DISSERTATION

Untersuchungen zur durchschnittlichen Gestalt gewisser Baumfamilien

Mit besonderer Berücksichtigung von Anwendungen in der Informatik

ausgeführt zum Zwecke der Erlangung
des akademischen Grades eines Doktors der technischen Wissenschaften
unter der Leitung von

Univ.-Prof. Dr. Helmut Prodinger
Institutsnummer 118
Institut für Algebra und Diskrete Mathematik

eingereicht an der Technischen Universität Wien
Technisch-Naturwissenschaftliche Fakultät

von

Alois Panholzer

Matrikelnummer 9125354
Rudolf-Köppl-Gasse 5/3/3
1220 Wien

Wien, im September 1997 ....

Besonders danken möchte ich meinem Betreuer der Dissertation Prof. Prodinger, der durch sein stetes Bemühen ein zügiges Voranschreiten dieser Arbeit ermöglichte.

Weiters bedanke ich mich bei meiner Lebensgefährtin Sheau-Wen für die Nachsicht und die Aufmunterungen während dieser Zeit.

Auch soll hier meine Familie erwähnt werden, die ebenfalls nicht unwesentlich zum Gelingen dieser Arbeit beitrug.

Kurzfassung

Die vorliegende Arbeit ist der Analyse von bestimmten, in der Informatik relevanten Baumstrukturen gewidmet. Verschiedenste Klassen von Bäumen treten als mathematische Objekte nicht nur direkt oder indirekt in vielen Algorithmen auf, sondern besitzen auch eine Reihe von Anwendungen außerhalb der Informatik.

Da Bäume im wesentlichen die einfachsten nichttrivialen rekursiv definierten Objekte darstellen, läßt sich dieser Sachverhalt bei der Analyse der Verteilung von bestimmten Baumparametern oftmals übersetzen in Rekursionen der betrachteten Größen. Unter Verwendung von – je nach Art der Rekursion unterschiedlich anzusetzenden – Erzeugenden Funktionen erhalten wir daraus Gleichungen für diese Funktionen. Mit Hilfe der sogenannten symbolischen Operatormethode sind diese Gleichungen meist auch direkt aus der formalen Gleichung der Bäume zu erhalten.

Nun kann man versuchen, die so erhaltenen Gleichungen dieser Erzeugenden Funktionen der Verteilungen zu lösen. Dies ist zwar oftmals nicht möglich, aber man kann aus diesen Beziehungen leicht Gleichungen für die Erzeugenden Funktionen der Faktoriellen Momente der betrachteten Größen gewinnen, welche einfacher zu lösen sind.

Schließlich bleibt noch das Ablesen der Koeffizienten, um zu exakten Resultaten für die gesuchten Größen zu gelangen. Gelingt dies nicht bzw. sind die erhaltenen Ergebnisse zu wenig aufschlußreich, kann man auch versuchen, direkt aus der Erzeugenden Funktion das asymptotische Verhalten der betrachteten Parameter zu bestimmen.

Dieses grobe Schema wird im wesentlichen bei der Analyse aller fünf hier behandelten Problemkreise verwendet. Als verwendete mathematische Methoden seien aus der Theorie der formalen Reihen die symbolische Operatormethode und die Inversionsformel von Lagrange erwähnt. Für die Lösung der auftretenden Differentialgleichungen werden verschiedene Methoden aus der Theorie der Gewöhnlichen Differentialgleichungen, wie beispielsweise die Variation der Konstanten, angewandt.

Für das Ablesen der Koeffizienten werden einerseits eine Reihe von Beziehungen für Harmonische Zahlen verwendet, andererseits wird für die Gewinnung von geschlossenen Formen bestimmter Summen Zeilberger's bzw. Gosper's Algorithmus benützt. Um asymptotische Ergebnisse zu erhalten, werden Methoden angewandt, die direkt von den exakten Resultaten ausgehen, wie beispielsweise die Euler'sche Summenformel oder das Arcussinusgesetz. Bei Bedarf wird allerdings ausgehend von den Erzeugenden Funktionen die Singularitätenanalyse durchgeführt. Weiters finden für die Bestimmung einer Grenzverteilung probabilistische Verfahren Verwendung, wie beispielsweise der Grenzwertsatz von Lévy.

Es sei auch angemerkt, daß in dieser Arbeit ausgiebig von den Möglichkeiten der zur Verfügung stehenden Formelmanipulationsprogramme, im speziellen des Softwarepaketes `Maple`, Gebrauch gemacht wurde. Ohne Softwareunterstützung ist die Berechnung eines Großteils der hier gebrachten Ergebnisse nicht mehr durchführbar.

Kapitel 1 befaßt sich mit der Analyse der *p-trees*, einer Datenstruktur, die bei der Implementierung von Prioritätswarteschlangen Verwendung findet. In solch einer Prioritätswarteschlange müssen die

ankommenden Elemente je nach ihrem Gewicht so eingefügt werden, daß auf das Element mit der höchsten Priorität rasch zugegriffen werden kann.

Die einzigen, aus den Siebziger-Jahren stammenden, Arbeiten über die *p-trees* von Jonasson und Dahl bzw. von Nevalainen und Teuhola werden unter Verwendung der heute zur Verfügung stehenden Methoden stark erweitert. Im speziellen werden Ergebnisse über die Höhe des Elementes der höchsten Priorität, die Anzahl der Vergleiche beim Einfügen eines Elementes bestimmter Priorität, die Höhe und die Anzahl der Nachfolger eines Elementes bestimmter Priorität erhalten. Weiters wird noch ein Algorithmus behandelt, der zum Aufteilen des Baumes an einer bestimmten Stelle dient. Insbesondere betrachtet man die Verteilung der Anzahl der Kinder eines speziellen Knotens.

Als Ergänzung des Artikels von Martínez und Prodinger über eine spezielle Art der balanzierten Suchbäume ist Kapitel 2 gedacht. Die wegen des hohen Rechenaufwandes in jenem Artikel ausgesparte Varianz der Höhe spezieller Knoten in dieser Baumstruktur wird hier explizit bzw. asymptotisch angegeben. Dies geschieht unter Zuhilfenahme von Beziehungen, welche in Kapitel 1 hergeleitet wurden.

In Kapitel 3 werden Fragestellungen bezüglich der Anzahl der Vorgänger bzw. Nachfolger von entsprechend der Reihenfolge gewisser Traversierungen markierten Knoten in Binärbäumen behandelt. Es werden exakte und asymptotische Ergebnisse für die drei in der Praxis meist verwendeten rekursiven Algorithmen zum Durchlaufen eines Baumes, Inorder-, Preorder- und Postorder-Traversierung geliefert.

Weiters wird ein in den Achtziger-Jahren von Kirschenhofer behandeltes Problem der Bestimmung der Höhe von Blättern in Binärbäumen untersucht, und es werden neben dem schon bekannten Erwartungswert auch exakte bzw. asymptotische Resultate für die Varianz geliefert.

Betrachtungen über die Anzahl der Kinder, also der direkten Nachfolger, eines Knotens in Bäumen werden in Kapitel 4 angestellt. Zunächst werden Ergebnisse von Prodinger bzw. Rote, betreffend die Anzahl der Binärbäume mit einer vorgegebenen Verteilung von Knoten bestimmter Kinderzahl auf eine Art wiedergewonnen, die eine Erweiterung der Fragestellungen auf viel allgemeinere Baumfamilien, hin zu den sogenannten einfach erzeugten Baumfamilien, zuläßt.

Anschließend wird ein von Sprugnoli behandeltes Problem wieder aufgegriffen. Er betrachtete die Anzahl der traversierten Binärbäume, bei der ein entsprechend der Reihenfolge der Traversierung markierter Knoten kein, ein nach links führendes, ein nach rechts führendes oder zwei Kinder hat. So werden durch eine einheitliche Methode die schon bekannten Resultate für die Inorder-Traversierung ergänzt durch entsprechende Ergebnisse für preorder- und postorder-traversierte Binärbäume.

Schließlich befaßt sich Kapitel 5 mit der Analyse von Multiple Quickselect, einem Algorithmus ähnlich dem Sortierprogramm Quicksort, der für die Ermittlung von Reihenfolgenstatistiken mehrerer Elemente herangezogen wird. Obgleich dieser rekursive Algorithmus angewandt wird auf in einer Tabelle gespeicherte Daten, besteht eine Verbindung zu den Binären Suchbäumen. Dieser Zusammenhang erlaubt es, die für die Leistungsfähigkeit von Multiple Quickselect interessanten Parameter der Anzahl der rekursiven Aufrufe bzw. der Anzahl der Vergleiche im Baummodell zu studieren und die hierfür entwickelten Methoden zu verwenden.

Weiters wird eine Modifikation des originalen Algorithmus, die sogenannte Median-of-three Partition, untersucht. Zu Vergleichszwecken werden auch Ergebnisse für die betrachteten Parameter im Modell der Binärbäume geliefert.

Inhaltsverzeichnis

Liste der verwendeten Symbole

$\log n$ *natürlicher Logarithmus*
$\log_e n$

$\binom{n}{k}$ *Binomialkoeffizient*
Anzahl der k-elementigen Teilmengen einer n-elementigen Menge

$\left[{n \atop k}\right]$ *vorzeichenlose Stirling-Zahlen erster Art*
Anzahl der n-elementigen Permutationen mit genau k Zyklen

$n^{\underline{k}}$ *fallende Faktorielle*
$n \cdot (n-1) \cdots (n-k+1)$

$n^{\overline{k}}$ *steigende Faktorielle*
$n \cdot (n+1) \cdots (n+k-1)$

H_n *Harmonische Zahlen*
$\sum_{k=1}^{n} \frac{1}{k}$

$H_n^{(m)}$ *Harmonische Zahlen der Ordnung m*
$\sum_{k=1}^{n} \frac{1}{k^m}, \quad H_n^{(1)} = H_n$

$\Gamma(z)$ *Gammafunktion*
$\int_0^\infty t^{z-1} e^{-t} dt$ für $\mathrm{Re}\, z > 0$, kann eindeutig fortgesetzt werden zu einer analytischen Funktion für $z \in \mathbb{C} \setminus \{0, -1, -2, \cdots\}$

$\Psi(z)$ *Psifunktion (Digammafunktion)*
$\frac{d}{dz} \log \Gamma(z)$

$\zeta(z)$ *Zetafunktion*
$\sum_{n \geq 1} \frac{1}{n^z}$ für $\mathrm{Re}\, z > 1$, kann eindeutig fortgesetzt werden zu einer analytischen Funktion für $z \in \mathbb{C} \setminus \{1\}$

$\mathrm{dilog}\, z$ *Dilogarithmus*
$\int_1^z \frac{1}{1-t} \log t \, dt$

$\mathrm{L}_m(z)$ *Polylogarithmen*
$\sum_{n \geq 1} \frac{z^n}{n^m}$ für $|z| < 1$

$\delta_{n,k}$ *Kroneckerdelta*
$\begin{cases} 1 & : \quad n=k \\ 0 & : \quad \text{sonst} \end{cases}$

$[A]$ *Wahrheitsfunktion*

$$\begin{cases} 1 & : \text{ Aussage } A \text{ ist wahr} \\ 0 & : \text{ sonst} \end{cases}$$

$[z^n]f(z)$ Koeffizient von z^n in $f(z)$

$\frac{\partial}{\partial z} f(z,u)\big|_{z=a}$ partielle Ableitung von $f(z,u)$ nach z und anschließendes Auswerten an der Stelle $z = a$

γ *Eulersche Konstante*

$$\gamma = \lim_{n\to\infty}\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} - \log n\right) = 0.5772156\ldots$$

Kapitel 1

Eine Analyse der p-trees

1.1 Einleitung

Eine zur Implementation von Prioritätswarteschlangen geeignete Datenstruktur sind die p-trees. Solch eine Art von Warteschlange muß die zwei grundlegenden Operationen des Einfügens eines Elements beliebiger Priorität (`Insert`) und des Entfernen des Elements mit der höchsten Priorität (`Delete`) unterstützen.

