f(z) := e^{z} \tilde{f}(z) = O\left({}^{(1-\varepsilon)|z|}\right), \tag{2.9}$$

for some $\varepsilon > 0$.

Since the conditions of admissibility are strong, we can now indeed justify the Poisson heuristic.

Proposition 1. *If $\tilde{f} \in \mathscr{J}\mathscr{S}_{\alpha,\beta}$, then a_n satisfies the asymptotic expansion*

$$a_n = \sum_{0 \leq j < 2k} \frac{\tilde{f}^{(j)}(n)}{j!} \tau_j(n) + O\left(n^{\alpha - k} \log^{\beta} n\right), \tag{2.10}$$

for $k = 0, 1, \ldots$, where the τ_j's are polynomials of n of degree $\lfloor j/2 \rfloor$ given by

$$\tau_j(n) = \sum_{0 \leq l \leq j} \binom{j}{l} (-n)^l \frac{n!}{(n-j+1)!}, \qquad (j = 0, 1, \ldots).$$

Remark 11. Note that we have the identity

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

for every entire function $\tilde{f}(z)$; see [8]. It is the asymptotic nature (2.10) that requires more regularity conditions.

Due to Proposition 1, in order to carry out the fifth step, we only have to check JS-admissibility. Note, however, that $\tilde{f}(z)$ is only given via the functional equation

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

Checking that $\tilde{f}(z)$ is JS-admissible is reduced to checking the same for $\tilde{g}(z)$ because of the following result (the proof can be found in [4])

Proposition 2. *Let $\tilde{f}$ and $\tilde{g}$ be entire functions satisfying*

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

with $f(0)$ given. Then

$$\tilde{f} \in \mathscr{J}\mathscr{S} \quad \textit{if and only if} \quad \tilde{g} \in \mathscr{J}\mathscr{S}$$

Remark 12. Proposition 2 is a generalization of the following real-valued asymptotic transfer (results of this type are often called master theorem in theoretical computer science) : if

$$\tilde{f}(x) = \tilde{f}(px) + \tilde{f}(qx) + \tilde{g}(x),$$

where

$$\tilde{g}(x) = O(x^{\alpha}(\log_{+} x)^{\beta} + 1)$$

with $\alpha > 0$, then

$$\tilde{f}(x) = \begin{cases} O(x), & \text{if } \alpha < 1; \\ O(x^{\alpha}(\log_{+} x)^{\beta+1}), & \text{if } \alpha = 1; \\ O(x^{\alpha}(\log_{+} x)^{\beta}), & \text{if } \alpha > 1. \end{cases}$$

Finally checking JS-admissibility of $\tilde{g}(z)$ is simple because of the following closure properties (for the proof see [8])

Proposition 3. *Let m be a non-negative integer and $\alpha \in (0, 1)$.*
(i) $z^m, e^{-\alpha z} \in \mathscr{J}\mathscr{S}$.
(ii) If $\tilde{f} \in \mathscr{J}\mathscr{S}$, then $P\tilde{f} \in \mathscr{J}\mathscr{S}$ for any polynomial $P(z)$.
(iii) If $\tilde{f} \in \mathscr{J}\mathscr{S}$, then $\tilde{f}(\alpha z) \in \mathscr{J}\mathscr{S}$.
(iv) If $\tilde{f}, \tilde{g} \in \mathscr{J}\mathscr{S}$, then $\tilde{f} + \tilde{g} \in \mathscr{J}\mathscr{S}$.
(v) If $\tilde{f}, \tilde{g} \in \mathscr{J}\mathscr{S}$, then $\tilde{f}(\alpha z), \tilde{g}((1-\alpha)z) \in \mathscr{J}\mathscr{S}$.
(vi) If $\tilde{f} \in \mathscr{J}\mathscr{S}$, then $\tilde{f}^{(m)} \in \mathscr{J}\mathscr{S}$.

Chapter 3

The number of 2-protected nodes in tries

Let $X_n^{(T)}$ be the number of 2-protected nodes in a random trie built from n records. Then, we have, for $n \geq 2$,

$$X_n^{(T)} \stackrel{d}{=} \begin{cases} X_{n-1}^{(T)}, & \text{when } I_n = 1 \text{ or } I_n = n-1; \\ X_{I_n}^{(T)} + X_{n-I_n}^{(T)*} + 1, & \text{otherwise,} \end{cases} \tag{3.1}$$

where notation is as in Section 1.4 and initial conditions are $X_0^{(T)} = X_1^{(T)} = 0$. Note that this can be rewritten to (1.2) with

$$T_n = \begin{cases} 0, & \text{when } I_n = 1 \text{ or } I_n = n-1; \\ 1, & \text{otherwise.} \end{cases}$$

3.1 The mean of the number of 2-protected nodes in tries

Let a_n be the mean of the number of 2-protected nodes. Then, by taking moments on both sides of (3.1), we obtain

$$a_n = \sum_{0 \leq k \leq n} \binom{n}{k} p^k q^{n-k} (a_k + a_{n-k}) + \left(1 - npq^{n-1} - np^{n-1}q\right), \quad (n \geq 3).$$

Note that if we set

$$b_n = \begin{cases} 0, & \text{when } n = 0 \text{ or } 1; \\ 1 - 2pq, & \text{when } n = 2; \\ 1 - npq^{n-1} - np^{n-1}q, & \text{otherwise,} \end{cases}$$

then, we have

$$a_n = \sum_{0\le k\le n} \pi_{n,k}(a_k + a_{n-k}) + b_n, \qquad (n \ge 0).$$

Next, we denote by $\tilde{f}_1(z)$ and $\tilde{g}_1(z)$ the Poisson generating functions of a_n and b_n. Then,

$$\tilde{f}_1(z) = \tilde{f}_1(pz) + \tilde{f}_1(qz) + \tilde{g}_1(z),$$

where

$$\begin{aligned}
\tilde{g}_1(z) &= e^{-z}\sum_{n\ge 0}\frac{b_n}{n!}z^n \\
&= e^{-z}\sum_{n\ge 3}\frac{1-npq^{n-1}-np^{n-1}q}{n!}z^n + e^{-z}\frac{1-2pq}{2}z^2 \\
&= e^{-z}\left(\sum_{n\ge 3}\frac{1}{n!}z^n - \sum_{n\ge 3}\frac{npq^{n-1}}{n!}z^n - \sum_{n\ge 3}\frac{np^{n-1}q}{n!}z^n + \frac{1-2pq}{2}z^2\right) \\
&= e^{-z}\left(\sum_{n\ge 0}\frac{z^n}{n!} - pz\sum_{(n-1)\ge 0}\frac{(qz)^{n-1}}{(n-1)!} - qz\sum_{(n-1)\ge 0}\frac{(pz)^{n-1}}{(n-1)!} - 1 + pqz^2\right) \\
&= e^{-z}\left(e^z - pze^{qz} - qze^{pz} - 1 + pqz^2\right) \\
&= 1 - e^{-z} + pqz^2e^{-z} - pze^{-pz} - qze^{-qz}.
\end{aligned}$$

The third step is applying Mellin transform. Therefore, observe that from Remark 12, we have

$$\tilde{f}_1(z) = \begin{cases} O(z^{1+\varepsilon}), & \text{as } z\to\infty; \\ O(z^2), & \text{as } z\to 0^+, \end{cases}$$

where $\varepsilon > 0$. Hence the Mellin transform of $\tilde{f}_1(z)$ exists in the strip $\langle -2,-1\rangle$. Applying Mellin transform yields

$$F_1^{(T)}(s) = \mathscr{M}[\tilde{f}_1(z);s] = \frac{G_1^{(T)}(s)}{1-p^{-s}-q^{-s}},$$

where

$$\begin{aligned}
G_1^{(T)}(s) &= \mathscr{M}[\tilde{g}_1(z);s] \\
&= \int_0^{+\infty}(1-e^{-z})z^{s-1}dz + \int_0^{+\infty} pqe^{-z}z^{s+1}dz \\
&\quad - \int_0^{+\infty} pe^{-pz}z^s dz - \int_0^{+\infty} pe^{-pz}z^s dz
\end{aligned}$$

$$= -\Gamma(s) + pq\Gamma(s+2) - p^{-s}\Gamma(s+1) - q^{-s}\Gamma(s+1)$$
$$= \Gamma(s)\left(-1 + pqs(s+1) - p^{-s}s - q^{-s}s\right).$$

Because

$$\tilde{g}_1(z) \overset{z\to 0^+}{=} O(z^2) \quad \text{and} \quad \tilde{g}_1(z) \overset{z\to +\infty}{=} O(z^0),$$

we get that $G_1^{(T)}(s)$ has fundamental strip $\langle -2, 0\rangle$. Plugging the expression for $G_1^{(T)}(s)$ into the one for $F_1^{(T)}(s)$, we obtain

$$F_1^{(T)}(s) = \frac{G_1^{(T)}(s)}{1 - p^{-s} - q^{-s}} = \frac{\Gamma(s)\left(-1 + pqs(s+1) - p^{-s}s - q^{-s}s\right)}{1 - p^{-s} - q^{-s}}. \tag{3.2}$$

In order to apply inverse Mellin transform, we have to understand the singularities of $F_1^{(T)}(s)$. Therefore, we need the following result on the zeros of $1 - p^{-s} - q^{-s}$; see [3].

Lemma 3.1.1. *The roots of* $1 - p^{-s} - q^{-s}$ *are all simple and have the following properties.*

(i) There are no roots with $\Re(s) < -1$.

(ii) On the line $\Re(s) = -1$, *the set of all roots is given by*

$$\begin{cases} \{-1\}, & \text{if } \frac{\log p}{\log q} \notin \mathbb{Q}; \\ \{-1 + \chi_k \ : \ k \in \mathbb{Z}\}, & \text{if } \frac{\log p}{\log q} \in \mathbb{Q}. \end{cases}$$

(iii) If $\Re(s) > -1$, *then the roots are uniformly separated, i.e., there exists an* $\epsilon > 0$ *such that for all roots* s_1 *and* s_2, *we have that* $|s_1 - s_2| \geq \varepsilon$. *Moreover,* $1-p^{-s}-q^{-s}$ *is uniformly bounded away from zero for all* s *having a distance of at least* $\varepsilon/2$ *to any root.*

Now, recall that we know that $G_1^{(T)}(s)$ exists in the strip $\langle -2, 0\rangle$. By the above lemma, all zeros of $1 - p^{-s} - q^{-s}$ generate simple poles and thus $F_1^{(T)}(s)$ is meromorphic in $\langle -2, -1+\varepsilon\rangle$.

Let $\mathcal{A} = \{\lambda_1, \lambda_2, \ldots\}$ be the simple poles of $F_1^{(T)}(s)$ in $-1 \leq \Re(s) < -1+\varepsilon$. We know that all poles come from zeros of $1 - p^{-s} - q^{-s}$. In order to compute the residue, we let

$$\frac{1}{1 - p^{-s} - q^{-s}} = \frac{a}{s - \lambda_k} + b + c(s - \lambda_k) + \cdots,$$

where

$$a = \lim_{s\to\lambda_k} \frac{s-\lambda_k}{1-p^{-s}-q^{-s}} = \lim_{s\to\lambda_k} \frac{1}{p^{-s}\log p + q^{-s}\log q}$$
$$= \frac{1}{p^{-\lambda_k}\log p + q^{-\lambda_k}\log q}.$$

Moreover, we have

$$G_1^{(T)}(s) = G_1^{(T)}(\lambda_k) + G_1^{(T)'}(\lambda_k)(s-\lambda_k) + \cdots.$$

This gives the following singularity expansion of $F_1^{(T)}(s)$

$$F_1^{(T)}(s) \asymp \sum_{\mathcal{A}} \frac{G_1^{(T)}(\lambda_k)}{p^{-\lambda_k}\log p + q^{-\lambda_k}\log q}(s-\lambda_k)^{-1}.$$

Now, we can apply Theorem 8. Note that assumption (2.6) of this theorem is satisfied due to the following well-known decay property of the Gamma function

$$|\Gamma(\sigma+it)| \sim \sqrt{2\pi}|t|^{\sigma-1/2}e^{-\pi|t|/2}. \tag{3.3}$$

Applying Theorem 8 yields

$$\tilde{f}_1(z) = -\sum_{\mathcal{A}} \frac{G_1^{(T)}(\lambda_k)}{p^{-\lambda_k}\log p + q^{-\lambda_k}\log q} z^{-\lambda_k} + O(z^{1-\varepsilon}).$$

We let $\mathcal{B} \subseteq \mathcal{A}$ be the poles with $\Re(s) = -1$. Then,

$$\tilde{f}_1(z) = -\sum_{\mathcal{B}} \frac{G_1^{(T)}(\lambda_k)}{p^{-\lambda_k}\log p + q^{-\lambda_k}\log q} z^{-\lambda_k}$$
$$-\sum_{\mathcal{A}\backslash\mathcal{B}} \frac{G_1^{(T)}(\lambda_k^*)}{p^{-\lambda_k^*}\log p + q^{-\lambda_k^*}\log q} z^{-\lambda_k^*} + O(z^{1-\varepsilon})$$

Because $-\Re(\lambda_k^*) < 1$, we obtain $|z^{-\lambda_k^*}| = o(z)$. Thus, by the dominated convergence theorem

$$\tilde{f}_1(z) = -\sum_{\mathcal{B}} \frac{G_1^{(T)}(\lambda_k)}{p^{-\lambda_k}\log p + q^{-\lambda_k}\log q} z^{-\lambda_k} + o(z)$$
$$= \frac{-G_1^{(T)}(-1)}{p\log p + q\log q} z$$
$$-\sum_{\mathcal{B}\backslash\{-1\}} \frac{G_1^{(T)}(\lambda_k)}{p^{-\lambda_k}\log p + q^{-\lambda_k}\log q} z^{-\lambda_k} + o(z), \tag{3.4}$$

where

$$G_1^{(T)}(-1) = \lim_{s\to -1} \Gamma(s)\left(-1 + pqs(s+1) - p^{-s}s - q^{-s}s\right)$$
$$= pq + 1 - h$$

with $h = -p\log p - q\log q$. Observe that due to (3.3), (3.4) holds also in a cone as explained in Section 2.1.3.

Finally, we apply depoissonization. Therefore, note that by the explicit form of $\tilde{g}_1(z)$ and Proposition 3, we obtain that $\tilde{g}_1(z) \in \mathscr{J}\mathscr{S}$ and by Proposition 2, $\tilde{f}_1(z) \in \mathscr{J}\mathscr{S}$ as well. Thus, from Proposition 1, we get the same result as already stated in Section 1.3.2,

Theorem 11. *The mean of the number of 2-protected nodes in tries satisfies*

$$\frac{\mathbb{E}(X_n^{(T)})}{n} = \frac{pq+1-h}{h} + \mathscr{F}[G_1^{(T)}](r\log_{1/p} n) + o(1), \tag{3.5}$$

where for $k \neq 0$ (when $\frac{\log p}{\log q} \in \mathbb{Q}$)

$$G_1^{(T)}(-1+\chi_k) = \Gamma(-1+\chi_k)\chi_k(\chi_k pq - pq - 1).$$

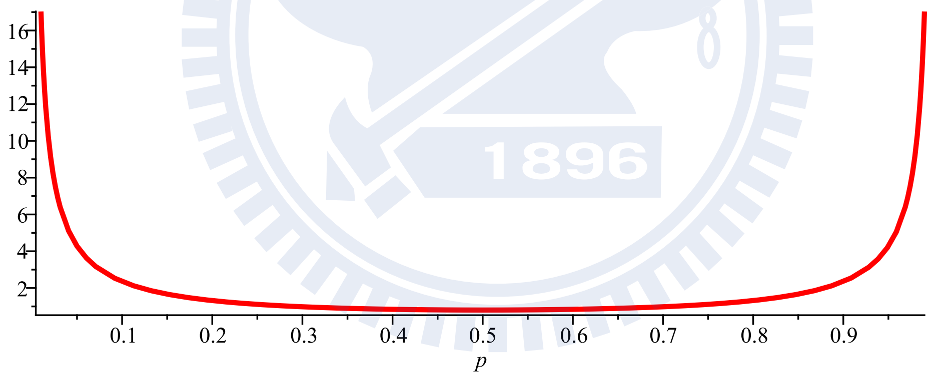


Figure 3.1: A plot of $(pq+1-h)/h$ which is the average value of the main term of $\mathbb{E}(X_n^{(T)})/n$.

Remark 13. Comparing with Theorem 3, note that we also obtained an explicit expression of the periodic function. This periodic function can be used to compute c_4 in Theorem 4. More precisely, the authors in [7] showed that c_4 is the 0-th Fourier coefficient of $\left(\mathscr{F}[G_1^{(T)}](r\log_{1/p} n)'\right)^2$. Thus, our expression above yields

$$c_4 = -\frac{2}{h^2}\sum_{j\geq 1} \frac{\frac{2rj\pi^2}{\log p}}{\sinh\left(\frac{2rj\pi^2}{\log p}\right)}\left(p^2q^2\left(\frac{2rj\pi}{\log p}\right)^2 + p^2q^2 + 2pq + 1\right).$$

Using Maple, we obtain for $p = q = 1/2$

$$c_4 \approx -6.810408446854064413349608724620905082923520900836 8 \cdot 10^{-10}$$

which is very small.

Remark 14. For a plot of the average value of $\mathbb{E}(X_n^{(T)})/n$ see Figure 3.1. Note that the average value becomes minimal if $p = q = 1/2$. This is intuitively clear because in the asymmetric case nodes with one-way branching (which are all 2-protected) become very likely.

3.2 The variance of the number of 2-protected nodes in tries

We start by considering the second moment of $X_n^{(T)}$. Therefore, we square and take moments on the both sides of (3.1). This yields

$$\begin{aligned}\mathbb{E}((X_n^{(T)})^2) =& \sum_{k=0}^{n} P(I_n = k)\Big(\mathbb{E}(X_{I_n}^{(T)}) + \mathbb{E}(X_{n-I_n}^{(T)}) + \mathbb{E}(T_n|I_n = k)\Big)^2 \\ =& \sum_{k=0}^{n} \pi_{n,k}\big(\mathbb{E}((X_k^{(T)})^2) + \mathbb{E}((X_{n-k}^{(T)})^2)\big) + 1 - npq^{n-1} - np^{n-1}q \\ &+ 2\sum_{k=0}^{n} \pi_{n,k}\mathbb{E}(X_k^{(T)})\mathbb{E}(X_{n-k}^{(T)}) \\ &+ 2\sum_{k=0}^{n} \pi_{n,k}\big(\mathbb{E}(X_k^{(T)}) + \mathbb{E}(X_{n-k}^{(T)})\big)\mathbb{E}(T_n|I_n = k), \quad (n \geq 3).\end{aligned} \tag{3.6}$$

Next, we let $s_n := \mathbb{E}((X_n^{(T)})^2)$. Then,

$$\begin{aligned}s_n =& \sum_{k=0}^{n} \pi_{n,k}(s_k + s_{n-k}) + b_n + 2\sum_{k=0}^{n} \pi_{n,k}a_k a_{n-k} \\ &+ 2\sum_{k=0}^{n} \pi_{n,k}(a_k + a_{n-k})\mathbb{E}(T_n|I_n = k), \quad (n \geq 0).\end{aligned} \tag{3.7}$$

As before, the next step is poissonization. Let $\tilde{f}_2(z)$ be the Poisson generating functions of s_n. Then,

$$\tilde{f}_2(z) = \tilde{f}_2(pz) + \tilde{f}_2(qz) + \tilde{g}_1(z) + 2\tilde{f}_1(pz)\tilde{f}_1(qz) + \tilde{h}_2(z),$$

where $\tilde{h}_2(z)$ is equal to

$$
\begin{aligned}
&2e^{-z}\sum_{n\geq 0}\sum_{0\leq k\leq n}\pi_{n,k}(a_k+a_{n-k})\mathbb{E}(T_n|I_n=k)\frac{z^n}{n!}\\
=&2e^{-z}\sum_{n\geq 3}\left(\sum_{0\leq k\leq n}\pi_{n,k}(a_k+a_{n-k})-\pi_{n,1}(a_1+a_{n-1})-\pi_{n,n-1}(a_{n-1}+a_1)\right)\frac{z^n}{n!}\\
&+2e^{-z}\sum_{0\leq k\leq 2}\pi_{2,k}(a_k+a_{2-k})\mathbb{E}(T_n|I_2=k)\frac{z^2}{2!}\\
=&2e^{-z}\sum_{n\geq 3}\sum_{0\leq k\leq n}\pi_{n,k}(a_k+a_{n-k})\frac{z^n}{n!}-2e^{-z}\sum_{n\geq 3}(\pi_{n,1}+\pi_{n,n-1})(a_1+a_{n-1})\frac{z^n}{n!}\\
&+2e^{-z}(\pi_{2,0}+\pi_{2,2})a_2\frac{z^2}{2!}\\
=&2e^{-z}\sum_{n\geq 2}\sum_{0\leq k\leq n}\binom{n}{k}p^kq^{n-k}(a_k+a_{n-k})\frac{z^n}{n!}\\
&-2e^{-z}\sum_{n\geq 3}(npq^{n-1}+np^{n-1}q)(a_1+a_{n-1})\frac{z^n}{n!}-2e^{-z}\pi_{2,1}(a_1+a_1)\frac{z^2}{2!}\\
=&2e^{-z}\sum_{n\geq 0}\sum_{0\leq k\leq n}p^kq^{n-k}(a_k+a_{n-k})\frac{z^n}{k!(n-k)!}\\
&-2e^{-z}\sum_{n\geq 3}(pq^{n-1}+p^{n-1}q)a_{n-1}\frac{z^n}{(n-1)!}\\
=&2e^{-z}\sum_{n\geq 0}\sum_{0\leq k\leq n}a_k\frac{(pz)^k}{k!}\frac{(qz)^{n-k}}{(n-k)!}+2e^{-z}\sum_{n\geq 0}\sum_{0\leq k\leq n}a_{n-k}\frac{(pz)^k}{k!}\frac{(qz)^{n-k}}{(n-k)!}\\
&-2e^{-z}\sum_{n-1\geq 2}pza_{n-1}\frac{(qz)^{n-1}}{(n-1)!}-2e^{-z}\sum_{n-1\geq 2}qza_{n-1}\frac{(pz)^{n-1}}{(n-1)!}\\
=&2e^{-z}\sum_{k\geq 0}a_k\frac{(pz)^k}{k!}\sum_{n\geq k}\frac{(qz)^{n-k}}{(n-k)!}+2e^{-z}\sum_{k\geq 0}a_k\frac{(qz)^k}{k!}\sum_{n\geq k}\frac{(pz)^{n-k}}{(n-k)!}\\
&-2e^{-z}pz\sum_{n-1\geq 0}a_{n-1}\frac{(qz)^{n-1}}{(n-1)!}-2e^{-z}qz\sum_{n-1\geq 0}a_{n-1}\frac{(pz)^{n-1}}{(n-1)!}\\
=&2\tilde{f}_1(pz)+2\tilde{f}_1(qz)-2pze^{-pz}\tilde{f}_1(qz)-2qze^{-qz}\tilde{f}_1(pz).
\end{aligned}
$$