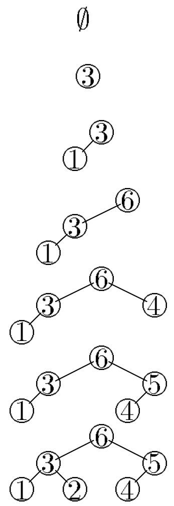
Ein p-tree ist nun entweder leer oder er besteht aus einer Folge von Knoten mit nicht anwachsenden Schlüsseln, dem sogenannten "linken Pfad", mit der Eigenschaft, daß jedem Knoten aus dem linken Pfad außer dem letzten, ein möglicherweiser leerer p-tree, der "rechte Unterbaum", zugeordnet ist. Ist nun z ein Knoten des linken Pfades mit Schlüssel l und besitzt dieser einen linken Nachfolger x mit Schlüssel k, so haben alle Knoten y_i des mit z verbundenen rechten Pfades Schlüssel s_i, die zwischen k und l liegen. Genauer gilt für die Schlüssel s_i der Knoten y_i des rechten Pfades von z, $k \leq s_i < l$.

Um in den p-tree T ein neues Element p einzufügen bedient man sich des nachfolgend angeführten rekursiven Algorithmus `Insert`:

- Wenn T leer ist, oder die Wurzel von T keinen größeren Schlüssel als p hat, so wird p zur neuen Wurzel und T ist sein linker Unterbaum.
- Sonst sucht man entlang des linken Pfades von T nach dem ersten Knoten x, der keinen größeren Schlüssel als p hat.
  - Falls man keinen solchen Knoten findet, wird p an den linken Pfad als linkes Blatt angehängt.
  - Ansonsten sei z der Vorgänger von x und es gilt, daß der Schlüssel von p zwischen denen von x und z liegt. In diesem Fall wird der Algorithmus `Insert` für den Knoten p rekursiv auf den rechten Unterbaum von z angewendet.

Als Beispiel für die Arbeitsweise des Algorithmus `Insert` diene der Aufbau eines p-tree aus der Permutation 3 1 6 4 5 2, wie in Abbildung 1.1 gezeigt.

Nachfolgend untersuchen wir Algorithmen für die p-trees und betrachten die Verteilungen bestimmter für diese Datenstruktur interessanter Parameter.



Schritt 0	(leerer Baum)
Schritt 1	(Einfügen von 3)
Schritt 2	(Einfügen von 1)
Schritt 3	(Einfügen von 6)
Schritt 4	(Einfügen von 4)
Schritt 5	(Einfügen von 5)
Schritt 6	(Einfügen von 2)

Abbildung 1.1: Aufbau eines p-tree.

1.2 Die verwendeten mathematischen Methoden

Für die mathematische Analyse werden wir nicht den in [17] bestrittenen Weg der Unterscheidung in linken Pfad und rechte Unterbäume einschlagen, da man damit schnell an Grenzen stößt, sondern lassen uns von einer Idee von H. Prodinger leiten. Man arbeitet dabei mit drei Familien von Objekten $\mathcal{A}$, $\mathcal{B}$ und $\mathcal{C}$. Während die $\mathcal{A}$-Familie die gewöhnlichen p-trees sind, stellt die $\mathcal{B}$-Familie p-trees dar, die aus einem zusätzlichen Element “$+\infty$” erzeugt werden, während die $\mathcal{C}$-Familie p-trees darstellen, welche aus einem zusätzlichen Element “$-\infty$” erzeugt werden.

Diese zusätzlichen Elemente werden bei der Analyse von Parametern nicht als solche gezählt. Das heißt wenn wir von der Größe eines Objektes sprechen, meinen wir die Anzahl der Knoten dieses Objektes ohne die Elemente $-\infty$ bzw. $+\infty$.

Bei unserer Analyse gehen wir von dem Modell aus, daß alle $n!$ Möglichkeiten von Permutationen der Zahlen von 1 bis n gleichwahrscheinlich für die Erzeugung von Objekten der Familien $\mathcal{A}$, $\mathcal{B}$ und $\mathcal{C}$ der Größe n sind.

Als entscheidend für alles weitere erweisen sich die Zerlegungen der Familien $\mathcal{A}$, $\mathcal{B}$ und $\mathcal{C}$ aufgrund des ersten Elementes k einer beliebigen Permutation der Zahlen $1, 2, \ldots, n$, welche in den Abbildungen 1.2, 1.3 und 1.4 dargestellt sind.

Daraus erhalten wir für die von uns untersuchten Parameter Rekursionen, welche sich in den entsprechenden erzeugenden Funktionen als ein System von Differentialgleichungen widerspiegelt. Obwohl sich aufgrund der Komplexität der erzeugenden Funktionen für die betrachteten Verteilungen diese meist nicht zu erhalten sind, lassen sich daraus jedoch zumindest die gewünschten Erwartungswerte bekommen.

Auf alle Fälle ist die hier durchgeführte Analyse der p-trees um vieles weitreichender als die vor 20 Jahren in [17] vorgenommene, da wir die uns heute zur Verfügung stehenden Hilfen beim Manipulieren komplexer Formeln ausgiebig verwenden. Insbesondere sei das Softwarepaket MAPLE erwähnt.

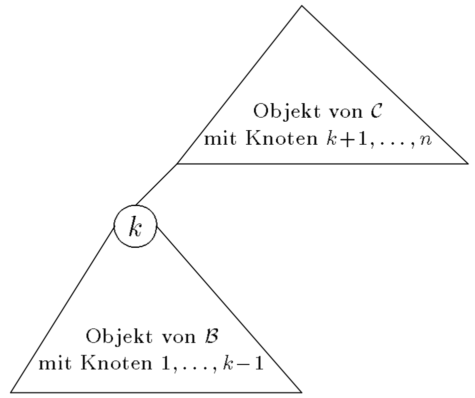


Abbildung 1.2: Zerlegung der Familie $\mathcal{A}$

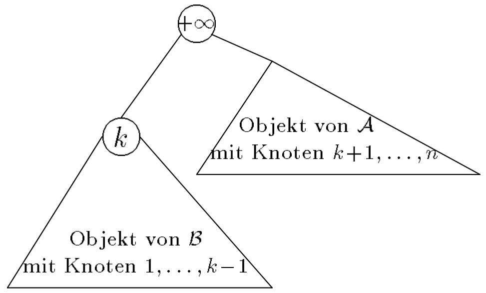


Abbildung 1.3: Zerlegung der Familie $\mathcal{B}$

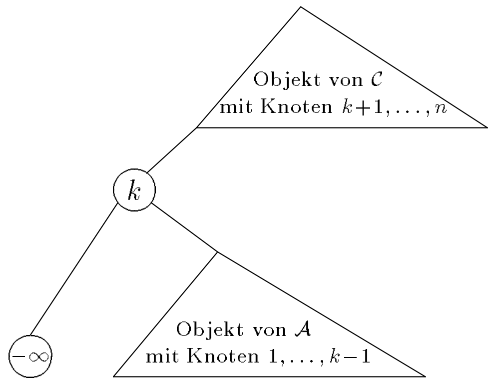


Abbildung 1.4: Zerlegung der Familie $\mathcal{C}$

1.3 Die Länge des linken Pfades

1.3.1 Die Verteilungen

Bei der Verwendung der p-trees als Datenstruktur für Prioritätswarteschlangenbehandlung ist es von Bedeutung, das Element mit der höchsten Priorität, das heißt also das Element mit kleinstem Schlüssel, bei Bedarf aus der Warteschlange zu entfernen. In einem p-tree ist dieses gesuchte Element das Blatt des oben definierten linken Pfades ausgehend von der Wurzel.

Nachfolgend behandeln wir die Verteilungen der Höhe des Blattes höchster Priorität in p-trees. In unseren Objektfamilien $\mathcal{A}$, $\mathcal{B}$ und $\mathcal{C}$ bedeutet dies, wir betrachten die Verteilungen der Längen des linken Pfades in den jeweiligen Objekten. Damit meinen wir die Anzahl der Knoten, die auf dem direkten Weg von der Wurzel bis zum sich am weitesten links befindlichen Blatt liegen. Aus technischen Gründen wird in der Objektfamilie $\mathcal{B}$ das Element $+\infty$ für die Länge des linken Pfades gezählt, während in der Objektfamilie $\mathcal{C}$ das Element $-\infty$ nicht gezählt wird.

Zerlegt man nun die Familien $\mathcal{A}$, $\mathcal{B}$ und $\mathcal{C}$ wie oben beschrieben aufgrund des ersten Elements k einer Zufallspermutation mit n Elementen, so ergibt sich für die Wahrscheinlichkeiten $A_{n,m}$, $B_{n,m}$ bzw. $C_{n,m}$, daß ein Objekt aus obigen Familien der Größe n einen linken Pfad der Länge m hat, folgende Rekursion:

$$A_{n,m} = \frac{1}{n}\sum_{k=1}^{n}\sum_{i=0}^{m} C_{n-k,i}\, B_{k-1,m-i} \quad \text{für } n \geq 1\,, \tag{1.1a}$$

$$B_{n,m} = \frac{1}{n}\sum_{k=1}^{n} B_{k-1,m-1} \quad \text{für } n \geq 1\,, \tag{1.1b}$$

$$C_{n,m} = \frac{1}{n}\sum_{k=1}^{n} C_{n-k,m-1} \quad \text{für } n \geq 1\,, \tag{1.1c}$$

$$A_{0,m} = \delta_{0,m}\,, \quad B_{0,m} = \delta_{1,m} \text{ und } C_{0,m} = \delta_{0,m}\,.$$

Dabei bezeichnet $\delta_{n,m}$ das sogenannte Kronecker-Delta, welches definiert ist durch

$$\delta_{n,m} = \begin{cases} 1 & \text{wenn } n = m\,, \\ 0 & \text{wenn } n \neq m\,. \end{cases}$$

Definiert man für $n \geq 0$ die erzeugenden Funktionen

$$A_n(v) = \sum_{m\geq 0} A_{n,m} v^m\,, \quad B_n(v) = \sum_{m\geq 0} B_{n,m} v^m\,, \quad C_n(v) = \sum_{m\geq 0} C_{n,m} v^m$$

und multipliziert obige Gleichungen (1.1) mit v^m, so ergibt sich nach dem Summieren über alle $m \geq 0$

$$nA_n(v) = \sum_{k=1}^{n} C_{n-k}(v) B_{k-1}(v) \quad \text{für } n \geq 1\,, \tag{1.2a}$$

$$nB_n(v) = v\sum_{k=1}^{n} B_{k-1}(v) \quad \text{für } n \geq 1\,, \tag{1.2b}$$

$$nC_n(v) = v\sum_{k=1}^{n} C_{k-1}(v) \quad \text{für } n \geq 1\,, \tag{1.2c}$$

$$A_0(v) = 1\,, \quad B_0(v) = v \text{ und } C_0(v) = 1\,.$$

Führt man weiters die erzeugenden Funktionen in zwei Variablen

$$A(z,v) = \sum_{n\geq 0} A_n(v)z^n\,, \quad B(z,v) = \sum_{n\geq 0} B_n(v)z^n \text{ und } C(z,v) = \sum_{n\geq 0} C_n(v)z^n$$

ein, so erhält man aus (1.2) durch Multiplizieren dieser Gleichungen mit z^{n-1} und Aufsummieren über alle $n \geq 1$ das System von Differentialgleichungen

$$\frac{\partial}{\partial z}A(z,v) = B(z,v)\,C(z,v)\,, \tag{1.3a}$$

$$\frac{\partial}{\partial z}B(z,v) = \frac{v}{1-z}B(z,v)\,, \tag{1.3b}$$

$$\frac{\partial}{\partial z}C(z,v) = \frac{v}{1-z}C(z,v)\,, \tag{1.3c}$$

mit den Anfangswerten

$$A(0,v) = 1\,, \quad B(0,v) = v \text{ und } C(0,v) = 1\,.$$

Aus (1.3b) erhalten wir durch Umformen

$$\frac{d\,(B(z,v))}{B(z,v)} = d\,(\log B(z,v)) = \frac{v}{1-z}dz\,. \tag{1.4}$$

Integration ergibt

$$\log B(z,v) = v\log\left(\frac{1}{1-z}\right) + c(v) \tag{1.5}$$

mit einer nur von v abhängigen Funktion $c(v)$ und weiter

$$B(z,v) = \frac{g(v)}{(1-z)^v} \tag{1.6}$$

mit einer Funktion $g(v)$.

Einsetzen in die Anfangsbedingungen ergibt

$$B(0,v) = g(v) = v\,,$$

woraus man schließlich als Lösung für $B(z,v)$

$$B(z,v) = \frac{v}{(1-z)^v} \tag{1.7}$$

erhält.

Analoges Vorgehen für (1.3c) führt zu

$$C(z,v) = \frac{g(v)}{(1-z)^v} \tag{1.8}$$

mit einer Funktion $g(v)$. Diese wird durch Einsetzen in die Anfangsbedingung

$$C(0,v) = g(v) = 1$$

bestimmt, und man bekommt so als Lösung für $C(z,v)$

$$C(z,v) = \frac{1}{(1-z)^v}\,. \tag{1.9}$$

Damit ergibt sich aus (1.3a) für $A(z,v)$ die Beziehung

$$\frac{\partial}{\partial z}A(z,v) = \frac{v}{(1-z)^{2v}}\,. \tag{1.10}$$

Integration führt zu

$$A(z,v) = \frac{v}{(2v-1)(1-z)^{2v-1}} + g(v)\,, \tag{1.11}$$

wobei $g(v)$ wegen der Anfangswerte durch

$$A(0,v) = \frac{v}{2v-1} + g(v) = 1$$

bestimmt ist.

Daraus ergibt sich als Lösung für $A(z,v)$

$$A(z,v) = \frac{1}{1-2v}\left(1 - v - \frac{v}{(1-z)^{2v-1}}\right)\,. \tag{1.12}$$

Bezeichnen wir die vorzeichenlosen Stirling-Zahlen 1. Art mit $\left[{n \atop m}\right]$, so sind diese gegeben durch die Beziehung (siehe beispielsweise [4] oder [12]):

$$\sum_{n\geq 0}\sum_{m=0}^{n} \left[{n \atop m}\right] \frac{z^n}{n!} v^m = \frac{1}{(1-z)^v}\,. \tag{1.13}$$

Dies ist nun beim Ablesen der Koeffizienten von (1.7), (1.9) und (1.10) hilfreich und wir erhalten

$$\begin{aligned} A_{n,m} &= [z^n v^m] \int_0^z \left(\frac{\partial}{\partial t}A(t,v)\right) dt = \frac{1}{n}[z^{n-1}v^m]\frac{\partial}{\partial z}A(z,v) = \frac{1}{n}[z^{n-1}v^m]\frac{v}{(1-z)^{2v}} \\ &= \frac{1}{n}[z^{n-1}v^{m-1}]\frac{1}{(1-z)^{2v}} = \frac{2^{m-1}}{n}[z^{n-1}v^{m-1}]\frac{1}{(1-z)^v} = \frac{2^{m-1}\left[{n-1 \atop m-1}\right]}{n!}\,, \end{aligned} \tag{1.14a}$$

$$B_{n,m} = [z^n v^m]\frac{v}{(1-z)^v} = [z^n v^{m-1}]\frac{1}{(1-z)^v} = \frac{\left[{n \atop m-1}\right]}{n!}\,, \tag{1.14b}$$

$$C_{n,m} = [z^n v^m]\frac{1}{(1-z)^v} = \frac{\left[{n \atop m}\right]}{n!}\,. \tag{1.14c}$$

Zusammengefaßt liefert dies

Satz 1.3.1. *Die Wahrscheinlichkeitsfunktionen $A_{n,m}$, $B_{n,m}$ und $C_{n,m}$ der Längen des linken Pfades in Objekten der Familien $\mathcal{A}$, $\mathcal{B}$ und $\mathcal{C}$ der Größe n, das heißt also die Wahrscheinlichkeiten, daß ein Objekt der Größe n eine linke Pfadlänge m besitzt, sind gegeben durch*

$$\boxed{A_{n,m} = \frac{2^{m-1}\left[{n-1 \atop m-1}\right]}{n!} \quad \text{für } n \geq 1\,.} \tag{1.15a}$$

$$\boxed{B_{n,m} = \frac{\left[{n \atop m-1}\right]}{n!} \quad \text{für } n \geq 1\,.} \tag{1.15b}$$

$$\boxed{C_{n,m} = \frac{\left[{n \atop m}\right]}{n!} \quad \text{für } n \geq 1\,.} \tag{1.15c}$$

1.3.2 Die Erwartungswerte

Bezeichnen wir mit A_n, B_n und C_n die Erwartungswerte der Längen der linken Pfade in Objekten der Größe n aus den Familien $\mathcal{A}$, $\mathcal{B}$ und $\mathcal{C}$.

Definieren wir weiters

$$A(z) = \frac{\partial}{\partial v}A(z,v)\bigg|_{v=1}\,, \quad B(z) = \frac{\partial}{\partial v}B(z,v)\bigg|_{v=1}\,, \quad C(z) = \frac{\partial}{\partial v}C(z,v)\bigg|_{v=1}\,, \tag{1.16}$$

dann ergeben sich die gesuchten Erwartungswerte als Koeffizienten dieser Funktionen $A(z)$, $B(z)$ und $C(z)$. Das heißt also

$$A_n = [z^n]A(z)\,, \quad B_n = [z^n]B(z)\,, \quad C_n = [z^n]C(z)\,. \tag{1.17}$$

Durch Ableiten der Gleichungen (1.7), (1.9) und (1.10) nach v und Auswerten bei $v = 1$ bekommen wir

$$\begin{aligned}\frac{\partial}{\partial z}A(z) &= \frac{\partial}{\partial v}\left(\frac{\partial}{\partial z}A(z,v)\right)\bigg|_{v=1} = \frac{\partial}{\partial v}\left(\frac{v}{(1-z)^{2v}}\right)\bigg|_{v=1}\\ &= \left(\frac{1}{(1-z)^{2v}} + \frac{2v}{(1-z)^{2v}}\log\left(\frac{1}{1-z}\right)\right)\bigg|_{v=1}\\ &= \frac{2}{(1-z)^2}\log\left(\frac{1}{1-z}\right) + \frac{1}{(1-z)^2}\,,\end{aligned} \tag{1.18a}$$

$$\begin{aligned}B(z) &= \frac{\partial}{\partial v}B(z,v)\bigg|_{v=1} = \left(\frac{1}{(1-z)^v} + \frac{v}{(1-z)^v}\log\left(\frac{1}{1-z}\right)\right)\bigg|_{v=1}\\ &= \frac{1}{1-z} + \frac{1}{1-z}\log\left(\frac{1}{1-z}\right)\,,\end{aligned} \tag{1.18b}$$

$$C(z) = \frac{\partial}{\partial v} C(z,v)\Big|_{v=1} = \frac{1}{(1-z)^v} \log\left(\frac{1}{1-z}\right)\Big|_{v=1} = \frac{1}{1-z} \log\left(\frac{1}{1-z}\right). \tag{1.18c}$$

Die Harmonischen Zahlen H_n bzw. $H_n^{(m)}$ seien wie üblich definiert durch

$$H_n = \sum_{k=1}^{n} \frac{1}{k} \quad \text{und} \quad H_n^{(m)} = \sum_{k=1}^{n} \frac{1}{k^m}.$$

Für das Ablesen der Koeffizienten von (1.18), vor allem aber von später auftretenden Funktionen, benötigen wir folgende Beziehungen zwischen speziellen Funktionen und den Harmonischen Zahlen (man findet diese teilweise in [11], vollständig in [50]):

$$[z^n]\,\frac{1}{(1-z)^{\alpha+1}} \log\left(\frac{1}{1-z}\right) = (H_{n+\alpha} - H_\alpha)\binom{n+\alpha}{n}, \tag{1.19a}$$

$$[z^n]\,\frac{1}{(1-z)^{\alpha+1}} \log^2\left(\frac{1}{1-z}\right) = \left((H_{n+\alpha} - H_\alpha)^2 - \left(H^{(2)}_{n+\alpha} - H^{(2)}_\alpha\right)\right)\binom{n+\alpha}{n}, \tag{1.19b}$$

$$\begin{aligned}[z^n]\,\frac{1}{(1-z)^{\alpha+1}} \log^3\left(\frac{1}{1-z}\right) = \Big((H_{n+\alpha} - H_\alpha) - 3\,(H_{n+\alpha} - H_\alpha)\left(H^{(2)}_{n+\alpha} - H^{(2)}_\alpha\right) + \\ + 2\left(H^{(3)}_{n+\alpha} - H^{(3)}_\alpha\right)\Big)\binom{n+\alpha}{n},\end{aligned} \tag{1.19c}$$

$$\begin{aligned}[z^n]\frac{1}{(1-z)^{\alpha+1}} \log^4\left(\frac{1}{1-z}\right) = \Big((H_{n+\alpha} - H_\alpha)^4 - 6\,(H_{n+\alpha} - H_\alpha)^2\left(H^{(2)}_{n+\alpha} - H^{(2)}_\alpha\right) + \\ + 8\,(H_{n+\alpha} - H_\alpha)\left(H^{(3)}_{n+\alpha} - H^{(3)}_\alpha\right) + \\ + 3\left(H^{(2)}_{n+\alpha} - H^{(2)}_\alpha\right)^2 - 6\left(H^{(4)}_{n+\alpha} - H^{(4)}_\alpha\right)\Big)\binom{n+\alpha}{n}.\end{aligned} \tag{1.19d}$$

Damit bekommt man aus (1.18)

$$\begin{aligned}A_n &= [z^n]A(z) = \frac{1}{n}[z^{n-1}]\frac{\partial}{\partial z}A(z) \\ &= \frac{2}{n}[z^{n-1}]\frac{1}{(1-z)^2}\log\left(\frac{1}{1-z}\right) + \frac{1}{n}[z^{n-1}]\frac{1}{(1-z)^2} = \frac{2}{n}\left(n\left(H_n - 1\right)\right) + 1 \\ &= 2H_n - 1\,,\end{aligned} \tag{1.20a}$$

$$B_n = [z^n]B(z) = [z^n]\frac{1}{1-z} + [z^n]\frac{1}{1-z}\log\left(\frac{1}{1-z}\right) = H_n + 1\,, \tag{1.20b}$$

$$C_n = [z^n]C(z) = [z^n]\frac{1}{1-z}\log\left(\frac{1}{1-z}\right) = H_n\,. \tag{1.20c}$$

Zusammengefaßt ergibt sich also

Satz 1.3.2. *Die Erwartungswerte A_n, B_n und C_n der Längen der linken Pfade in Objekten der Größe n aus den Familien $\mathcal{A}$, $\mathcal{B}$ und $\mathcal{C}$ sind gegeben durch*

$$\boxed{\begin{array}{l} A_n = 2H_n - 1 \quad \text{für } n \geq 1\,, \\ A_0 = 0\,. \end{array}} \tag{1.21a}$$

$$\boxed{\begin{array}{l} B_n = H_n + 1 \quad \text{für } n \geq 1\,, \\ B_0 = 1\,. \end{array}} \tag{1.21b}$$

$$\boxed{\begin{array}{l} C_n = H_n \quad \text{für } n \geq 1\,, \\ C_0 = 0\,. \end{array}} \tag{1.21c}$$

Das Resultat für den Erwartungswert der Länge des linken Pfades findet sich bereits in [17].