Recall the poissonized variance from Section 2.2:

$$\tilde{V}_X^{(T)}(z):=\tilde{f}_2(z)-\tilde{f}_1(z)^2-z\tilde{f}_2'(z)^2$$

which satisfies the functional equation

$$\tilde{V}_X^{(T)}(z) = \tilde{V}_X^{(T)}(pz) + \tilde{V}_X^{(T)}(qz) + \tilde{V}_T(z) + \tilde{\phi}_1(z) + \tilde{\phi}_2(z), \tag{3.8}$$

where $\tilde{V}_T(z)$ is given by

$$\begin{aligned}
&\tilde{g}_1(z) - \tilde{g}_1(z)^2 - z\tilde{g}_1'(z)^2\\
=&\tilde{g}_1(z)\left(1-\tilde{g}_1(z)\right) - z\tilde{g}_1'(z)^2\\
=&\left(1-e^{-z}+pqz^2e^{-z}-pze^{-pz}-qze^{-qz}\right)\left(e^{-z}-pqz^2e^{-z}+pze^{-pz}+qze^{-qz}\right)\\
&-z\left(e^{-z}+2zpqe^{-z}-pqz^2e^{-z}-pe^{-pz}+p^2ze^{-pz}-qe^{-qz}+q^2ze^{-qz}\right)^2
\end{aligned}$$

and $\tilde{\phi}_1(z)$ is defined by

$$\begin{aligned}
&\tilde{h}_2(z) - 2\tilde{g}_1(z)\left(\tilde{f}_1(pz)+\tilde{f}_1(qz)\right) - 2z\tilde{g}_1'(z)\left(p\tilde{f}_1'(pz)+q\tilde{f}_1'(qz)\right)\\
=&2\tilde{f}_1(pz)+2\tilde{f}_1(qz)-2pze^{-pz}\tilde{f}_1(qz)-2qze^{-qz}\tilde{f}_1(pz)\\
&-2\left(1-e^{-z}+pqz^2e^{-z}-pze^{-pz}-qze^{-qz}\right)\left(\tilde{f}_1(pz)+\tilde{f}_1(qz)\right)\\
&-2z\left(e^{-z}+2pqze^{-z}-pqz^2e^{-z}-pe^{-pz}+p^2ze^{-pz}-qe^{-qz}+q^2ze^{-qz}\right)\\
&\quad\cdot\left(p\tilde{f}_1'(pz)+q\tilde{f}_1'(qz)\right)\\
=&2(e^{-z}-pqz^2e^{-z}+pze^{-pz})\tilde{f}_1(pz)+2(e^{-z}-pqz^2e^{-z}+qze^{-qz})\tilde{f}_1(qz)\\
&-2z\left(e^{-z}+2pqze^{-z}-pqz^2e^{-z}-pe^{-pz}+p^2ze^{-pz}-qe^{-qz}+q^2ze^{-qz}\right)\\
&\quad\cdot\left(p\tilde{f}_1'(pz)+q\tilde{f}_1'(qz)\right).
\end{aligned}$$

Finally,

$$\tilde{\phi}_2(z) := pqz\big(\tilde{f}_1'(pz)-\tilde{f}_1'(qz)\big)^2.$$

Before we apply Mellin transform observe that from (3.4) and Ritt's theorem,

$$\tilde{V}_T(z)+\tilde{\phi}_1(z)+\tilde{\phi}_2(z) = O(z), \text{ as } z\to\infty.$$

Thus, by Remark 12, the Mellin transform of $\tilde{V}_X^{(T)}(z)$ exists in the strip $\langle -2,-1\rangle$. Now, applying Mellin transform gives

$$\left(1-p^{-s}-q^{-s}\right)\mathscr{M}[\tilde{V}_X^{(T)}(z);s] = \mathscr{M}[\tilde{V}_T(z);s] + \mathscr{M}[\tilde{\phi}_1(z);s] + \mathscr{M}[\tilde{\phi}_2(z);s].$$

We denote by $G_2^{(T)}(s) = \Phi(s)+\Phi_1(s)+\Phi_2(s)$, where $\Phi(s) = \mathscr{M}[\tilde{V}_T(z);s]$, $\Phi_1(s) = \mathscr{M}[\tilde{\phi}_1(z);s]$ and $\Phi_2(s) = \mathscr{M}[\tilde{\phi}_2(z);s]$. Then,

$$F_2^{(T)}(s) = \frac{G_2^{(T)}(s)}{1-p^{-s}-q^{-s}},$$

where

$$F_2^{(T)}(s) := \mathscr{M}[\tilde{V}_X^{(T)}(z); s].$$

We next discuss the fundamental strip of $G_2^{(T)}(s)$. First observe that

$$\tilde{V}_T(z), \tilde{\phi}_1(z) = \begin{cases} O(z^{-b}), & \text{as } z \to \infty; \\ O(z^2), & \text{as } z \to 0^+. \end{cases}$$

for arbitrary $b > 0$. Thus the fundamental strip of $\Phi(s)$ and $\Phi_1(s)$ is $\langle -2, \infty \rangle$. As for $\Phi_2(s)$, it was proved in [4] that this function has fundamental strip $\langle -2, -1 \rangle$ and is analytic on the line $\Re(\sigma) = -1$.

Next, we apply again Theorem 8 (that condition (2.6) is satisfied will become clear from the explicit expressions for $G_2^{(T)}(s)$ below). This yields

$$\tilde{V}_X^{(T)}(z) = -\frac{G_2^{(T)}(-1)}{p\log p + q\log q} z - \sum_{\mathcal{B}\setminus\{-1\}} \frac{G_2^{(T)}(\lambda_k)}{p^{-\lambda_k}\log p + q^{-\lambda_k}\log q} z^{-\lambda_k} + o(z).$$

Note that this again holds in some cone.

Finally, we apply the theory of JS-admissibility and obtain

$$\frac{\mathbb{V}(X_n^{(T)})}{n} = \frac{G_2^{(T)}(-1)}{h} + \mathscr{F}[G_2^{(T)}](r\log_{1/p} n) + o(1).$$

We conclude this section by computing an explicit expression for

$$G_2^{(T)}(s) = \Phi(s) + \Phi_1(s) + \Phi_2(s)$$

at $s = -1 + \chi_k$ (note that only $k = 0$ makes sense in the case $\frac{\log p}{\log q} \notin \mathbb{Q}$)

First, we compute $\Phi(s)$:

$$\begin{aligned}
\Phi(s) =& \mathscr{M}[\tilde{V}_T(z); s] \\
=& \int_0^{+\infty} \left(1 - e^{-z} + pqz^2e^{-z} - pze^{-pz} - qze^{-qz}\right) \cdot \left(e^{-z} - pqz^2e^{-z}\right. \\
&\left. + pze^{-pz} + qze^{-qz}\right) z^{s-1} dz - \int_0^{+\infty} \tilde{g}_1'(z)^2 z^s dz \\
=& \Gamma(s+4)\left(-2^{-s-4}p^2q^2\right) \\
&+ \Gamma(s+3)\left(2p^2q(1+p)^{-s-3} + 2pq^2(1+q)^{-s-3}\right) \\
&+ \Gamma(s+2)\left(-3pq + 2^{-s-1}pq - 2^{-s-2}p^{-s} - 2^{-s-2}q^{-s}\right) \\
&+ \Gamma(s+1)\left(p^{-s} + q^{-s} - 2p(1+p)^{-s-1} - 2q(1+q)^{-s-1}\right) \\
&+ \Gamma(s)\left(1 - 2^{-s}\right) - \int_0^{+\infty} \tilde{g}_1'(z)^2 z^s dz,
\end{aligned}$$

where $\int_0^{+\infty} \tilde{g}_1'(z)^2 z^s dz$ is given by

$$\begin{aligned}
&\Gamma(s+5)\left(-2^{-s-5}p^2q^2\right)\\
&-\Gamma(s+4)\left(2^{-s-2}p^2q^2+2p^3q(1+p)^{-s-4}+2pq^3(1+q)^{-s-4}\right)\\
&+\Gamma(s+3)\Big((2p^2q+4p^3q)(1+p)^{-s-3}+(2pq^2+4pq^3)(1+q)^{-s-3}\\
&+2p^2q^2+2^{-s-3}(p^{1-s}+p^{1-s})-2^{-s-2}pq+2^{-s-1}p^2q^2\Big)\\
&+\Gamma(s+2)\Big((2p^2-4p^2q)(1+p)^{-s-2}+(2q^2-4pq^2)(1+q)^{-s-2}\\
&-2^{-s-1}(p^{1-s}+q^{1-s})-2p^2q-2pq^2+2^{-s}pq\Big)\\
&+\Gamma(s+1)\left(2pq+2^{-1-s}(1+p^{1-s}+q^{1-s})-2p(1+p)^{-s-1}-2q(1+q)^{-s-1}\right).
\end{aligned}$$

(The reason why we do not plug the last integral into the above expression for $\Phi(s)$ will become clear below.) In particular, when $s=-1+\chi_k$, we obtain the following result: if $k \neq 0$

$$\begin{aligned}
\Phi(-1+\chi_k)=&\Gamma(3+\chi_k)\left(-2^{-3-\chi_k}p^2q^2\right)\\
&+\Gamma(\chi_k+2)\left(2p^2q(1+p)^{-2-\chi_k}+2pq^2(1+q)^{-2-\chi_k}\right)\\
&+\Gamma(\chi_k+1)\left(-3pq+2^{-\chi_k}pq-2^{-1-\chi_k}\right)\\
&+\Gamma(\chi_k)\left(1-2p(1+p)^{-\chi_k}-2q(1+q)^{-\chi_k}\right)\\
&+\Gamma(\chi_k-1)\left(1-2^{1-\chi_k}\right)-\int_0^{+\infty}\tilde{g}_1'(z)^2z^{-1+\chi_k}dz,
\end{aligned}$$

and if $k=0$

$$\begin{aligned}
\Phi(-1)=&\frac{2p^2q}{(1+p)^2}+\frac{2pq^2}{(1+q)^2}-2pq-\frac{p^2q^2}{4}+\frac{1}{2}-p\log p-q\log q\\
&+2p\log(1+p)+2q\log(1+q)-2\log 2-\int_0^{+\infty}\tilde{g}_1'(z)^2z^{-1}dz.
\end{aligned}$$

Next, we turn to $\Phi_1(s)$ which is the Mellin transform of $\phi_1(z)$. We first rewrite $\phi_1(z)$ before we compute its Mellin transform.

$$\begin{aligned}
\phi_1(z)=&2\left(e^{-z}-pqz^2e^{-z}\right)\left(\tilde{f}_1(pz)+\tilde{f}_1(qz)\right)+2pze^{-pz}\tilde{f}_1(pz)+2qze^{-qz}\tilde{f}_1(qz)\\
&-2z\tilde{g}_1'(z)\left(p\tilde{f}_1'(pz)+q\tilde{f}_1'(qz)\right)\\
=&2\left(e^{-z}-pqz^2e^{-z}\right)\left(\frac{1}{2\pi i}\int_{(-1-\varepsilon)}\frac{G_1^{(T)}(w)(p^{-w}+q^{-w})}{1-p^{-w}-q^{-w}}z^{-w}dw\right)
\end{aligned}$$

$$
\begin{aligned}
&+ 2e^{-pz}\left(\frac{1}{2\pi i}\int_{(-1-\varepsilon)}\frac{G_1^{(T)}(w)p^{-w+1}}{1-p^{-w}-q^{-w}}z^{-w+1}dw\right)\\
&+ 2e^{-qz}\left(\frac{1}{2\pi i}\int_{(-1-\varepsilon)}\frac{G_1^{(T)}(w)q^{-w+1}}{1-p^{-w}-q^{-w}}z^{-w+1}dw\right)\\
&- 2\tilde{g}_1'(z)\left(\frac{1}{2\pi i}\int_{(-1-\varepsilon)}\frac{G_1^{(T)}(w)(p^{-w}+q^{-w})(-w)}{1-p^{-w}-q^{-w}}z^{-w}dw\right).
\end{aligned}
$$