1.3.3 Die Varianz

Bezeichnen wir hier mit $\breve{A}_n$ die zweiten Faktoriellen Momente der Längen der linken Pfade in p-trees der Größe n.

Weiters definieren wir uns

$$\breve{A}(z) = \frac{\partial^2}{\partial v^2} A(z,v)\bigg|_{v=1} . \tag{1.22}$$

Daraus erhalten wir die Hilfsgrößen $\breve{A}_n$ mittels

$$\breve{A}_n = [z^n]\breve{A}(z)\,. \tag{1.23}$$

Differenzieren wir also (1.10) zweimal nach v und setzen wir anschließend $v = 1$, bekommen wir

$$\frac{\partial}{\partial z}\breve{A}(z) = \frac{4}{(1-z)^2}\log^2\left(\frac{1}{1-z}\right) + \frac{4}{(1-z)^2}\log\left(\frac{1}{1-z}\right) . \tag{1.24}$$

Ablesen der Koeffizienten von (1.24) führt unter Zuhilfenahme der Beziehungen (1.19) zu

Hilfssatz 1.3.3. *Das zweite Faktorielle Moment $\breve{A}_n$ der Länge des linken Pfades in p-trees der Größe n, also in Objekten der Größe n aus der Familie $\mathcal{A}$, ist gegeben durch*

$$\boxed{\begin{array}{l} \breve{A}_n = 4H_n^2 - 4H_n - 4H_n^{(2)} + 4 \quad \text{für } n \geq 1\,, \\ \breve{A}_0 = 0\,. \end{array}} \tag{1.25}$$

Bezeichnen wir die Varianz der Länge des linken Pfades in p-trees der Größe n mit $\hat{A}_n$, so läßt sich diese aufgrund der Beziehung

$$\hat{A}_n = \breve{A}_n + A_n - A_n^2$$

leicht aus den vorangegangenen Resultaten erhalten.

Wir bekommen also aus (1.21) und (1.25)

Satz 1.3.4. *Die Varianz $\hat{A}_n$ der Länge des linken Pfades in p-trees der Größe n, also in Objekten der Größe n aus der Familie $\mathcal{A}$, ist gegeben durch*

$$\boxed{\begin{array}{l} \hat{A}_n = 2H_n - 4H_n^{(2)} + 2 \quad \text{für } n \geq 1\,, \\ \hat{A}_0 = 0\,. \end{array}} \tag{1.26}$$

1.3.4 Asymptotische Betrachtungen

Es zeigt sich, daß die Folge der Verteilungen der Längen der linken Pfade der p-trees für wachsende Baumgröße n schwach gegen eine Normalverteilung konvergiert. Genauer gilt

Satz 1.3.5. *Die Folge der Zufallsvariablen*

$$\Omega_n^* = \frac{\Omega_n - \mu_n}{\sigma_n}$$

ist schwach konvergent gegen die standardisierte Normalverteilung $\mathcal{N}(0,1)$ mit Mittelwert 0 und Varianz 1. Dabei bezeichnen die Zufallsvariablen Ω_n die Verteilungen der Längen der linken Pfade in p-trees der Größe n mit Mittelwert μ_n und Varianz σ_n^2. Es gilt also

$$\mathbb{P}\left(a < \Omega_n^* < b\right) \to \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{t^2}{2}} dt$$

und weiters

$$\mu_n = 2\log n + \mathcal{O}(1) \;\text{ und }\; \sigma_n^2 = 2\log n + \mathcal{O}(1)$$

für $n \to \infty$.

Beweis. Aus den exakten Formeln (1.21a) und (1.26) für den Erwartungswert A_n bzw. die Varianz $\hat{A}_n$ der Längen der linken Pfade in p-trees erhält man unter Benützung der später angegebenen asymptotischen Entwicklungen der Harmonischen Zahlen (1.92) unmittelbar

$$\mu_n = A_n = 2\log n + \mathcal{O}(1) \;\text{ und }\; \sigma_n^2 = \hat{A}_n = 2\log n + \mathcal{O}(1)\,.$$

Zum Beweis des Satzes verwenden wir nun die von Flajolet/Soria in [8] beschriebene Methode. Wir gehen dazu von der Folge der charakteristischen Funktionen $\phi_{\Omega_n^*}(t) = \mathbb{E}(e^{i\Omega_n^* t})$ der Ω_n^* aus und zeigen, daß diese Folge punktweise gegen die charakteristische Funktion der standardisierten Normalverteilung konvergiert, mit anderen Worten: wir werden zeigen, daß

$$\phi_{\Omega_n^*}(t) \to e^{-\frac{t^2}{2}}$$

für $n \to \infty$ gilt. Nach dem sogenannten Stetigkeitssatz von Levy (siehe beispielsweise [3]) kann nun von der punktweisen Konvergenz der charakteristischen Funktionen auf die schwache Konvergenz der entsprechenden Verteilungsfunktionen geschlossen werden. Genauer verwenden wir

Satz 1.3.6 (Stetigkeitssatz von Levy). *Seien $\{\mu_n, 1 \le n < \infty\}$ Wahrscheinlichkeitsmaße auf $\mathbb{R}$ mit charakteristischen Funktionen $\{\phi_n(t), 1 \le n < \infty\}$. Wenn die Bedingungen*

- *$\phi_n(t)$ konvergiert punktweise in ganz $\mathbb{R}$ gegen die Grenzfunktion $\phi_\infty(t)$,*
- *$\phi_\infty(t)$ ist stetig bei $t = 0$*

erfüllt sind, dann gilt

- *μ_n konvergiert schwach gegen μ_∞, wobei μ_∞ ein Wahrscheinlichkeitsmaß darstellt,*

- $\phi_\infty(t)$ *ist die charakteristische Funktion von* μ_n.

Aus der durch (1.12) gegebenen bivariaten erzeugenden Funktion $A(z,v)$ der Wahrscheinlichkeiten $A_{n,m}$, daß die Länge des linken Pfades in einem p-tree der Größe n genau m beträgt, erhalten wir durch

$$A_n(v) = [z^n]A(z,v)$$

die wahrscheinlichkeitserzeugende Funktion der Ω_n. Wir erhalten so zunächst

$$A_n(v) = \frac{1}{n}[z^{n-1}]\frac{\partial}{\partial z}A(z,v) = \frac{1}{n}[z^{n-1}]\frac{v}{(1-z)^{2v}} = \frac{v}{n}\binom{n+2v-2}{n-1} = \frac{v}{2v-1}\binom{n+2v-2}{n}.$$

Da für festes α die asymptotische Entwicklung

$$\begin{aligned}\binom{n+\alpha}{n} &= \frac{(n+\alpha)\cdot(n+\alpha-1)\cdots(n+1)}{\Gamma(\alpha+1)} = \frac{n^\alpha}{\Gamma(\alpha+1)}\cdot(1+\frac{\alpha}{n})\cdot(1+\frac{\alpha-1}{n})\cdots(1+\frac{1}{n})\\ &= \frac{n^\alpha}{\Gamma(\alpha+1)}\cdot(1+\mathcal{O}(1/n))\end{aligned}$$

gilt, folgert man weiter

$$A_n(v) = \frac{v}{2v-1}\binom{n+2v-2}{n} = \frac{v}{2v-1}\frac{n^{2v-2}}{\Gamma(2v-1)}\cdot(1+\mathcal{O}(1/n)) = \frac{v\,n^{2v-2}}{\Gamma(2v)}\cdot(1+\mathcal{O}(1/n))$$

gleichmäßig in einer hinreichend nahen Umgebung von $v=1$ für $n\to\infty$.

Für die charakteristische Funktion $\phi_{\Omega_n^*}(t)$ der Ω_n^* gilt nun unter Verwendung der $A_{n,m}$ bzw. $A_n(v)$:

$$\begin{aligned}\phi_{\Omega_n^*}(t) &= \mathbb{E}(e^{i\Omega_n^* t}) = \sum_m \mathbb{P}(\Omega_n^*=m)\cdot e^{imt} = \sum_m \mathbb{P}(\Omega_n=\mu_n+\sigma_n m)\cdot e^{imt}\\ &= \sum_{m\ge 0}\mathbb{P}(\Omega_n=m)\cdot e^{it\frac{m-\mu_n}{\sigma_n}} = \sum_{m\ge 0}A_{n,m}e^{it\frac{m-\mu_n}{\sigma_n}} = e^{-\frac{i\mu_n t}{\sigma_n}}A_n(e^{\frac{it}{\sigma_n}}).\end{aligned}$$

Wir entwickeln zunächst

$$A_n(e^{\frac{it}{\sigma_n}}) = \frac{e^{i\frac{t}{\sigma_n}}}{\Gamma(2e^{i\frac{t}{\sigma_n}})}\exp\left(2\log n(e^{i\frac{t}{\sigma_n}}-1)\right)\cdot(1+\mathcal{O}(1/n))$$

und folgern wegen

$$\Gamma(2e^s) = 1+s(-1+2\gamma)+\mathcal{O}(s^2)$$

weiter

$$A_n(e^{\frac{it}{\sigma_n}}) = \exp\left(2\log n(e^{i\frac{t}{\sigma_n}}-1)\right)\cdot(1+o(1))$$

$$
\begin{aligned}
&= \exp\left(2\log n\left(\frac{it}{\sigma_n} - \frac{t^2}{2\sigma_n^2} + \mathcal{O}\left(\frac{t^3}{\sigma_n^3}\right)\right)\right)\cdot(1+o(1)) \\
&= \exp\left(\sqrt{2\log n}\, it - \frac{t^2}{2} + \mathcal{O}\left(\frac{t^3}{\sqrt{\log n}}\right)\right)\cdot(1+o(1)) \\
&= \exp\left(\sqrt{2\log n}\, it - \frac{t^2}{2}\right)\cdot(1+o(1))\,.
\end{aligned}
$$

Damit bekommen wir nun

$$
\begin{aligned}
\phi_{\Omega_n^*}(t) &= e^{-\frac{i\mu_n t}{\sigma_n}} A_n(e^{\frac{it}{\sigma_n}}) = \exp\left(-\sqrt{2\log n}\, it\right)\exp\left(\sqrt{2\log n}\, it - \frac{t^2}{2}\right)\cdot(1+o(1)) \\
&= e^{-\frac{t^2}{2}}\cdot(1+o(1)) \to e^{-\frac{t^2}{2}}
\end{aligned}
$$

gleichmäßig in einer hinreichend nahen Umgebung von $t = 0$ für $n \to \infty$.

Zusammen mit dem Stetigkeitssatz von Levy komplettiert dies den Beweis. □

1.4 Die Höhe der Knoten

1.4.1 Die Verteilungen

Im folgenden betrachten wir die Verteilungen der Höhen der Knoten mit Schlüssel j in Objekten der Größe n aus den Familien $\mathcal{A}$, $\mathcal{B}$ und $\mathcal{C}$. Unter der Höhe eines Knotens verstehen wir immer die Anzahl der Knoten auf dem direkten Weg von der Wurzel hin zum betrachteten Knoten.

Seien also $A_{n,j,m}$, $B_{n,j,m}$ und $C_{n,j,m}$ die Wahrscheinlichkeiten, daß der Knoten mit Schlüssel j im Objekt der Größe n der entsprechenden Familie eine Höhe m besitzt.

Zum Aufstellen einer Rekursionsformel für diese Größen benötigen wir die vorhin behandelte Verteilung der Länge des linken Pfades in Objekten der Familie $\mathcal{C}$. Um eine Namenskollision zu vermeiden, bezeichnen wir in diesem Abschnitt die Wahrscheinlichkeit, daß die Länge des linken Pfades in einem Objekt der Größe n aus der Familie $\mathcal{C}$ genau m ist, mit $p_{n,m}$. Wir benötigen weiters die erzeugenden Funktionen

$$
p_n(v) = \sum_{m\geq 0} p_{n,m} v^m \quad \text{und} \quad p(z,v) = \sum_{n\geq 0}\sum_{m\geq 0} p_{n,m} z^n v^m\,.
$$

Mit den Gleichungen (1.9) und (1.15c) ergibt sich

$$
p_{n,m} = \frac{\left[{n \atop m}\right]}{n!}\,, \tag{1.27a}
$$

$$
p(z,v) = \frac{1}{(1-z)^v}\,, \tag{1.27b}
$$

$$
p_n(v) = [z^n]p(z,v) = \binom{v+n-1}{v-1}\,. \tag{1.27c}
$$

Teilt man nun die Familien $\mathcal{A}$, $\mathcal{B}$ und $\mathcal{C}$ aufgrund des ersten Elements einer Zufallspermutation auf, so ergeben sich die Rekursionen

$$A_{n,j,m} = \frac{1}{n}\left(\sum_{k=1}^{j-1} C_{n-k,j-k,m} + p_{n-j,m-1} + \sum_{k=j+1}^{n}\sum_{i=0}^{m} p_{n-k,i}\, B_{k-1,j,m-i}\right) \quad \text{für } n \ge j \ge 1\,, \tag{1.28a}$$

$$B_{n,j,m} = \frac{1}{n}\left(\sum_{k=1}^{j-1} A_{n-k,j-k,m-1} + \delta_{m,2} + \sum_{k=j+1}^{n} B_{k-1,j,m-1}\right) \quad \text{für } n \ge j \ge 1\,, \tag{1.28b}$$

$$C_{n,j,m} = \frac{1}{n}\left(\sum_{k=1}^{j-1} C_{n-k,j-k,m} + p_{n-j,m-1} + \sum_{k=j+1}^{n}\sum_{i=0}^{m-1} p_{n-k,i}\, A_{k-1,j,m-1-i}\right) \quad \text{für } n \ge j \ge 1\,. \tag{1.28c}$$

Definiert man

$$A_{n,j}(v) = \sum_{m\ge 0} A_{n,j,m} v^m\,, \quad B_{n,j}(v) = \sum_{m\ge 0} B_{n,j,m} v^m\,, \quad C_{n,j}(v) = \sum_{m\ge 0} C_{n,j,m} v^m\,,$$

dann ergeben sich aus obigen Gleichungen (1.28) durch Multiplikation mit v^m und Summation über alle $m \ge 0$

$$A_{n,j}(v) = \frac{1}{n}\left(\sum_{k=1}^{j-1} C_{n-k,j-k}(v) + v\, p_{n-j}(v) + \sum_{k=j+1}^{n} p_{n-k}(v)\, B_{k-1,j}(v)\right)\,, \tag{1.29a}$$

$$B_{n,j}(v) = \frac{1}{n}\left(\sum_{k=1}^{j-1} v\, A_{n-k,j-k}(v) + v^2 + \sum_{k=j+1}^{n} v\, B_{k-1,j}(v)\right)\,, \tag{1.29b}$$

$$C_{n,j}(v) = \frac{1}{n}\left(\sum_{k=1}^{j-1} C_{n-k,j-k}(v) + v\, p_{n-j}(v) + \sum_{k=j+1}^{n} v\, p_{n-k}(v)\, A_{k-1,j}(v)\right)\,. \tag{1.29c}$$

Führt man nun die erzeugenden Funktionen in drei Variablen z, u und v ein durch

$$A(z,u,v) = \sum_{j\ge 1}\sum_{n\ge j} A_{n,j}(v) z^n u^j\,, \quad B(z,u,v) = \sum_{j\ge 1}\sum_{n\ge j} B_{n,j}(v) z^n u^j\,,$$
$$C(z,u,v) = \sum_{j\ge 1}\sum_{n\ge j} C_{n,j}(v) z^n u^j\,,$$

so läßt sich obiges Gleichungssystem (1.29) durch Multiplizieren mit $nz^{n-1}u^j$ und anschließende Summation über alle $n \ge j \ge 1$ transformieren zu

$$\frac{\partial}{\partial z} A(z,u,v) = \frac{u}{1-uz} C(z,u,v) + \frac{vu}{1-uz} p(z,v) + p(z,v)\, B(z,u,v)\,, \tag{1.30a}$$

$$\frac{\partial}{\partial z} B(z,u,v) = \frac{vu}{1-uz} A(z,u,v) + \frac{v^2 u}{(1-z)(1-uz)} + \frac{v}{1-z} B(z,u,v)\,, \tag{1.30b}$$

$$\frac{\partial}{\partial z}C(z,u,v) = \frac{u}{1-uz}C(z,u,v) + \frac{vu}{1-uz}p(z,v) + v\,p(z,v)\,A(z,u,v)\,. \tag{1.30c}$$

Diese Transformation der Rekursionsformeln (1.28) durch erzeugende Funktionen führt also zu einem System von Differentialgleichungen für die erzeugenden Funktionen der Verteilungen der Höhen der Knoten in Objekten aus den Familien $\mathcal{A}$, $\mathcal{B}$ und $\mathcal{C}$. Dieses Gleichungssystem wird sich wohl in seiner Allgemeinheit einer exakten Lösung entziehen, aber man kann daran gehen, aus diesem statistische Kenngrößen abzuleiten.

1.4.2 Die Erwartungswerte

Die gewöhnlichen erzeugenden Funktionen der Erwartungswerte $A_{n,j}$, $B_{n,j}$ und $C_{n,j}$ für die Abstände des Knotens j in einem Objekt der Größe n aus den Familien $\mathcal{A}$, $\mathcal{B}$ und $\mathcal{C}$ bezeichnen wir hier mit $A(z,u)$, $B(z,u)$ und $C(z,u)$. Sie sind also gegeben durch

$$A(z,u) = \sum_{n\geq 1}\sum_{j=1}^{n} A_{n,j}z^n u^j\,, \quad B(z,u) = \sum_{n\geq 1}\sum_{j=1}^{n} B_{n,j}z^n u^j\,, \quad C(z,u) = \sum_{n\geq 1}\sum_{j=1}^{n} C_{n,j}z^n u^j\,.$$

Mit den vorher definierten erzeugenden Funktionen für die Verteilungen bestehen dann die Beziehungen

$$A(z,u) = \frac{\partial}{\partial v}A(z,u,v)\Big|_{v=1}\,, \quad B(z,u) = \frac{\partial}{\partial v}B(z,u,v)\Big|_{v=1}\,, \quad C(z,u) = \frac{\partial}{\partial v}C(z,u,v)\Big|_{v=1}\,. \tag{1.31}$$

Mit der aus (1.27b) erhaltenen Beziehung

$$\frac{\partial}{\partial v}p(z,v) = \log\left(\frac{1}{1-z}\right)\frac{1}{(1-z)^v} \tag{1.32}$$

ergibt sich durch Ableiten von (1.30) nach v und Auswerten bei $v=1$ folgendes System von Differentialgleichungen für die erzeugenden Funktionen $A(z,u)$, $B(z,u)$ und $C(z,u)$ der Erwartungswerte der Höhe der Knoten in Objeken aus den Familien $\mathcal{A}$, $\mathcal{B}$ und $\mathcal{C}$

$$\begin{aligned}
\frac{\partial}{\partial z}A(z,u) &= \frac{u}{1-uz}C(z,u) + \frac{u}{(1-z)(1-uz)} + \frac{u}{(1-z)(1-uz)}\log\left(\frac{1}{1-z}\right)\\
&\quad + \frac{uz}{(1-z)^2(1-uz)}\log\left(\frac{1}{1-z}\right) + \frac{1}{1-z}B(z,u)\,, && (1.33a)\\
\frac{\partial}{\partial z}B(z,u) &= \frac{u^2z}{(1-z)(1-uz)^2} + \frac{u}{1-uz}A(z,u) + \frac{2u}{(1-z)(1-uz)}\\
&\quad + \frac{uz}{(1-z)^2(1-uz)} + \frac{1}{1-z}B(z,u)\,, && (1.33b)\\
\frac{\partial}{\partial z}C(z,u) &= \frac{u}{1-uz}C(z,u) + \frac{u}{(1-z)(1-uz)} + \frac{u}{(1-z)(1-uz)}\log\left(\frac{1}{1-z}\right)\\
&\quad + \frac{uz}{(1-z)^2(1-uz)} + \frac{uz}{(1-z)^2(1-uz)}\log\left(\frac{1}{1-z}\right)\\
&\quad + \frac{1}{1-z}A(z,u)\,. && (1.33c)
\end{aligned}$$

Mit den Abkürzungen

$$\begin{aligned}
f_1(z,u) &= \frac{u}{(1-z)(1-uz)} + \frac{u}{(1-z)(1-uz)}\log\left(\frac{1}{1-z}\right) \\
&\quad + \frac{uz}{(1-z)^2(1-uz)}\log\left(\frac{1}{1-z}\right), \\
f_2(z,u) &= \frac{u^2 z}{(1-z)(1-uz)^2} + \frac{2u}{(1-z)(1-uz)} + \frac{uz}{(1-z)^2(1-uz)}, \\
f_3(z,u) &= \frac{u}{(1-z)(1-uz)} + \frac{u}{(1-z)(1-uz)}\log\left(\frac{1}{1-z}\right) + \frac{uz}{(1-z)^2(1-uz)} \\
&\quad + \frac{uz}{(1-z)^2(1-uz)}\log\left(\frac{1}{1-z}\right)
\end{aligned}$$

schreibt sich das Gleichungssystem (1.33) als

$$\frac{\partial}{\partial z}A(z,u) = \frac{u}{1-uz}C(z,u) + \frac{1}{1-z}B(z,u) + f_1(z,u)\,, \tag{1.34a}$$

$$\frac{\partial}{\partial z}B(z,u) = \frac{u}{1-uz}A(z,u) + \frac{1}{1-z}B(z,u) + f_2(z,u)\,, \tag{1.34b}$$

$$\frac{\partial}{\partial z}C(z,u) = \frac{u}{1-uz}C(z,u) + \frac{1}{1-z}A(z,u) + f_3(z,u)\,. \tag{1.34c}$$

Dieses Gleichungssystem vereinfacht sich, indem man neue Funktionen $\widetilde{B}(z,u)$ und $\widetilde{C}(z,u)$ einführt durch

$$B(z,u) = \frac{\widetilde{B}(z,u)}{1-z}\,, \quad C(z,u) = \frac{\widetilde{C}(z,u)}{1-uz}\,. \tag{1.35}$$

Man bekommt dann nämlich aus (1.34)

$$\frac{\partial}{\partial z}A(z,u) = \frac{u}{(1-uz)^2}\widetilde{C}(z,u) + \frac{1}{(1-z)^2}\widetilde{B}(z,u) + f_1(z,u)\,, \tag{1.36a}$$

$$\frac{1}{1-z}\frac{\partial}{\partial z}\widetilde{B}(z,u) = \frac{u}{1-uz}A(z,u) + f_2(z,u)\,, \tag{1.36b}$$

$$\frac{1}{1-uz}\frac{\partial}{\partial z}\widetilde{C}(z,u) = \frac{1}{1-z}A(z,u) + f_3(z,u)\,. \tag{1.36c}$$

Wir werden dieses Gleichungssystem nun schrittweise überführen in eine einzige Differentialgleichung für $A(z,u)$. Dazu werden wir zuerst $\widetilde{B}(z,u)$ eliminieren, indem wir aus (1.36a) $\widetilde{B}(z,u)$ herausheben, nach z ableiten und mit $\frac{1}{1-z}$ multiplizieren. Dies liefert die Gleichungen

$$\widetilde{B}(z,u) = (1-z)^2\frac{\partial}{\partial z}A(z,u) - \frac{u(1-z)^2}{(1-uz)^2}\widetilde{C}(z,u) - (1-z)^2 f_1(z,u)\,, \tag{1.37a}$$

$$\begin{aligned}
\frac{1}{1-z}\frac{\partial}{\partial z}\widetilde{B}(z,u) &= -2\frac{\partial}{\partial z}A(z,u) + (1-z)\frac{\partial^2}{\partial z^2}A(z,u) - \frac{u(1-z)}{(1-uz)^2}\frac{\partial}{\partial z}\widetilde{C}(z,u) \\
&\quad - \frac{2u(u-1)}{(1-uz)^3}\widetilde{C}(z,u) - \frac{1}{1-z}\frac{\partial}{\partial z}\left((1-z)^2 f_1(z,u)\right)\,.
\end{aligned} \tag{1.37b}$$