Now, we shift the line of integration from $-1-\varepsilon$ to $-\infty$. Note that $G_1^{(T)}(z)$ has poles at -2, -3, -4, $\cdots$. We set $K_1(s) := -1 + pqs(s+1) - p^{-s}s - q^{-s}s$. Then, by using the residue theorem, we have

$$
\begin{aligned}
\phi_1(z) =& 2\left(e^{-z} - pqz^2e^{-z}\right)\sum_{\ell\geq 2}\frac{(-1)^\ell}{\ell!}\frac{K_1(-\ell)(p^\ell+q^\ell)}{1-p^\ell-q^\ell}z^\ell\\
&+ 2e^{-pz}\sum_{\ell\geq 2}\frac{(-1)^\ell}{\ell!}\frac{K_1(-\ell)p^{\ell+1}}{1-p^\ell-q^\ell}z^{\ell+1} + 2e^{-qz}\sum_{\ell\geq 2}\frac{(-1)^\ell}{\ell!}\frac{K_1(-\ell)q^{\ell+1}}{1-p^\ell-q^\ell}z^{\ell+1}\\
&- 2\tilde{g}_1'(z)\sum_{\ell\geq 2}\frac{(-1)^\ell}{\ell!}\frac{K_1(-\ell)\ell(p^\ell+q^\ell)}{1-p^\ell-q^\ell}z^\ell\\
=& 2\sum_{\ell\geq 2}\frac{(-1)^\ell}{\ell!}\frac{K_1(-\ell)}{1-p^\ell-q^\ell}\Big(\left(e^{-z}-pqz^2e^{-z}\right)(p^\ell+q^\ell)z^\ell + e^{-pz}p^{\ell+1}z^{\ell+1}\\
&+ e^{-qz}q^{\ell+1}z^{\ell+1} - \tilde{g}_1'(z)\ell(p^\ell+q^\ell)z^\ell\Big).
\end{aligned}
$$

Thus, its Mellin transform becomes

$$
\begin{aligned}
\Phi_1(s) =& \mathscr{M}[\phi_1(z);s]\\
=& 2\sum_{\ell\geq 2}\frac{(-1)^\ell}{\ell!}\frac{K_1(-\ell)}{1-p^\ell-q^\ell}\int_0^\infty\Big(\left(e^{-z}-pqz^2e^{-z}\right)(p^\ell+q^\ell)z^\ell\\
&+ e^{-pz}p^{\ell+1}z^{\ell+1} + e^{-qz}q^{\ell+1}z^{\ell+1} - \tilde{g}_1'(z)\ell(p^\ell+q^\ell)z^\ell\Big)z^{s-1}dz\\
=& 2\sum_{\ell\geq 2}\frac{(-1)^\ell}{\ell!}\frac{K_1(-\ell)}{1-p^\ell-q^\ell}\Big((p^\ell+q^\ell)\int_0^\infty\left(e^{-z}-pqz^2e^{-z}\right)z^{s+\ell-1}dz\\
&+ p^{\ell+1}\int_0^\infty e^{-pz}z^{s+\ell}dz + q^{\ell+1}\int_0^\infty e^{-qz}z^{s+\ell}dz\\
&- \ell(p^\ell+q^\ell)\int_0^\infty \tilde{g}_1'(z)z^{s+\ell-1}dz\Big)
\end{aligned}
$$

$$=2\sum_{\ell\geq 2}\frac{(-1)^{\ell}}{\ell!}\frac{K_1(-\ell)}{1-p^{\ell}-q^{\ell}}\Big((p^{\ell}+q^{\ell})\left(\Gamma(s+\ell)-pq\Gamma(s+\ell+2)\right)$$
$$+p^{-s}\Gamma(s+\ell+1)+q^{-s}\Gamma(s+\ell+1)-\ell(p^{\ell}+q^{\ell})\Big(2pq\Gamma(s+\ell+1)$$
$$+\Gamma(s+\ell)-pq\Gamma(s+\ell+2)-p^{-\ell-s+1}\Gamma(s+\ell)-q^{-\ell-s+1}\Gamma(s+\ell)$$
$$+p^{-\ell-s+1}\Gamma(s+\ell+1)+q^{-\ell-s+1}\Gamma(s+\ell+1)\Big)\Big).$$

So, when $s=-1+\chi_k$

$$\Phi_1(-1+\chi_k)=2\sum_{\ell\geq 2}\frac{(-1)^{\ell}}{\ell!}\frac{K_1(-\ell)}{1-p^{\ell}-q^{\ell}}\Big(pq(p^{\ell}+q^{\ell})(\ell-1)\Gamma(\chi_k+\ell+1)$$
$$+\left(1-\ell(p^{\ell}+q^{\ell})(2pq+p^{-\ell+2}+q^{-\ell+2})\right)\Gamma(\chi_k+\ell)$$
$$+(p^{\ell}+q^{\ell})(1-\ell+p^{-\ell+2}\ell+q^{-\ell+2}\ell)\Gamma(\chi_k+\ell-1)\Big).$$

Remark 15. The reader should note that the above series representation does not converge in the usual sense. However, convergence is granted if one uses Abel summability and indeed the Abel sum gives the correct result (an explanation for this will be given at the end of this section). Thus, from now on, convergence of series will always be understood to be in the sense of Abel summability.

Finally, we consider $\Phi_2(s)$:

$$\Phi_2(s)=\mathscr{M}[\phi_2(z);s]$$
$$=\mathscr{M}[pqz\big(\tilde{f}_1'(pz)-\tilde{f}_1'(qz)\big)^2;s]$$
$$=\frac{pq}{2\pi i}\int_{(-\frac{1}{2})}\mathscr{M}[\tilde{f}_1'(pz)-\tilde{f}_1'(qz);w]\mathscr{M}[z\left(\tilde{f}_1'(pz)-\tilde{f}_1'(qz)\right);s-w]dw$$
$$=\frac{pq}{2\pi i}\int_{(-\frac{1}{2})}\frac{(1-w)G_1^{(T)}(w-1)\left(p^{-w}-q^{-w}\right)}{1-p^{1-w}-q^{1-w}}$$
$$\cdot\frac{(w-s)G_1^{(T)}(s-w)\left(p^{w-s-1}-q^{w-s-1}\right)}{1-p^{w-s}-q^{w-s}}dw,$$

so, we know that when $s=-1+\chi_k$

$$\Phi_2(-1+\chi_k)=\frac{pq}{2\pi i}\int_{(-\frac{1}{2})}\frac{\left(p^{-w}-q^{-w}\right)\left(p^{w}-q^{w}\right)}{\left(1-p^{1-w}-q^{1-w}\right)\left(1-p^{1+w}-q^{1+w}\right)}(1-w)$$
$$\cdot G_1^{(T)}(w-1)(1+w-\chi_k)G_1^{(T)}(-1-w+\chi_k)dw$$

$$
\begin{aligned}
=&\frac{1}{2\pi i}\int_{(0)^+}\left(\frac{1}{1-p^{1-w}-q^{1-w}}+\frac{p^{1+w}+q^{1+w}}{1-p^{1+w}-q^{1+w}}\right)(1-w)\\
&\qquad\cdot G_1^{(T)}(w-1)(1+w-\chi_k)G_1^{(T)}(-1-w+\chi_k)dw\\
=&\frac{1}{2\pi i}\int_{(0)^+}\frac{1}{1-p^{1-w}-q^{1-w}}(1-w)G_1^{(T)}(w-1)\\
&\qquad\cdot (1+w-\chi_k)G_1^{(T)}(-1-w+\chi_k)dw\\
&+\frac{1}{2\pi i}\int_{(0)^+}\frac{p^{1+w}+q^{1+w}}{1-p^{1+w}-q^{1+w}}(1-w)G_1^{(T)}(w-1)\\
&\qquad\cdot (1+w-\chi_k)G_1^{(T)}(-1-w+\chi_k)dw\\
:=&P_1+P_2,
\end{aligned}
$$

where P_1 and P_2 is the first term and second term of $\Phi_2(-1+\chi_k)$, respectively, and $\int_{(0)^+}$ is the integral along the line $\Re(w)=0$ with a small indentation to the right at poles. In the sequel, we will mainly focus on the case $\frac{\log p}{\log q}\in\mathbb{Q}$. The irrational case is treated similarly.

By the change of variables $w\mapsto\chi_k-w$ and then by moving the line of integration to the right, we have

$$
\begin{aligned}
P_1=&\frac{1}{2\pi i}\int_{(0)^-}\frac{(1+w-\chi_k)G_1^{(T)}(-w-1+\chi_k)(1-w)G_1^{(T)}(w-1)}{1-p^{1+w}-q^{1+w}}dw\\
=&-\frac{1}{h}\sum_{j\in\mathbb{Z}}(\chi_j-1)G_1^{(T)}(\chi_j-1)(-1+\chi_{k-j})G_1^{(T)}(-1+\chi_{k-j})\\
&+\frac{1}{2\pi i}\int_{(0)^+}\frac{(1+w-\chi_k)G_1^{(T)}(-w-1+\chi_k)(1-w)G_1^{(T)}(w-1)}{1-p^{1+w}-q^{1+w}}dw.
\end{aligned}
$$

The last integral equals

$$
\frac{1}{2\pi i}\int_{(0)^+}(1+w-\chi_k)G_1^{(T)}(-w-1+\chi_k)(1-w)G_1^{(T)}(w-1)dw+P_2.
$$

Note that

$$
\begin{aligned}
&\frac{1}{2\pi i}\int_{(0)^+}(1+w-\chi_k)G_1^{(T)}(-w-1+\chi_k)(1-w)G_1^{(T)}(w-1)dw\\
&=\int_0^\infty \tilde{g}_1'(z)^2 z^{-1+\chi_k}dz.
\end{aligned}
$$

Combining these relations, we get

$$\Phi_2(-1+\chi_k) = 2P_2 - \frac{1}{h}\sum_{j\in\mathbb{Z}}(\chi_j-1)G_1^{(T)}(\chi_j-1)(-1+\chi_{k-j})G_1^{(T)}(-1+\chi_{k-j}) + \int_0^\infty \tilde{g}_1'(z)^2 z^{-1+\chi_k}dz,$$

where

$$P_2 = \frac{1}{2\pi i}\int_{(0)^+}\frac{p^{1+w}+q^{1+w}}{1-p^{1+w}-q^{1+w}}(1-w)G_1^{(T)}(w-1) \cdot (1+w-\chi_k)G_1^{(T)}(-1-w+\chi_k)dw.$$

Now, we shift the line of integration of P_2 from $(0)^+$ to ∞. Note that $G_1^{(T)}(-1-w+\chi_k)$ has poles at $w=1+\chi_k, 2+\chi_k, 3+\chi_k, \cdots$ and recall that $K_1(s) := -1+pqs(s+1)-p^{-s}s-q^{-s}s$. Then, by using the residue theorem, we have

$$P_2 = \sum_{\ell\geq 1}\frac{(-1)^\ell}{\ell!}\frac{p^{1+\ell}+q^{1+\ell}}{1-p^{1+\ell}-q^{1+\ell}}K_1(\ell+\chi_k-1)K_1(-1-\ell)\Gamma(\ell+\chi_k).$$