Nun können wir die Gleichungen (1.37) in obiges Gleichungssystem (1.36) einsetzen und bekommen als neues Differentialgleichungsystem in $A(z,u)$ und $\widetilde{C}(z,u)$

$$\begin{aligned}&\frac{u}{1-uz}A(z,u)+2\frac{\partial}{\partial z}A(z,u)-(1-z)\frac{\partial^2}{\partial z^2}A(z,u)\\&+\frac{u(1-z)}{(1-uz)^2}\frac{\partial}{\partial z}\widetilde{C}(z,u)+\frac{2u(u-1)}{(1-uz)^3}\widetilde{C}(z,u)=-f_2(z,u)-\frac{1}{1-z}\frac{\partial}{\partial z}\left((1-z)^2f_1(z,u)\right)\,,\end{aligned} \tag{1.38a}$$

$$\frac{1}{1-uz}\frac{\partial}{\partial z}\widetilde{C}(z,u)=\frac{1}{1-z}A(z,u)+f_3(z,u)\,. \tag{1.38b}$$

Nun können wir daran gehen, $\widetilde{C}(z,u)$ zu eliminieren. Dazu setzten wir Gleichung (1.38b) ein in Gleichung (1.38a) und lösen nach $\widetilde{C}(z,u)$ auf. Man erhält so

$$\begin{aligned}\widetilde{C}(z,u)=&-\frac{(1-uz)^2}{u-1}A(z,u)-\frac{(1-uz)^3}{u(u-1)}\frac{\partial}{\partial z}A(z,u)+\frac{(1-z)(1-uz)^3}{2u(u-1)}\frac{\partial^2}{\partial z^2}A(z,u)\\&-\frac{(1-uz)^3}{2u(u-1)}f_2(z,u)-\frac{(1-uz)^3}{2u(u-1)(1-z)}\frac{\partial}{\partial z}\left((1-z)^2f_1(z,u)\right)\\&-\frac{(1-uz)^2(1-z)}{2(u-1)}f_3(z,u)\,.\end{aligned} \tag{1.39}$$

Leitet man diese Gleichung nach z ab, multipliziert mit $\frac{1}{1-uz}$ und setzt das Ergebnis in Gleichung (1.38b) ein, dann erhält man als Differentialgleichung für $A(z,u)$

$$\begin{aligned}&\frac{1}{1-uz}\frac{\partial}{\partial z}\left[\frac{(1-uz)^2}{u-1}A(z,u)+\frac{(1-uz)^3}{u(u-1)}\frac{\partial}{\partial z}A(z,u)-\right.\\&\left.-\frac{(1-z)(1-uz)^3}{2u(u-1)}\frac{\partial^2}{\partial z^2}A(z,u)\right]+\frac{1}{1-z}A(z,u)\\&=\;-f_3(z,u)-\frac{1}{1-uz}\frac{\partial}{\partial z}\left[\frac{(1-uz)^3}{2u(u-1)}f_2(z,u)+\right.\\&\left.+\frac{(1-uz)^3}{2u(u-1)(1-z)}\frac{\partial}{\partial z}\left((1-z)^2f_1(z,u)\right)+\frac{(1-uz)^2(1-z)}{2(u-1)}f_3(z,u)\right]\,.\end{aligned} \tag{1.40}$$

Vereinfacht man dies nun und beachtet die Anfangswerte $A_{0,0}=0$, $A_{1,1}=1$, $A_{2,1}=2$ und $A_{2,2}=1$, so ergibt sich folgende lineare Differentialgleichung dritter Ordnung für $A(z,u)$ mit den angegebenen Anfangsbedingungen:

$$\begin{aligned}&\frac{2uz-u-1}{(1-z)(1-u)}A(z,u)-2\frac{1-uz}{1-u}\frac{\partial}{\partial z}A(z,u)-\\&-\frac{3}{2}\frac{(1-uz)(2uz-u-1)}{u(1-u)}\frac{\partial^2}{\partial z^2}A(z,u)-\frac{1}{2}\frac{(1-uz)^2(1-z)}{u(1-u)}\frac{\partial^3}{\partial z^3}A(z,u)=\frac{uz+u-2}{(1-z)^3(1-u)}\,,\\&A(0,u)=0\,,\quad \left.\frac{\partial}{\partial z}A(z,u)\right|_{z=0}=u\,,\quad \left.\frac{\partial^2}{\partial z^2}A(z,u)\right|_{z=0}=2\left(2u+u^2\right)\,.\end{aligned} \tag{1.41}$$

Durch ein Studium der der Gleichung (1.41) entsprechenden homogenen Differentialgleichung

$$\begin{aligned} &\frac{2uz-u-1}{(1-z)(1-u)}A_h(z,u) - 2\frac{1-uz}{1-u}\frac{\partial}{\partial z}A_h(z,u) - \\ &-\frac{3}{2}\frac{(1-uz)(2uz-u-1)}{u(1-u)}\frac{\partial^2}{\partial z^2}A_h(z,u) - \frac{1}{2}\frac{(1-uz)^2(1-z)}{u(1-u)}\frac{\partial^3}{\partial z^3}A_h(z,u) = 0 \end{aligned} \tag{1.42}$$

findet man, daß eine Lösung derselben gegeben ist durch

$$A_{h,1}(z,u) = \frac{1}{(1-z)(1-uz)}\,. \tag{1.43}$$

Um nun die Ordnung der Differentialgleichung (1.42) zu reduzieren, substituieren wir

$$A(z,u) = \frac{E(z,u)}{(1-z)(1-uz)}$$

mit einer so definierten Funktion $E(z,u)$ (Verfahren von d'Alembert [18]).
Gleichung (1.42) transformiert sich dann zu

$$\frac{1}{(1-z)(1-u)}\frac{\partial}{\partial z}E(z,u) - \frac{1}{2}\frac{1-uz}{(1-u)u}\frac{\partial^3}{\partial z^3}E(z,u) = 0\,. \tag{1.44}$$

Durch Einführen einer durch

$$F(z,u) = \frac{\partial}{\partial z}E(z,u)$$

definierten Funktion $F(z,u)$ schreibt sich die Differentialgleichung (1.44) als

$$\frac{1}{(1-z)(1-u)}F(z,u) - \frac{1}{2}\frac{1-uz}{(1-u)u}\frac{\partial^2}{\partial z^2}F(z,u) = 0\,. \tag{1.45}$$

Nun findet man eine Lösung von (1.44) durch

$$F_{h,1}(z,u) = (1-z)(1-uz)\,. \tag{1.46}$$

Substituieren wir nun weiter

$$F(z,u) = (1-z)(1-uz)G(z,u)\,,$$

mit einer so definierten Funktion $G(z,u)$, dann vereinfacht sich (1.45) zu

$$\begin{aligned} &\frac{2u^2z^2-u^2z-3uz+u+1}{(1-u)u}\frac{\partial}{\partial z}G(z,u) + \\ &+\frac{1}{2}\frac{u^2z^3-u^2z^2-2uz^2+2uz+z-1}{(1-u)u}\frac{\partial^2}{\partial z^2}G(z,u) = 0\,. \end{aligned} \tag{1.47}$$

Schließlich kommt man durch Einführen einer Funktion $H(z,u)$ mit der Substitution

$$H(z,u) = \frac{\partial}{\partial z}G(z,u)$$

zu einer linearen homogenen Differentialgleichung erster Ordnung

$$-\frac{(2uz-u-1)(1-uz)}{(1-u)u}H(z,u)-\frac{1}{2}\frac{(1-z)(1-uz)^2}{(1-u)u}\frac{\partial}{\partial z}H(z,u)=0\,. \tag{1.48}$$

Diese Gleichung besitzt nun die Lösung

$$H_h(z,u)=\frac{c(u)}{(1-z)^2(1-uz)^2}\,, \tag{1.49}$$

mit einer nur von u abhängigen Funktion $c(u)$.

Analoge Umformungen für die entsprechende inhomogene Differentialgleichung (1.41) führen nun zu

$$\frac{\partial}{\partial z}H(z,u)-2\frac{1+u-2uz}{(1-z)(1-uz)}H(z,u)=-2\frac{u(uz+u-2)}{(1-uz)^2(1-z)^4}\,. \tag{1.50}$$

Mittels der Methode der Variation der Konstanten (siehe beispielsweise [18]) von Gleichung (1.49) ermittelt man aus dem Ansatz

$$H_p(z,u)=\frac{g(z,u)}{(1-z)^2(1-uz)^2}$$

durch Differenzieren von (1.50) als Bedingung für die Funktion $g(z,u)$

$$\frac{\partial}{\partial z}g(z,u)=-\frac{2u(uz+u-2)}{(1-z)^2}\,,$$

woraus man schließlich zu folgender Partikulärlösung $H_p(z,u)$ für die Differentialgleichung (1.50) kommt:

$$H_p(z,u)=2\frac{u^2}{(1-z)^2(1-uz)^2}\log\left(\frac{1}{1-z}\right)-4\frac{u(u-1)}{(1-z)^3(1-uz)^2}\,. \tag{1.51}$$

Durch Rücksubstitution bekommt man nun eine Partikulärlösung $A_{p,a}(z,u)$ für $A(z,u)$. Sie ist gegeben ist durch

$$A_{p,a}(z,u)=\frac{1}{(1-z)(1-uz)}\int_0^z\left[(1-x)(1-ux)\int_0^x H_p(t,u)\,dt\right]dx\,. \tag{1.52}$$

Die allgemeine Lösung der homogenen Differentialgleichung für $A(z,u)$ ist nun ebenfalls aus (1.49) durch Rücksubstitution zu erhalten. Mit drei von u abhängigen Funktionen $c_1(u)$, $c_2(u)$ und $c_3(u)$ lautet sie

$$\begin{aligned}A_h(z,u)&=\frac{c_1(u)}{(1-z)(1-uz)}+c_2(u)\frac{z(6-3uz-3z+2uz^2)}{(1-z)(1-uz)}+\\&\quad+\frac{c_3(u)}{(1-z)(1-uz)}\int_0^z\left[(1-x)(1-ux)\int_0^x\frac{1}{(1-t)^2(1-ut)^2}\,dt\right]dx\,.\end{aligned} \tag{1.53}$$

Führt man diese Integrationen aus, schreibt sich diese Gleichung als

$$\begin{aligned} A_h(z,u) = {} & \frac{c_1(u)}{(1-z)(1-uz)} + c_2(u)\frac{z(6-3uz-3z+2uz^2)}{(1-z)(1-uz)} \\ & + c_3(u)\left[-\frac{6(1-uz)(2uz-3u+1)}{1-z}\log\left(\frac{1}{1-uz}\right)\right. \\ & \qquad + \frac{6(1-z)(2uz-3+u)u^2}{1-uz}\log\left(\frac{1}{1-z}\right) \\ & \qquad \left. - \frac{(u-1)(12u^2z^2-12u(u+1)z-5u^2+4u-5)}{(1-z)(1-uz)}\right]. \end{aligned} \tag{1.54}$$

Um die Koeffizienten anschließend besser ablesen zu können, formen wir die durch (1.52) gegebene Partikulärlösung $A_p(z,u)$ durch partielle Integration und anschließendes Vereinfachen etwas um und erhalten

$$\begin{aligned} A_p(z,u) = {} & \frac{1}{3}\frac{(2uz-3u+1)(1-uz)}{1-z}\int_0^z \frac{1}{(1-t)^2(1-ut)^2}\log\left(\frac{1}{1-t}\right)dt \\ & + \frac{1}{3}\frac{u}{(1-z)(1-uz)}\log^2\left(\frac{1}{1-z}\right) \\ & - \frac{1}{3}\left[2u^4z^4-5u^4z^3-8u^3z^3-10u+6uz+12u^3z^2+10u^3z+9u^2z^2-3u^2z^3\right. \\ & \quad \left.-27u^2z+27u^2-20u^3+1+6u^4-5u^4z+2u^3z^4+3u^4z^2\right] \times \\ & \times \frac{1}{(1-z)^2(1-u)^3(1-uz)}\log\left(\frac{1}{1-z}\right) \\ & - \frac{1}{3}\frac{(u+1)(2uz-3u+1)(1-uz)}{(1-u)^3(1-z)}\log\left(\frac{1}{1-uz}\right) \\ & - \frac{1}{3}\left[6u^3z^4+3u^3z-13u^3z^3+5u^3z^2-7u^2z^3+21u^2z^2-17u^2z+2uz\right. \\ & \quad \left.-2uz^2+3u-1\right]\frac{1}{(1-uz)(1-u)^2(1-z)^2}. \end{aligned} \tag{1.55, 1.56}$$

Paßt man nun in der allgemeinen Lösung der Differentialgleichung (1.40)

$$A(z,u) = A_p(z,u) + A_h(z,u) \tag{1.57}$$

die Funktionen $c_1(u)$, $c_2(u)$ und $c_3(u)$ an die Anfangsbedingungen an, bekommt man für diese die Lösungen

$$c_1(u) = \frac{1}{3}\frac{3u-1}{(1-u)^2}, \quad c_2(u) = -\frac{1}{3}\frac{u^2}{(1-u)^2}, \quad c_3(u) = 0. \tag{1.58}$$

Das ergibt insgesamt die vollständige Lösung für die erzeugende Funktion $A(z,u)$ der Erwartungswerte der Höhe des Knoten j im p-tree der Größe n:

$$A(z,u) = \frac{1}{3}\frac{(2uz-3u+1)(1-uz)}{1-z}\int_0^z \frac{1}{(1-t)^2(1-ut)^2}\log\left(\frac{1}{1-t}\right)dt$$

$$
\begin{aligned}
&+\frac{1}{3}\frac{u}{(1-z)(1-uz)}\log^2\left(\frac{1}{1-z}\right)\\
&-\frac{1}{3}\left[2u^4z^4-5u^4z^3-8u^3z^3-10u+6uz+12u^3z^2+10u^3z+9u^2z^2-3u^2z^3\right.\\
&\qquad\left.-27u^2z+27u^2-20u^3+1+6u^4-5u^4z+2u^3z^4+3u^4z^2\right]\times \qquad (1.59)\\
&\quad\times\frac{1}{(1-z)^2(1-u)^3(1-uz)}\log\left(\frac{1}{1-z}\right)\\
&-\frac{1}{3}\frac{(u+1)(2uz-3u+1)(1-uz)}{(1-u)^3(1-z)}\log\left(\frac{1}{1-uz}\right)\\
&-\frac{1}{3}\left[-8u^3z^2+2u^3z+3u^3+4u^3z^3-4u^2z^2-11u^2+12u^2z-2uz+\right.\\
&\qquad\left.+5u-1\right]\frac{z}{(1-uz)(1-u)^2(1-z)^2}\,. \qquad (1.60)
\end{aligned}
$$

Das Koffizientenablesen gestaltet sich leichter, wenn man noch die Partialbruchzerlegung durchführt und erst dann von den einzelnen Summanden abliest. Dem Rechnung getragen schreibt sich $A(z,u)$ als

$$
\begin{aligned}
A(z,u)=&\frac{1}{3}(2u^2z-u^2-u)\int_0^z\frac{1}{(1-t)^2(1-ut)^2}\log\left(\frac{1}{1-t}\right)dt\\
&+\frac{1}{3}\frac{(1-u)^2}{1-z}\int_0^z\frac{1}{(1-t)^2(1-ut)^2}\log\left(\frac{1}{1-t}\right)dt\\
&+\frac{1}{3}\frac{u}{(1-z)(1-uz)}\log^2\left(\frac{1}{1-z}\right)\\
&+\frac{1}{3}\frac{u(2u^2z-u^2+2uz-2u+1)}{(1-u)^3}\log\left(\frac{1}{1-z}\right)\\
&+\frac{7}{3}\frac{u}{(1-u)(1-z)}\log\left(\frac{1}{1-z}\right)-\frac{1}{3}\frac{1}{(1-z)^2}\log\left(\frac{1}{1-z}\right)\\
&-\frac{1}{3}\frac{(6u-1)u}{(1-u)(1-uz)}\log\left(\frac{1}{1-z}\right)\\
&-\frac{1}{3}\frac{u(u+1)(2uz-u-1)}{(1-u)^3}\log\left(\frac{1}{1-uz}\right)-\frac{1}{3}\frac{u+1}{(1-u)(1-z)}\log\left(\frac{1}{1-uz}\right)\\
&-\frac{4}{3}\frac{u^2z}{(1-u)^2}+\frac{u}{(1-u)(1-uz)}-\frac{1}{3}\frac{2u+1}{(1-u)(1-z)}+\frac{1}{3}\frac{1}{(1-z)^2}\,. \qquad (1.61)
\end{aligned}
$$

Es soll hier angemerkt werden, daß das in diesem Ausdruck auftretende Integral mittels

$$
\begin{aligned}
&\int_0^z\frac{1}{(1-t)^2(1-ut)^2}\log\left(\frac{1}{1-t}\right)dt=\frac{2u}{(1-u)^3}\left(\mathrm{L}_2\left(\frac{1-uz}{u(1-z)}\right)-\mathrm{L}_2\left(\frac{1}{u}\right)\right)\\
&+\frac{2u}{(1-u)^3}\log\left(\frac{u}{u-1}\right)\log\left(\frac{1-z}{1-uz}\right)+\frac{z^2u^2-3zu^2+zu+u^2+u-1}{(1-u)^3(1-z)(1-uz)}\log(1-z)\\
&-\frac{u}{(1-u)^3}\log(1-uz)-\frac{z}{(1-u)^2(1-z)}\,.
\end{aligned}
$$

in Beziehung zu dem in Unterabschnitt 1.4.5 definiertem Polylogarithmus $\mathrm{L}_2(z)$ steht. Für das anschließende Koeffizientenablesen ist die gewählte Darstellung allerdings besser geeignet.

Zum Ablesen der Koeffizienten benötigt man weitere Beziehungen mit Harmonischen Zahlen, welche man im wesentlichen aus (1.19) und eventuelles Aufsummieren erhalten kann. Sie seien nachfolgend aufgelistet.

$$[z^n u^j] \int_0^z \frac{1}{(1-t)^2(1-ut)^2} \log\left(\frac{1}{1-t}\right) dt = \frac{(j+1)(n-j)}{n}(H_{n-j}-1)\,, \tag{1.62a}$$

$$\begin{aligned}[z^n u^j] \frac{1}{1-z} \int_0^z \frac{1}{(1-t)^2(1-ut)^2} \log\left(\frac{1}{1-t}\right) dt \\ = (j+1)\left((n-j+1)H_{n-j} - 2(n-j)\right) - j(j+1)\left(\sum_{k=j+1}^{n} \frac{H_{k-j}}{k} - (H_n - H_j)\right),\end{aligned} \tag{1.62b}$$

$$[z^n u^j] \frac{1}{(1-z)(1-uz)} \log^2\left(\frac{1}{1-z}\right) = H_{n-j}^2 - H_{n-j}^{(2)}\,, \tag{1.62c}$$

$$[z^n u^j] \frac{1}{(1-z)(1-u)} \log\left(\frac{1}{1-uz}\right) = H_j \quad \text{für } 1 \le j \le n\,, \tag{1.62d}$$

$$[z^n u^j] \frac{1}{(1-u)(1-z)} = 1 \quad \text{für } 0 \le j \le n\,, \tag{1.62e}$$

$$[z^n u^j] \frac{1}{(1-u)^3} \log\left(\frac{1}{1-z}\right) = \frac{1}{n}\binom{j+2}{2}\,, \tag{1.62f}$$

$$[z^n u^j] \frac{1}{(1-u)(1-z)} \log\left(\frac{1}{1-z}\right) = H_n\,, \tag{1.62g}$$

$$[z^n u^j] \frac{1}{(1-u)(1-uz)} \log\left(\frac{1}{1-z}\right) = H_n - H_{n-j-1} \quad \text{für } n \ge j+1\,. \tag{1.62h}$$

Liest man nun die Koeffizienten von $A(z,u)$ unter Zuhilfenahme obiger Beziehungen ab, ergibt sich

Satz 1.4.1. *Der Erwartungswert $A_{n,j}$ der Höhe des Knotens mit Schlüssel j in einem p-tree der Größe n, also einem Objekt der Größe n aus der Familie $\mathcal{A}$, ist gegeben durch*

$$\boxed{\begin{aligned} A_{n,j} = -\frac{2}{3}\sum_{k=j+1}^{n}\frac{H_{k-j}}{k} + \frac{1}{3}(H_n - H_j) + \frac{-\frac{2}{3}l(l+1)H_l + \frac{4}{3}l^2}{n-1} \\ + \frac{\frac{2}{3}(l+1)^2 H_l - \frac{4}{3}l^2 - l}{n} + \frac{1}{3}(H_{l+1}^2 - H_{l+1}^{(2)}) + \frac{5}{3}H_{l+1} - 1 \\ + \frac{1}{3(l+1)} \quad \text{für } 1 \le j \le n \quad \text{mit } \ l := n-j\,. \end{aligned}} \tag{1.63}$$

Um nun zu $C(z,u)$ zu kommen, setzen wir das aus (1.61) erhaltene $A(z,u)$ in Gleichung (1.39) ein. Mit (1.35) erhalten wir nach etlichen Umformungen

$$\begin{aligned} C(z,u) = \frac{1}{3}\frac{(1-uz)^2}{1-z} \int_0^z \frac{1}{(1-t)^2(1-ut)^2} \log\left(\frac{1}{1-t}\right) dt \\ + \frac{1}{3}\frac{u}{(1-uz)(1-z)} \log^2\left(\frac{1}{1-z}\right) \end{aligned}$$

$$
\begin{aligned}
&+\left[-\frac{1}{3}\frac{u(u^2z+u^2-u+uz-2)}{(1-u)^3}+\frac{7}{3}\frac{u}{(1-u)(1-z)}-\frac{1}{3}\frac{1}{(1-z)^2}\right.\\
&\qquad\left.-\frac{1}{3}\frac{u(6u-1)}{(1-u)(1-uz)}\right]\log\left(\frac{1}{1-z}\right)\\
&+\left(\frac{1}{3}\frac{u(u+1)(uz+u-2)}{(1-u)^3}-\frac{1}{3}\frac{u+1}{(1-u)(1-z)}\right)\log\left(\frac{1}{1-uz}\right)\\
&-\frac{1}{3}\frac{u(3u+2uz-3)}{(1-u)^2}-\frac{1}{3}\frac{5u+1}{(1-u)(1-z)}+\frac{1}{3}\frac{1}{(1-z)^2}+\frac{u}{(1-u)(1-uz)}\,.
\end{aligned}
\tag{1.64}
$$

Mit den oben angegebenen Beziehungen (1.62) erhalten wir durch Koeffizientenablesen