Now, we can return to $G_2^{(T)}(-1+\chi_k)$. Putting everything together yields that $G_2^{(T)}(-1+\chi_k)$ for $k\neq 0$ is given by

$$\begin{aligned}
&\Phi(-1+\chi_k)+\Phi_1(-1+\chi_k)+\Phi_2(-1+\chi_k)\\
=&\Gamma(3+\chi_k)\left(-2^{-3-\chi_k}p^2q^2\right)\\
&+\Gamma(\chi_k+2)\left(2p^2q(1+p)^{-2-\chi_k}+2pq^2(1+q)^{-2-\chi_k}\right)\\
&+\Gamma(\chi_k+1)\left(-3pq+2^{-\chi_k}pq-2^{-1-\chi_k}\right)\\
&+\Gamma(\chi_k)\left(1-2p(1+p)^{-\chi_k}-2q(1+q)^{-\chi_k}\right)\\
&+\Gamma(\chi_k-1)\left(1-2^{1-\chi_k}\right)\\
&+2\sum_{\ell\geq 2}\frac{(-1)^\ell}{\ell!}\frac{K_1(-\ell)}{1-p^\ell-q^\ell}\Bigg(pq(p^\ell+q^\ell)(\ell-1)\Gamma(\chi_k+\ell+1)\\
&+\left(1-\ell(p^\ell+q^\ell)(2pq+p^{-\ell+2}+q^{-\ell+2})\right)\Gamma(\chi_k+\ell)\\
&+(p^\ell+q^\ell)(1-\ell+p^{-\ell+2}\ell+q^{-\ell+2}\ell)\Gamma(\chi_k+\ell-1)\Bigg)\\
&+2\sum_{\ell\geq 1}\frac{(-1)^\ell}{\ell!}\frac{p^{1+\ell}+q^{1+\ell}}{1-p^{1+\ell}-q^{1+\ell}}K_1(\ell+\chi_k-1)K_1(-1-\ell)\Gamma(\ell+\chi_k)\\
&-\frac{1}{h}\sum_{j\in\mathbb{Z}}(\chi_j-1)G_1^{(T)}(\chi_j-1)(-1+\chi_{k-j})G_1^{(T)}(-1+\chi_{k-j}),
\end{aligned}$$

and for $k=0$, we obtain

$$\begin{aligned}
G_2^{(T)}(-1) =& \frac{2p^2q}{(1+p)^2} + \frac{2pq^2}{(1+q)^2} - 2pq - \frac{p^2q^2}{4} + 2p\log(1+p) + 2q\log(1+q) \\
&+ \frac{1}{2} + h - 2\log 2 + 2\sum_{\ell\geq 2}(-1)^\ell \frac{K_1(-\ell)}{1-p^\ell-q^\ell}\Bigg((p^\ell+q^\ell)\Bigg(pq\ell - 3pq \\
&- p^{-\ell+2} - q^{-\ell+2} + \frac{1-\ell+p^{-\ell+2}\ell+q^{-\ell+2}\ell}{\ell(\ell-1)}\Bigg) + \frac{1}{\ell}\Bigg) \\
&+ 2\sum_{\ell\geq 1}\frac{(-1)^\ell}{\ell}\frac{p^{1+\ell}+q^{1+\ell}}{1-p^{1+\ell}-q^{1+\ell}} K_1(\ell-1)K_1(-1-\ell) \\
&- \frac{1}{h}\sum_{j\in\mathbb{Z}\setminus\{0\}}(\chi_j-1)G_1^{(T)}(\chi_j-1)(-1-\chi_j)G_1^{(T)}(-1-\chi_j) \\
&- \frac{1}{h}(pq+1-h)^2.
\end{aligned}$$

Finally, we further simplify the second last term of $G_2^{(T)}(-1)$ (which we call M^*). M^* is equal to

$$\begin{aligned}
&-\frac{1}{h}\sum_{j\in\mathbb{Z}\setminus\{0\}}(\chi_j-1)G_1^{(T)}(\chi_j-1)(-1+\chi_{-j})G_1^{(T)}(-1-\chi_j) \\
=&-\frac{1}{h}\sum_{j\in\mathbb{Z}\setminus\{0\}}(\chi_j-1)K_1(\chi_j-1)(-1-\chi_j)K_1(-1-\chi_j)\Gamma(\chi_j-1)\Gamma(-1-\chi_j).
\end{aligned}$$

We know that $K_1(\chi_j-1) = \chi_j\,(pq(\chi_j-1)-1)$ and $K_1(-1-\chi_j) = \chi_j(pq(\chi_j+1)+1)$, So we can rewrite the previous expression into

$$-\frac{1}{h}\sum_{j\in\mathbb{Z}\setminus\{0\}}(pq(\chi_j-1)-1)\,\chi_j\,(pq(\chi_j+1)+1)\,\Gamma(\chi_j+1)\Gamma(-\chi_j).$$

Next, recall that $\Gamma(1-z)\Gamma(z) = \pi/\sin\pi z$ which yields

$$\begin{aligned}
&-\frac{1}{h}\sum_{j\in\mathbb{Z}\setminus\{0\}}\frac{\pi}{\sin\pi(-\chi_j)}(pq(\chi_j-1)-1)\,\chi_j\,(pq(\chi_j+1)+1) \\
=&\frac{1}{h}\sum_{j\in\mathbb{Z}\setminus\{0\}}\frac{\pi\chi_j}{\sin\pi\chi_j}(pq(\chi_j-1)-1)\,(pq(\chi_j+1)+1)\,.
\end{aligned}$$

Finally, we substitute $\frac{2rj\pi i}{\log p}$ for χ_j, and obtain

$$
\begin{aligned}
M^* &= -\frac{1}{h}\sum_{j\in\mathbb{Z}\setminus\{0\}} \frac{\frac{2rj\pi^2}{\log p}}{\sinh\left(\frac{2rj\pi^2}{\log p}\right)} \left(p^2q^2\left(\frac{2rj\pi}{\log p}\right)^2 + p^2q^2 + 2pq + 1\right)\\
&= -\frac{2}{h}\sum_{j\geq 1} \frac{\frac{2rj\pi^2}{\log p}}{\sinh\left(\frac{2rj\pi^2}{\log p}\right)} \left(p^2q^2\left(\frac{2rj\pi}{\log p}\right)^2 + p^2q^2 + 2pq + 1\right).
\end{aligned}
$$

Overall, we have proved the following result,

Theorem 12. *The variance of the number of 2-protected nodes in tries satisfies*

$$
\frac{\mathbb{V}(X_n^{(T)})}{n} = \frac{G_2^{(T)}(-1)}{h} + \mathscr{F}[G_2^{(T)}](r\log_{1/p} n) + o(1),
$$

where $G_2^{(T)}(-1+\chi_k)$ for $k \neq 0$ (when $\frac{\log p}{\log q} \in \mathbb{Q}$) is given by

$$
\begin{aligned}
&\Gamma(3+\chi_k)\left(-2^{-3-\chi_k}p^2q^2\right)\\
&+\Gamma(\chi_k+2)\left(2p^2q(1+p)^{-2-\chi_k} + 2pq^2(1+q)^{-2-\chi_k}\right)\\
&+\Gamma(\chi_k+1)\left(-3pq + 2^{-\chi_k}pq - 2^{-1-\chi_k}\right)\\
&+\Gamma(\chi_k)\left(1 - 2p(1+p)^{-\chi_k} - 2q(1+q)^{-\chi_k}\right)\\
&+\Gamma(\chi_k-1)\left(1-2^{1-\chi_k}\right)\\
&+2\sum_{\ell\geq 2}\frac{(-1)^\ell}{\ell!}\frac{K_1(-\ell)}{1-p^\ell-q^\ell}\bigg(pq(p^\ell+q^\ell)(\ell-1)\Gamma(\chi_k+\ell+1)\\
&+\left(1-\ell(p^\ell+q^\ell)(2pq+p^{-\ell+2}+q^{-\ell+2})\right)\Gamma(\chi_k+\ell)\\
&+(p^\ell+q^\ell)(1-\ell+p^{-\ell+2}\ell+q^{-\ell+2}\ell)\Gamma(\chi_k+\ell-1)\bigg)\\
&+2\sum_{\ell\geq 1}\frac{(-1)^\ell}{\ell!}\frac{p^{1+\ell}+q^{1+\ell}}{1-p^{1+\ell}-q^{1+\ell}}K_1(\ell+\chi_k-1)K_1(-1-\ell)\Gamma(\ell+\chi_k)\\
&-\frac{1}{h}\sum_{j\in\mathbb{Z}}(\chi_j-1)G_1^{(T)}(\chi_j-1)(-1+\chi_{k-j})G_1^{(T)}(-1+\chi_{k-j})
\end{aligned}
$$

and $G_2^{(T)}(-1)$ is given by

$$
\begin{aligned}
&\frac{2p^2q}{(1+p)^2} + \frac{2pq^2}{(1+q)^2} - 2pq - \frac{p^2q^2}{4} + 2p\log(1+p) + 2q\log(1+q)\\
&+\frac{1}{2} + h - 2\log 2 + 2\sum_{\ell\geq 2}(-1)^\ell\frac{K_1(-\ell)}{1-p^\ell-q^\ell}\bigg((p^\ell+q^\ell)\Big(pq\ell - 3pq
\end{aligned}
$$

$$
\left. - p^{-\ell+2} - q^{-\ell+2} + \frac{1-\ell+p^{-\ell+2}\ell+q^{-\ell+2}\ell}{\ell(\ell-1)}\right) + \frac{1}{\ell}\Bigg)
$$

$$
+ 2\sum_{\ell\geq 1}\frac{(-1)^\ell}{\ell}\frac{p^{1+\ell}+q^{1+\ell}}{1-p^{1+\ell}-q^{1+\ell}}K_1(\ell-1)K_1(-1-\ell)
$$

$$
- \frac{1}{h}(pq+1-h)^2
$$

$$
- \begin{cases} \dfrac{1}{h\log p}\displaystyle\sum_{j\geq 1}\frac{4rj\pi^2}{\sinh\left(\frac{2rj\pi^2}{\log p}\right)}\left(p^2q^2\left(\frac{2rj\pi}{\log p}\right)^2+(pq+1)^2\right), & \textit{if } \frac{\log p}{\log q}\in\mathbb{Q}; \\ 0, & \textit{if } \frac{\log p}{\log q}\notin\mathbb{Q}. \end{cases}
$$

Remark 16. Comparing with Theorem 4, we obtained an explicit expression of the period function and our expression of $G_2^{(T)}(-1)$ is also different. Thus, we have $G_2^{(T)}(-1)/h = c_1 + c_2 - c_3^2 + c_4$ and this identity can also be proved directly (however, the computation is long). More precisely, we have

$$
\begin{aligned}
(c_1+c_2)h =& \frac{2p^2q}{(1+p)^2} + \frac{2pq^2}{(1+q)^2} - 2pq - \frac{p^2q^2}{4} + 2p\log(1+p) + 2q\log(1+q) \\
&+ \frac{1}{2} + h - 2\log 2 + 2\sum_{\ell\geq 2}(-1)^\ell\frac{K_1(-\ell)}{1-p^\ell-q^\ell}\Bigg((p^\ell+q^\ell)\Bigg(pq\ell - 3pq \\
&- p^{-\ell+2} - q^{-\ell+2} + \frac{1-\ell+p^{-\ell+2}\ell+q^{-\ell+2}\ell}{\ell(\ell-1)}\Bigg) + \frac{1}{\ell}\Bigg) \\
&+ 2\sum_{\ell\geq 1}\frac{(-1)^\ell}{\ell}\frac{p^{1+\ell}+q^{1+\ell}}{1-p^{1+\ell}-q^{1+\ell}}K_1(\ell-1)K_1(-1-\ell)).
\end{aligned}
$$

Finally, we remark that if we consider the symmetric case ($p = q = 1/2$), our analysis will become much easier because $\tilde{\phi}_2(z)$ equals to 0. This yields the following result with a somewhat different expression for the periodic function.

Theorem 13. *The variance of the number of 2-protected nodes in symmetric tries satisfies*

$$
\frac{\mathbb{V}(X_n^{(T)})}{n} = \frac{G_2^{(T)}(-1)}{\log 2} + \mathscr{F}[G_2^{(T)}](r\log_2 n) + o(1),
$$

where for $k \neq 0$ (when $\frac{\log p}{\log q}\in\mathbb{Q}$)

$$
G_2^{(T)}(-1+\chi_k) = \Phi(-1+\chi_k) + \Phi_1(-1+\chi_k)
$$

$$
\begin{aligned}
=&\frac{-\Gamma(\chi_k+4)}{256}+\Gamma(\chi_k+3)\left(\frac{3}{128}+2\cdot 3^{-\chi_k-3}\right)-\Gamma(\chi_k+2)\\
&\cdot\left(\frac{3}{16}+2\cdot 3^{-\chi_k-2}\right)-\frac{1}{2}\Gamma(\chi_k+1)-\Gamma(\chi_k)-\Gamma(\chi_k-1)\\
&+2\sum_{\ell\geq 2}\frac{(-1)^\ell}{\ell!}\frac{-1+\frac{\ell(\ell+1)}{4}+2^{-\ell+1}\ell}{1-2^{-\ell+1}}\cdot\Bigg(\Gamma(\ell+\chi_k-1)\\
&\left(2^{-\ell+1}-2^{-\ell+1}\ell+\ell\right)+\Gamma(\ell+\chi_k)\left(1-2^{-\ell}\ell-\ell\right)\\
&+\Gamma(\ell+\chi_k+1)2^{-\ell-1}(\ell-1)\Bigg)
\end{aligned}
$$

and for $k=0$

$$
\begin{aligned}
G_2^{(T)}(-1)=&\frac{905}{3456}+2\sum_{\ell\geq 2}(-1)^\ell\frac{-1+\frac{\ell(\ell+1)}{4}+2^{-\ell+1}\ell}{1-2^{-\ell+1}}\cdot\Bigg(2^{-\ell-1}(\ell-1)\\
&+\frac{2^{-\ell+1}}{\ell(\ell-1)}+\frac{1-2^{-\ell+1}}{\ell-1}+\frac{1}{\ell}-2^{-\ell}-1\Bigg).
\end{aligned}
$$

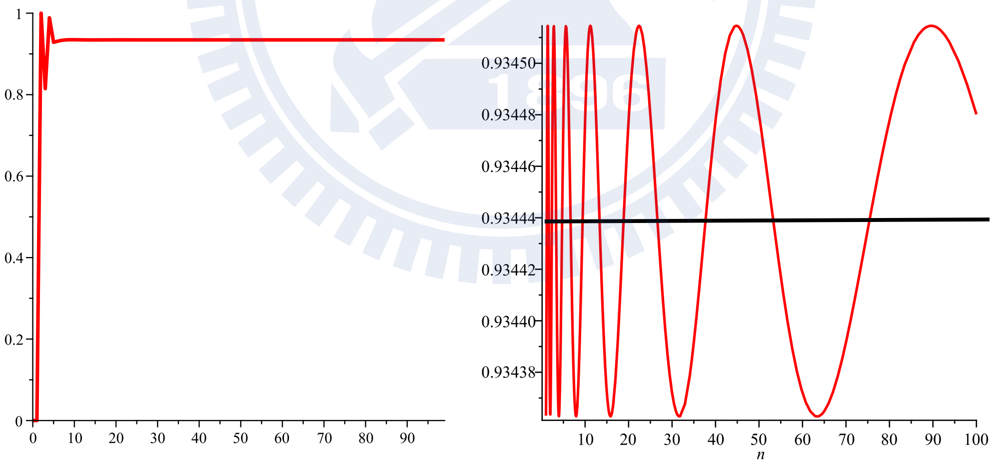


Figure 3.2: Left is a plot of $\mathbb{V}(X_n^{(T)})/n$ for $n=1\dots100$; right is a plot of the periodic function of the main term in the asymptotic expansion of $\mathbb{V}(X_n^{(T)})/n$.

Remark 17. The above result gives

$$
\frac{G_2^{(T)}(-1)}{\log 2}\approx 0.934438704470192498535678544772878810582742224974 62\cdots.
$$

which coincides with the result from Theorem 12 and the one from [7]; see also Figure 3.2.

Remark 18. We now give an explanation why Abel summability gives the correct result in all the integral evaluations above. Therefore, consider the following integral

$$\frac{1}{2\pi i}\int_{(0)+}\frac{\Gamma(-w+1)\Gamma(w)}{1-2^{-w}}dw.$$

All of the above integrals are of a similar type (only more involved).

Observe that moving the line of integration to infinity yields the divergent series (with the usual notion of convergence)

$$\sum_{\ell\geq 1}\frac{(-1)^{\ell-1}}{1-2^{-\ell}}. \tag{3.9}$$

Thus, this step is not jusified. However, if we consider Abel summability then we have to replace the above integral by

$$\frac{1}{2\pi i}\int_{(0)+}\frac{\Gamma(-w+1)\Gamma(w)}{1-2^{-w}}x^{\omega}dw$$

with $|x|<1$. Now, due to the appearance of x^{ω} which produces exponential decay, we are indeed allowed to move the line to infinity and obtain

$$\sum_{\ell\geq 1}\frac{(-1)^{\ell-1}}{1-2^{-\ell}}x^{\ell}.$$

Consequently, we have the identity

$$\frac{1}{2\pi i}\int_{(0)+}\frac{\Gamma(-w+1)\Gamma(w)}{1-2^{-w}}x^{\omega}dw=\sum_{\ell\geq 1}\frac{(-1)^{\ell-1}}{1-2^{-\ell}}x^{\ell}.$$

Taking limits on both sides produces our original integral on the left-hand side and the Abel sum of (3.9) on the right-hand side. This yields our claimed result.

Note that alternatively one could move the line of integration of our original integral to $-\infty$ (this only works in the symmetric case; in the asymmetric case the resulting expression would become much more complicated). Then, we obtain the convergent sum

$$\frac{1}{2}+\sum_{\ell\geq 1}\frac{(-1)^{\ell-1}}{2^{\ell}-1}.$$

Thus, we have the interesting identiy

$$\frac{1}{2} + \sum_{\ell\geq 1} \frac{(-1)^{\ell-1}}{2^\ell - 1} = \sum_{\ell\geq 1} \frac{(-1)^{\ell-1}}{1-2^{-\ell}}$$

with a rapidly converging series on the left-hand side and a divergent series on the right-hand side (whose Abel sum equals the value on the left-hand side). We leave it as an exercise to the reader to prove this identity directly.

3.3 Central limit theorem (CLT)

In this section, we are going to prove a central limit theorem for the number of 2-protected nodes. We will use the tools from [5] which can be applied to the current problem. First, we have to show that the variance grows at least linearly for n large enough. For this purpose, we recall the following proposition from [5].

Proposition 4. *Let f_n be a sequence satisfying a recurrence of the form*

$$f_n = \sum_{j=0}^{n} \binom{n}{j} p^j q^{n-j} (f_j + f_{n-j}) + g_n, \qquad (n \geq 2),$$

with $f_0 = f_1 = 0$. Assume that g_n is non-negative and $g_{n_0} > 0$ for some $n_0 \geq 2$. Then, $f_n = \Omega(n)$.

From this proposition, we deduce the following lemma.

Lemma 3.3.1. *We have that $\mathbb{V}(X_n^{(T)}) = \Omega(n)$.*

Proof. We first derive a recurrence for $\mathbb{V}(X_n^{(T)})$. Therefore, set

$$M_n(y) = \mathbb{E}(e^{(X_n^{(T)} - a_n)y})$$

where a_n denotes the mean of $X_n^{(T)}$. Then, from (1.2), we obtain

$$M_n(y) = \sum_{k=0}^{n} \pi_{n,k} M_k M_{n-k} \mathbb{E}(e^{(T_n - a_n + a_k + a_{n-k})} | I_n = k)$$

with $n \geq 2$ and $M_0(y) = M_1(y) = 1$. Observe that

$$\sigma_n^2 = \mathbb{V}(X_n^{(T)}) = M_n(y)''|_{y=0}.$$

Taking derivatives on both side of the above recurrence yields

$$\sigma_n^2 = \sum_{k=0}^{n} \pi_{n,k}(\sigma_k^2 + \sigma_{n-k}^2) + \eta_n, \qquad (n \geq 2),$$

where $\sigma_0^2 = \sigma_1^2 = 0$ and

$$\eta_n = \sum_{k=0}^{n} \pi_{n,k} \mathbb{E}((T_n - a_n + a_k + a_{n-k})^2 | I_n = k).$$

From the expression of η_n, we see that $\eta_n \geq 0$ because of the square. Then, by Proposition 4, either $\mathbb{V}(X_n^{(T)})$ grows at least linearly or is identical to 0. The latter case, however, cannot happen since $X_n^{(T)}$ is easily seen to be not deterministic for $n \geq 2$. ∎

Now, we can state the main result of this section (this result proves a conjecture from [7]).

Theorem 14. *We have,*

$$\frac{X_n^{(T)} - \mathbb{E}(X_n^{(T)})}{\sqrt{\mathbb{V}(X_n^{(T)})}} \xrightarrow{d} N(0,1),$$

where $\xrightarrow{d}$ *denotes convergence in distribution and* $N(0,1)$ *is the standard normal distribution.*

Proof. This follows from the above lemma and our expansions for the mean and the variance from Section 3.1 and Section 3.2 by an application of the contraction method along the same lines as in the proof of Theorem 2 in [5] (note that the theorem does not directly apply to the current situation due to dependency between T_n and I_n). ∎

Chapter 4

The number of 2-protected nodes in PATRICIA tries

Let $X_n^{(P)}$ be the number of 2-protected nodes in a random PATRICIA trie built from n records. Then, we have, for $n \geq 2$,

$$X_n^{(P)} \stackrel{d}{=} \begin{cases} X_n^{(P)}, & \text{when } I_n = 0 \text{ or } n; \\ X_{n-1}^{(P)}, & \text{when } I_n = 1 \text{ or } n-1; \\ X_{I_n}^{(P)} + X_{n-I_n}^{(P)*} + 1, & \text{otherwise}, \end{cases} \tag{4.1}$$

where notation is as in Section 1.4 and initial conditions are $X_0^{(P)} = X_1^{(P)} = 0$. Again this can be rewritten to (1.2) with

$$T_n = \begin{cases} 0, & \text{if } I_n \in \{0, 1, n-1, n\}; \\ 1, & \text{otherwise}. \end{cases} \tag{4.2}$$

We start with the mean. Note that we could in principle repeat the analysis of Section 3.1. However, we do not have to do so due to the following relationship between 2-protected nodes in tries and PATRICIA tries via the number of internal nodes:

$$X_n^{(P)} = X_n^{(T)} - N_n + n - 1. \tag{4.3}$$

Thus, we obtain the result for PATRICIA tries from the result for tries (Section 3.1) and the known result for the number of internal nodes (Section 1.3).

$$\frac{\mathbb{E}(X_n^{(P)})}{n} = \frac{\mathbb{E}(X_n^{(T)})}{n} - \frac{\mathbb{E}(N_n)}{n} + 1 - \frac{1}{n} \tag{4.4}$$

$$
\begin{aligned}
&= \frac{pq+1-h}{h} + \mathscr{F}[G_1^{(T)}](r\log_{1/p} n) + o(1) \\
&\quad - \frac{1}{h} - \mathscr{F}[G_1^{(N)}](r\log_{1/p} n) - o(1) + 1 - \frac{1}{n} \\
&= \frac{pq}{h} + \mathscr{F}[G_1^{(P)}](r\log_{1/p} n) + o(1),
\end{aligned}
$$

where

$$
\begin{aligned}
G_1^{(P)}(s) &= G_1^{(T)}(s) - G_1^{(N)}(s) \\
&= \Gamma(s)\left(-1 + pqs(s+1) - p^{-s}s - q^{-s}s\right) - \Gamma(s)(-s-1) \\
&= \Gamma(s)\left(s + pqs(s+1) - p^{-s}s - q^{-s}s\right) \\
&= \Gamma(s+1)\left(1 - p^{-s} - q^{-s} + pq(s+1)\right).
\end{aligned}
$$

This gives the following result,

Theorem 15. *The mean of the number of 2-protected nodes in PATRICIA tries satisfies*

$$
\frac{\mathbb{E}(X_n^{(P)})}{n} = \frac{pq}{h} + \mathscr{F}[G_1^{(P)}](r\log_{1/p} n) + o(1), \tag{4.5}
$$

where for $k \neq 0$ *(when* $\frac{\log p}{\log q} \in \mathbb{Q}$*)*

$$
G_1^{(P)}(-1+\chi_k) = pq\Gamma(\chi_k + 1).
$$

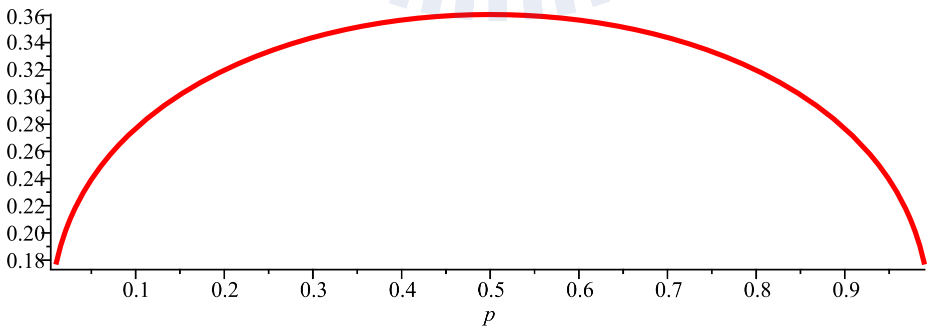


Figure 4.1: A plot of pq/h which is the average value of the main term of $\mathbb{E}(X_n^{(P)})/n$.

Remark 19. For a plot of the average value of $\mathbb{E}(X_n^{(P)})/n$ see Figure 4.1. Note that in contrast to tries, the symmetric PATRICIA trie now attains the maximum. Again this is intuitively clear since in the asymmetric case most 2-protected nodes come from one-way branching which have been removed.

Next, for the variance, we first point out that we can not use the same idea as for the mean because of the appearance of the covariance of the number of 2-protected nodes and the number of internal nodes in tries.

So, we have to return to the analysis from Section 3.2. Due to similarities and in order to keep the thesis short, we do not give details and only show results.

First, as in Section 3.1, one derives for the Poisson generating function of mean (which we again denote by $\tilde{f}_1(z)$)

$$\tilde{f}_1(z) = \frac{1}{2\pi i}\int_{-1-\varepsilon} \frac{G_1^{(P)}(s)}{1-p^{-s}-q^{-s}}ds,$$

where

$$G_1^{(P)}(s) = (pqs + 1 - p^{-s} - q^{-s})(s+1)\Gamma(s).$$

This is then used as in Section 3.2 to show that

$$G_2^{(P)}(-1+\chi_k) = \Phi(-1+\chi_k) + \Phi_1(-1+\chi_k) + \Phi_2(-1+\chi_k),$$

where

$$\begin{aligned}\Phi(-1+\chi_k) =& \Gamma(3+\chi_k)\left(-2^{-3-\chi_k}p^2q^2\right)\\ &+\Gamma(\chi_k+2)\left(2p^2q(1+p)^{-2-\chi_k} + 2pq^2(1+q)^{-2-\chi_k} - 2^{-1-\chi_k}pq\right)\\ &+\Gamma(\chi_k+1)\big(-3pq - 2^{-1-\chi_k}(1+2pq+p^{1-\chi_k}+q^{1-\chi_k})\\ &+2p(1+q)(1+p)^{-1-\chi_k} + 2q(1+p)(1+q)^{-1-\chi_k}\big)\\ &+(\Gamma(\chi_k)+\Gamma(\chi_k-1))\big(2(1+p)^{1-\chi_k} + 2(1+q)^{1-\chi_k} - 2^{1-\chi_k}\\ &+(1-2^{1-\chi_k})(p^{1-\chi_k}+q^{1-\chi_k}) - 3\big)\\ &-\int_0^{+\infty}\tilde{g}_1'(z)^2 z^{-1+\chi_k}dz\end{aligned}$$

and $\int_0^{+\infty}\tilde{g}_1'(z)^2 z^s dz$ is given by

$$\begin{aligned}&\Gamma(s+5)\left(-2^{-s-5}p^2q^2\right)\\ &-\Gamma(s+4)\left(2^{-s-2}p^2q^2 + 2p^3q(1+p)^{-s-4} + 2pq^3(1+q)^{-s-4} - 2^{-s-3}pq\right)\\ &+\Gamma(s+3)\Big((4p^3q-2p^2)(1+p)^{-s-3} + (4pq^3-2q^2)(1+q)^{-s-3}\\ &+2p^2q^2 + 2^{-s-3}(p^{1-s}+p^{1-s}) + 2^{-s-2}(2pq-1)^2\Big).\end{aligned}$$

Moreover, the same as in the previous chapter, we set $K_1(s) := (pqs + 1 - p^{-s} - q^{-s})(s+1)$, so we have $\Phi_1(-1+\chi_k)$ is equal to

$$
\begin{aligned}
&2\sum_{\ell\geq 2}\frac{(-1)^\ell}{\ell!}\frac{K_1(-\ell)}{1-p^\ell-q^\ell}\Big(pq(p^\ell+q^\ell)(\ell-1)\Gamma(\chi_k+\ell+1)\\
&+\big((p^\ell+q^\ell)(\ell-2pq\ell-p^{-\ell+2}\ell-q^{-\ell+2}\ell-1)+1\big)\,\Gamma(\chi_k+\ell)\\
&+(1-p^\ell-q^\ell)\Gamma(\chi_k+\ell-1)\Big)
\end{aligned}
$$

and $\Phi_2(-1+\chi_k)$ is equal to (in the case $\frac{\log p}{\log q}\in\mathbb{Q}$; a similar expression can be given for the case $\frac{\log p}{\log q}\notin\mathbb{Q}$)

$$
\begin{aligned}
&2\sum_{\ell\geq 1}\frac{(-1)^\ell}{\ell!}\frac{p^{1+\ell}+q^{1+\ell}}{1-p^{1+\ell}-q^{1+\ell}}K_1(\ell+\chi_k-1)K_1(-1-\ell)\Gamma(\ell+\chi_k)\\
&-\frac{1}{h}\sum_{j\in\mathbb{Z}}(\chi_j-1)G_1^{(P)}(\chi_j-1)(-1+\chi_{k-j})G_1^{(P)}(-1+\chi_{k-j})\\
&+\int_0^\infty \tilde{g}_1'(z)^2 z^{-1+\chi_k}\,dz.
\end{aligned}
$$

Combining everything gives the following result.

Theorem 16. *The variance of the number of 2-protected nodes in PATRICIA tries satisfies*

$$
\frac{\mathbb{V}(X_n^{(P)})}{n}=\frac{G_2^{(P)}(-1)}{h}+\mathscr{F}[G_2^{(P)}](r\log_{1/p}n)+o(1),
$$

where $G_2^{(P)}(-1+\chi_k)$ for $k\neq 0$ (when $\frac{\log p}{\log q}\in\mathbb{Q}$) is given by

$$
\begin{aligned}
&\Gamma(3+\chi_k)\left(-2^{-3-\chi_k}p^2q^2\right)\\
&+\Gamma(\chi_k+2)\left(2p^2q(1+p)^{-2-\chi_k}+2pq^2(1+q)^{-2-\chi_k}-2^{-1-\chi_k}pq\right)\\
&+\Gamma(\chi_k+1)\big(-3pq-2^{-1-\chi_k}(1+2pq+p^{1-\chi_k}+q^{1-\chi_k})\\
&+2p(1+q)(1+p)^{-1-\chi_k}+2q(1+p)(1+q)^{-1-\chi_k}\big)\\
&+(\Gamma(\chi_k)+\Gamma(\chi_k-1))\big(2(1+p)^{1-\chi_k}+2(1+q)^{1-\chi_k}-2^{1-\chi_k}\\
&+(1-2^{1-\chi_k})(p^{1-\chi_k}+q^{1-\chi_k})-3\big)\\
&+2\sum_{\ell\geq 2}\frac{(-1)^\ell}{\ell!}\frac{K_1(-\ell)}{1-p^\ell-q^\ell}\Big(pq(p^\ell+q^\ell)(\ell-1)\Gamma(\chi_k+\ell+1)\\
&+\big((p^\ell+q^\ell)(\ell-2pq\ell-p^{-\ell+2}\ell-q^{-\ell+2}\ell-1)+1\big)\,\Gamma(\chi_k+\ell)\\
&+(1-p^\ell-q^\ell)\Gamma(\chi_k+\ell-1)\Big)
\end{aligned}
$$

$$+ 2\sum_{\ell\geq 1}\frac{(-1)^\ell}{\ell!}\frac{p^{1+\ell}+q^{1+\ell}}{1-p^{1+\ell}-q^{1+\ell}}K_1(\ell+\chi_k-1)K_1(-1-\ell)\Gamma(\ell+\chi_k)$$
$$-\frac{1}{h}\sum_{j\in\mathbb{Z}}(\chi_j-1)G_1^{(P)}(\chi_j-1)(-1+\chi_{k-j})G_1^{(P)}(-1+\chi_{k-j})$$

and $G_2^{(P)}(-1)$ *is given by*

$$-\frac{p^2q^2}{4}-\frac{9pq}{2}-1+\frac{2p(1+q)}{1+p}+\frac{2q(1+p)}{1+q}+\frac{2p^2q}{(1+p)^2}+\frac{2pq^2}{(1+q)^2}$$
$$+2\sum_{\ell\geq 2}\frac{(-1)^\ell K_1(-\ell)}{1-p^\ell-q^\ell}\left((p^\ell+q^\ell)\left(1-p^{-\ell+2}-q^{-\ell+2}-\frac{1}{\ell-1}+pq\ell-3pq\right)\right.$$
$$\left.+\frac{1}{\ell-1}\right)+2\sum_{\ell\geq 1}\frac{(-1)^\ell}{\ell}\frac{p^{1+\ell}+q^{1+\ell}}{1-p^{1+\ell}-q^{1+\ell}}K_1(\ell-1)K_1(-1-\ell)-\frac{p^2q^2}{h}$$
$$-\begin{cases}\dfrac{1}{h\log p}\displaystyle\sum_{j\geq 1}\frac{4rj\pi^2}{\sinh\left(\frac{2rj\pi^2}{\log p}\right)}\left(p^2q^2\left(\frac{2rj\pi}{\log p}\right)^2+p^2q^2\right), & \textit{if } \frac{\log p}{\log q}\in\mathbb{Q};\\ 0, & \textit{if } \frac{\log p}{\log q}\notin\mathbb{Q}.\end{cases}$$

Moreover, if we consider the symmetric case ($p=q=1/2$), we have again the following result with different Fourier coefficients.

Theorem 17. *The variance of the number of 2-protected nodes in symmetric PATRICIA tries satisfies*

$$\frac{\mathbb{V}(X_n^{(P)})}{n}=\frac{G_2^{(P)}(-1)}{\log 2}+\mathscr{F}[G_2^{(P)}](r\log_2 n)+o(1),$$

where for $k\neq 0$ *(when* $\frac{\log p}{\log q}\in\mathbb{Q}$*)*

$$\begin{aligned}G_2^{(P)}(-1+\chi_k)=&\Phi(-1+\chi_k)+\Phi_1(-1+\chi_k)\\ =&\frac{-\Gamma(\chi_k+4)}{256}-\Gamma(\chi_k+3)\left(\frac{5}{128}-2\cdot 3^{-\chi_k-3}\right)\\ &-\Gamma(\chi_k+2)\left(\frac{7}{16}+2\cdot 3^{-\chi_k-2}-2\cdot 3^{-\chi_k-1}\right)\\ &-(\Gamma(\chi_k+1)+3\Gamma(\chi_k)+3\Gamma(\chi_k-1))\left(2-2\cdot 3^{-\chi_k}\right)\\ &+2\sum_{\ell\geq 2}\frac{(-1)^\ell}{\ell!}\frac{(\ell-1)\left(-1+\frac{\ell}{4}+2^{-\ell+1}\right)}{1-2^{-\ell+1}}\cdot\Bigg(\Gamma(\ell+\chi_k-1)\end{aligned}$$

$$
(1-2^{-\ell+1}) + \Gamma(\ell+\chi_k)\Big(1 + 2^{-\ell}\ell - \ell - 2^{-\ell+1}\Big)
$$
$$
+ \Gamma(\ell+\chi_k+1)2^{-\ell-1}(\ell-1)\Big)
$$

and

$$
G_2^{(P)}(-1) = \frac{185}{3456} + 2\sum_{\ell\geq 2}(-1)^{\ell}\frac{(\ell-1)\left(-1+\frac{\ell}{4}+2^{-\ell+1}\right)}{1-2^{-\ell+1}} \cdot \Bigg(2^{-\ell-1}(\ell+1) + \frac{1-2^{-\ell+1}}{\ell-1} - 1\Bigg).
$$

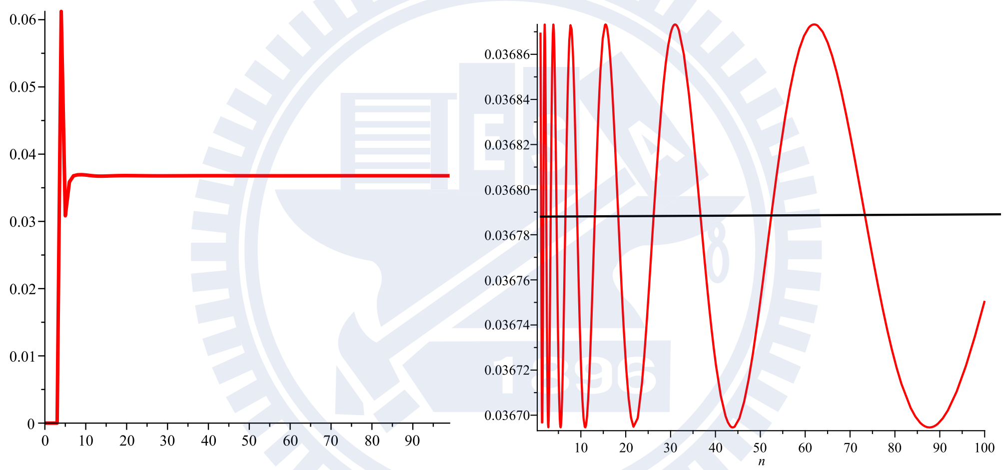


Figure 4.2: Left is a plot of $\mathbb{V}(X_n^{(P)})/n$ for $n = 1 \ldots 100$; right is a plot of the periodic function of the main term in the asymptotic expansion of $\mathbb{V}(X_n^{(P)})/n$.

Remark 20. The above result gives

$$
\frac{G_2^{(P)}(-1)}{\log 2} \approx 0.03678391067977127535980507235349874149194893879\cdots.
$$

which is relatively small; see Figure 4.2. (This is intuitively clear since most contribution for the number of 2-protected nodes in tries is coming from the nodes with only one child and these nodes are removed in the PATRICIA tries.)

Remark 21. As for tries one could also proved a central limit theorem for $X_n^{(P)}$. However, we do not do this because we are going to prove a much stranger result in the next chapter.

Chapter 5

A bivariate central limit theorem

Here, we are going to show a bivariate central limit theorem for the number N_n of internal nodes and the number $X_n^{(T)}$ of 2-protected nodes in a random trie. We first need to compute the covariance. Therefore, recall from Chapter 4 that

$$X_n^{(P)} = X_n^{(T)} - N_n + n - 1.$$

Thus,

$$\mathbb{V}(X_n^{(P)}) = \mathbb{V}(X_n^{(T)}) + \mathbb{V}(N_n) - 2\mathrm{Cov}(N_n, X_n^{(T)}).$$

Rewriting this gives

$$\mathrm{Cov}(N_n, X_n^{(T)}) = \frac{1}{2}\left(\mathbb{V}(X_n^{(T)}) + \mathbb{V}(N_n) - \mathbb{V}(X_n^{(P)})\right).$$

Consequently, by plugging into this the result for N_n from Section 1.3.1 and our results from Chapter 3 and Chapter 4, we obtain

$$\frac{\mathrm{Cov}(N_n, X_n^{(T)})}{n} = \frac{H_2(-1)}{h} + \mathscr{F}[H_2](r\log_{1/p} n) + o(1),$$

where $H(x)$ is some function. We state this as a theorem.

Theorem 18. *The covariance of the number of internal nodes and the number of 2-protected nodes in tries satisfies*

$$\frac{\mathrm{Cov}(N_n, X_n^{(T)})}{n} = \frac{H_2(-1)}{h} + \mathscr{F}[H_2](r\log_{1/p} n) + o(1),$$

where

$$H_2(x) = \frac{G_2^{(T)}(x) + G_2^{(N)}(x) - G_2^{(P)}(x)}{2}.$$

Next, we consider the matrix Σ_n which is defined by

$$n\begin{pmatrix} G_2^{(N)}(-1)/h + \mathscr{F}[G_2^{(N)}](r\log_{1/p} n) & H_2(-1)/h + \mathscr{F}[H_2](r\log_{1/p} n) \\ H_2(-1)/h + \mathscr{F}[H_2](r\log_{1/p} n) & G_2^{(T)}(-1)/h + \mathscr{F}[G_2^{(T)}](r\log_{1/p} n) \end{pmatrix}.$$

We have to show that this matrix is positive definite for all large n. We first show the following lemma.

Lemma 2. *The correlation coefficient of N_4 and $X_4^{(T)}$ is neither -1 nor 1.*

Proof. If the claim is wrong, then we have a, b with $N_4 = aX_4^{(T)} + b$. By looking at the two tries in Figure 5.1 this is clearly impossible. ∎

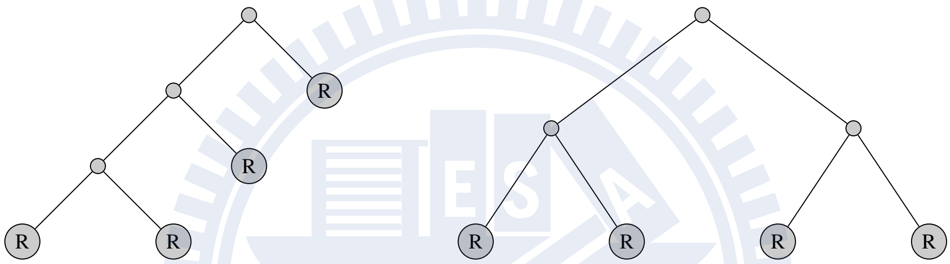


Figure 5.1: The two tries of size 4 from the proof of Lemma 2

Lemma 3. *For all n large enough, we have that Σ_n is positive definite.*

Proof. It is sufficient to show that $\det(\Sigma_n) > 0$ for all n large enough. We will show that in fact $\det(\Sigma_n) = \Omega(n^2)$. In order to show this, note that in the proof of Proposition 3 in [5], the authors showed that the claim holds provided that the correlation coefficient of N_n and $X_n^{(T)}$ is not in $\{-1, 1\}$ for all $n \geq 2$. Thus, the claim follows from the previous lemma. ∎

Consequently, we can consider $\Sigma_n^{-1/2}$ for all n large enough. Our main result in this section is the following bivariate limit law.

Theorem 19. *We have,*

$$\Sigma_n^{-1/2}\begin{pmatrix} N_n - \mathbb{E}(N_n) \\ X_n^{(T)} - \mathbb{E}(X_n^{(T)}) \end{pmatrix} \xrightarrow{d} N(0, I_2),$$

where I_2 denotes the 2×2 unity matrix and $N(0, I_2)$ is the standard two-dimensional normal distribution.

Proof. This follows from the multivariate contraction method with a similar proof as of Theorem 4 in [5]. ∎

Note that this result implies the central limit theorem from Section 3.3. Moreover, due to the above relation between $X_n^{(P)}, X_n^{(T)}$ and N_n, the result also yields the following central limit theorem for the number of 2-protected nodes in PATRICIA tries as consequence.

Theorem 20. *We have,*

$$\frac{X_n^{(P)} - \mathbb{E}(X_n^{(P)})}{\sqrt{\mathbb{V}(X_n^{(P)})}} \xrightarrow{d} N(0,1).$$

Chapter 6

Conclusion

In this chapter, we first summarize our main findings.

In 2012, J. Gaither, H. Homma, M. Sellke, and M. D. Ward [6] derived an asymptotic expansion for the mean of the number of 2-protected nodes in random tries. Recently, J. Gaither and M. D. Ward [7] also gave a similar result for the variance and conjectured a central limit theorem.

The main purpose of this thesis was as follows: first, we re-derived (and corrected) previous results by using a recent systematic method of M. Fuchs, H.-K. Hwang, and V. Zacharovas [4]. Then, we proved the conjectured central limit theorem. Moreover, we derived a bivariate central limit theorem for mean and variance of the number of internal nodes and the number of 2-protected nodes in random tries. This result contains the central limit theorem for PATRICIA tries. Finally, we also derived asymptotic expansions for the number of 2-protected nodes in PATRICIA tries.

Overall, our results complete the analysis of the number of 2-protected nodes in random tries and PATRICIA tries. The reader might wonder how about corresponding results for the random digital search tree which was also introduced in Chapter 1? In fact, the mean for this class of random digital trees was the first instance for which the number of 2-protected nodes in random digital trees was studied; see the paper of R. R.-X. Du and H. Prodinger [1]. Moreover, asymptotic expansions for the variance and a central limit theorem have been also derived; see C.-K. Lee's Ph.D. thesis [5].

Bibliography

[1] R. R.-X. Du and H. Prodinger. Notes on protected nodes in digital search trees. *Applied Mathematics Letters*, 25(6):1025–1028, 2012.

[2] P. Flajolet, X. Gourdon, and P. Dumas. Mellin transforms and asymptotics: Harmonic sums. *Theoretical Computer Science*, 144(1):3–58, 1995.

[3] P. Flajolet, M. Roux, and B. Vallée. Digital tree and memoryless sources: from arithmetics to analysis. *In Proceedings of the 21st International Meeting on Probabilistic, Combinatorial, and Asymptotic Methods in the Analysis of Algorithm (AofA'10), Discrete Mathematics and Theoretical Computer Science Proceedings*, 2010.

[4] M. Fuchs, H.-K. Hwang, and V. Zacharovas. An analytic approach to the asymptotic variance of trie statistics and related structures. *Theoretical Computer Science*, 21(1), 26 pages, 2014.

[5] M. Fuchs and C.-K. Lee. A general central limit theorem for shape parameters of m-ary tries and PATRICIA tries. *Electronic Journal of Combinatorics*, 21(1), 2014.

[6] J. Gaither, Y. Homma, M. Sellke, and M. D. Ward. On the number of 2-protected nodes in tries and suffix trees. *In Proceedings of the 23rd International Meeting on Probabilistic, Combinatorial, and Asymptotic Methods in the Analysis of Algorithm (AofA'12), Discrete Mathematics and Theoretical Computer Science Proceedings*, 2012.

[7] J. Gaither and M. D. Ward. The variance of the number of 2-protected nodes in a trie. In ANALCO, page 43–51, SIAM, 2013.

[8] H.-K. Hwang, M. Fuchs, and V. Zacharovas. Asymptotic variance of random symmetric digital search trees. *Discrete Mathematics and Theoretical Computer Science (Special Issue in Honor of Philippe Flajolet)*, 12(2):103–166, 2010.

[9] P. Jacquet and M. Régnier. Normal limiting distribution of the size of tries. *In Proceedings of the 12th IFIP WG 7.3 International Symposium on Computer Performance Modelling, Measurement and Evaluation*, page 209–223, North-Holland Publishing Co., 1987.

[10] P. Jacquet and W. Szpankowski. Analytical de-Poissonization and its applications. *Theoretical Computer Science*, 201(1):1–62, 1998.

[11] P. Kirschenhofer and H. Prodinger. On some applications of formulae of Ramanujan in the analysis of algorithms. *Mathematika*, 38(1):14–33, 1991.

[12] D. E. Knuth. *The Art of Computer Programming. Volume 3: Searching and Sorting*. Addison Wesley Publishing Co., Second edition, 1998.

[13] C.-K. Lee. *Probabilistic Analysis of Additive Shape Parameters in Random Digital Tree*. National Chiao Tung University, Ph. D. Thesis, 2014.

[14] F. W. J. Olver. *Asymptotic and Special Functions.* Akademic Press, 1974.

[15] M. Régnier and P. Jacquet. New results on the size of tries. *IEEE Transaction on Information Theory*, 35(1):203–205, 1989.