Korollar 1.4.2. *Der Erwartungswert $C_{n,j}$ der Höhe des Knotens mit Schlüssel j in einem Objekt der Größe n aus der Familie $\mathcal{C}$ ist gegeben durch*

$$
\boxed{
\begin{aligned}
C_{n,j}=&-\frac{2}{3}\sum_{k=j+1}^{n}\frac{H_{k-j}}{k}+\frac{1}{3}(H_n-H_j)+\frac{\frac{1}{3}l(l+1)H_l-\frac{2}{3}l^2}{n-1}\\
&+\frac{-\frac{1}{3}(l-2)(l+1)H_l+\frac{2}{3}l^2-l}{n}+\frac{1}{3}(H_{l+1}^2-H_{l+1}^{(2)})\\
&+\frac{5}{3}H_{l+1}-1+\frac{1}{3(l+1)}\qquad\text{für } 1\le j\le n \quad \text{mit } l:=n-j\,.
\end{aligned}
}
\tag{1.65}
$$

Für die Berechnung von $B(z,u)$ setzen wir $A(z,u)$ und $C(z,u)$ aus Gleichung (1.61) bzw. (1.64) in (1.37a) ein. Unter Verwendung von (1.35) erhalten wir damit

$$
\begin{aligned}
B(z,u)=&\left(-\frac{1}{3}u^2z+\frac{2}{3}u^2-\frac{1}{3}u+\frac{1}{3}\frac{(1-u)^2}{1-z}\right)\int_0^z\frac{1}{(1-t)^2(1-ut)^2}\log\left(\frac{1}{1-t}\right)dt\\
&+\frac{1}{3}\frac{u}{(1-z)(1-uz)}\log^2\left(\frac{1}{1-z}\right)\\
&+\left[-\frac{1}{3}\frac{u(u^2z-2u^2-u+uz+1)}{(1-u)^3}+\frac{4}{3}\frac{u}{(1-u)(1-z)}-\frac{1}{3}\frac{1}{(1-z)^2}\right.\\
&\qquad\left.-\frac{1}{3}\frac{u(3u-1)}{(1-u)(1-uz)}\right]\log\left(\frac{1}{1-z}\right)\\
&+\left(\frac{1}{3}\frac{u(u+1)(uz-2u+1)}{(1-u)^3}-\frac{1}{3}\frac{u+1}{(1-u)(1-z)}\right)\log\left(\frac{1}{1-uz}\right)\\
&-\frac{1}{3}\frac{u(2uz-3u+3)}{(1-u)^2}+\frac{1}{3}\frac{1}{(1-z)^2}+\frac{1}{3}\frac{4u-1}{(1-u)(1-z)}\,.
\end{aligned}
\tag{1.66}
$$

Das Ablesen der Koeffizienten geschieht wie gehabt unter Benützung der Formeln (1.62) und man bekommt

Korollar 1.4.3. *Der Erwartungswert $B_{n,j}$ der Höhe des Knotens mit Schlüssel j in Objekten* *aus der Familie $\mathcal{B}$ mit Größe n ist gegeben durch*

$$\boxed{\begin{aligned} B_{n,j} = &-\frac{2}{3}\sum_{k=j+1}^{n}\frac{H_{k-j}}{k} + \frac{1}{3}(H_n - H_j) + \frac{\frac{1}{3}l(l+1)H_l - \frac{2}{3}l^2}{n-1} \\ &+ \frac{-\frac{1}{3}(l+4)(l+1)H_l + \frac{2}{3}l^2 + 2l}{n} + \frac{1}{3}(H_{l+1}^2 - H_{l+1}^{(2)}) \\ &+ \frac{5}{3}H_{l+1} + \frac{1}{3(l+1)} \quad \text{für } 1 \le j \le n \quad \text{mit} \quad l := n-j\,. \end{aligned}} \tag{1.67}$$

1.4.3 Das zweite Faktorielle Moment

Dies dient als Vorbereitung für die Berechnung der eigentlich gesuchten Varianz der Höhe der Knoten in einem zufälligen p-tree.

Die gewöhnlichen erzeugenden Funktionen der zweiten Faktoriellen Momente $\breve{A}_{n,j}$, $\breve{B}_{n,j}$ und $\breve{C}_{n,j}$ für die Abstände des Knotens j in einem Objekt der Größe n aus den Familien $\mathcal{A}$, $\mathcal{B}$ und $\mathcal{C}$ bezeichnen wir hier mit $\breve{A}(z,u)$, $\breve{B}(z,u)$ und $\breve{C}(z,u)$. Sie sind also definiert durch

$$\breve{A}(z,u) = \sum_{n\ge1}\sum_{j=1}^{n}\breve{A}_{n,j}z^n u^j\,, \quad \breve{B}(z,u) = \sum_{n\ge1}\sum_{j=1}^{n}\breve{B}_{n,j}z^n u^j\,, \quad \breve{C}(z,u) = \sum_{n\ge1}\sum_{j=1}^{n}\breve{C}_{n,j}z^n u^j\,.$$

Des weiteren gelten die Bezeichnungen der vorangegangenen Unterabschnitte 1.4.1 und 1.4.2.

Es gelten somit folgende Beziehungen mit den erzeugenden Funktionen der Verteilungen:

$$\breve{A}(z,u) = \frac{\partial^2}{\partial v^2}A(z,u,v)\bigg|_{v=1}, \quad \breve{B}(z,u) = \frac{\partial^2}{\partial v^2}B(z,u,v)\bigg|_{v=1}, \quad \breve{C}(z,u) = \frac{\partial^2}{\partial v^2}C(z,u,v)\bigg|_{v=1}. \tag{1.68}$$

Mit der aus (1.27b) erhaltenen Beziehung für die erzeugende Funktion der Länge des linken Pfades in Objekten der Familie $\mathcal{C}$

$$\frac{\partial^2}{\partial v^2}p(z,v) = \log^2\left(\frac{1}{1-z}\right)\frac{1}{(1-z)^v} \tag{1.69}$$

erhalten wir durch zweimaliges Differenzieren des Gleichungssystems (1.30) nach v und anschließendes Auswerten an der Stelle $v=1$ folgendes Gleichungssystem für die $\breve{A}(z,u)$, $\breve{B}(z,u)$ und $\breve{C}(z,u)$:

$$\begin{aligned} \frac{\partial}{\partial z}\breve{A}(z,u) = &\frac{u}{1-uz}\breve{C}(z,u) + \frac{2u}{(1-z)(1-uz)}\log\left(\frac{1}{1-z}\right) \\ &+ \frac{u}{(1-z)(1-uz)}\log^2\left(\frac{1}{1-z}\right) + \frac{uz}{(1-z)^2(1-uz)}\log^2\left(\frac{1}{1-z}\right) \\ &+ \frac{2}{1-z}\log\left(\frac{1}{1-z}\right)B(z,u) + \frac{1}{1-z}\breve{B}(z,u)\,, \end{aligned} \tag{1.70a}$$

$$
\begin{aligned}
\frac{\partial}{\partial z}\breve{B}(z,u) &= \frac{u}{1-uz}\breve{A}(z,u) + \frac{2u}{1-uz}A(z,u) + \frac{2u}{(1-z)(1-uz)} \\
&\quad + \frac{2}{1-z}B(z,u) + \frac{1}{1-z}\breve{B}(z,u)\,, \qquad (1.70b)\\
\frac{\partial}{\partial z}\breve{C}(z,u) &= \frac{u}{1-uz}\breve{C}(z,u) + \frac{2u}{(1-z)(1-uz)}\log\left(\frac{1}{1-z}\right) \\
&\quad + \frac{u}{(1-z)(1-uz)}\log^2\left(\frac{1}{1-z}\right) + \frac{uz}{(1-z)^2(1-uz)}\log^2\left(\frac{1}{1-z}\right) \\
&\quad + \frac{2uz}{(1-z)^2(1-uz)}\log\left(\frac{1}{1-z}\right) + \frac{2}{1-z}\log\left(\frac{1}{1-z}\right)A(z,u) \\
&\quad + \frac{2}{1-z}A(z,u) + \frac{1}{1-z}\breve{A}(z,u)\,. \qquad (1.70c)
\end{aligned}
$$

Mit den Bezeichnungen

$$
\begin{aligned}
\breve{f}_1(z,u) &= \frac{2u}{(1-z)(1-uz)}\log\left(\frac{1}{1-z}\right) + \frac{u}{(1-z)(1-uz)}\log^2\left(\frac{1}{1-z}\right) \\
&\quad + \frac{uz}{(1-z)^2(1-uz)}\log^2\left(\frac{1}{1-z}\right) + \frac{2}{1-z}\log\left(\frac{1}{1-z}\right)B(z,u)\,, \\
\breve{f}_2(z,u) &= \frac{u}{1-uz}\breve{A}(z,u) + \frac{2u}{1-uz}A(z,u) + \frac{2u}{(1-z)(1-uz)} + \frac{2}{1-z}B(z,u)\,, \\
\breve{f}_3(z,u) &= \frac{2u}{(1-z)(1-uz)}\log\left(\frac{1}{1-z}\right) + \frac{u}{(1-z)(1-uz)}\log^2\left(\frac{1}{1-z}\right) \\
&\quad + \frac{uz}{(1-z)^2(1-uz)}\log^2\left(\frac{1}{1-z}\right) + \frac{2uz}{(1-z)^2(1-uz)}\log\left(\frac{1}{1-z}\right) \\
&\quad + \frac{2}{1-z}\log\left(\frac{1}{1-z}\right)A(z,u) + \frac{2}{1-z}A(z,u)\,,
\end{aligned}
$$

schreibt sich das Gleichungssystem (1.70) als

$$
\begin{aligned}
\frac{\partial}{\partial z}\breve{A}(z,u) &= \frac{u}{1-uz}\breve{C}(z,u) + \frac{1}{1-z}\breve{B}(z,u) + \breve{f}_1(z,u)\,, \qquad (1.71a)\\
\frac{\partial}{\partial z}\breve{B}(z,u) &= \frac{u}{1-uz}\breve{A}(z,u) + \frac{1}{1-z}\breve{B}(z,u) + \breve{f}_2(z,u)\,, \qquad (1.71b)\\
\frac{\partial}{\partial z}\breve{C}(z,u) &= \frac{u}{1-uz}\breve{C}(z,u) + \frac{1}{1-z}\breve{A}(z,u) + \breve{f}_3(z,u)\,. \qquad (1.71c)
\end{aligned}
$$

Dieses System von Differentialgleichungen ist von gleicher Bauart, wie das für die Erwartungswerte gefundene (1.36). Somit können die selben Umformungen wie dort angewandt werden, um zu einer Differentialgleichung dritter Ordnung für $\breve{A}(z,u)$ zu gelangen.

Unter Berücksichtigung der Anfangswerte $\breve{A}_{0,0} = 0$, $\breve{A}_{1,1} = 0$, $\breve{A}_{2,1} = 2$, $\breve{A}_{2,2} = 0$, $\breve{A}_{3,1} = \frac{14}{3}$,

$\breve{A}_{3,2} = 2$ und $\breve{A}_{3,3} = 0$ findet man als zu lösende Differentialgleichung für $\breve{A}(z,u)$

$$
\begin{aligned}
&\frac{2uz-u-1}{(1-z)(1-u)}\breve{A}(z,u) - 2\frac{1-uz}{1-u}\frac{\partial}{\partial z}\breve{A}(z,u) \\
&- \frac{3}{2}\frac{(1-uz)(2uz-u-1)}{u(1-u)}\frac{\partial^2}{\partial z^2}\breve{A}(z,u) - \frac{1}{2}\frac{(1-uz)^2(1-z)}{u(1-u)}\frac{\partial^3}{\partial z^3}\breve{A}(z,u) \\
&\quad = +\breve{f}_3(z,u) + \frac{1}{1-uz}\frac{\partial}{\partial z}\left(\frac{(1-uz)^3}{2u(u-1)}\breve{f}_2(z,u)\right. \\
&\qquad \left. + \frac{(1-uz)^3}{2u(u-1)(1-z)}\frac{\partial}{\partial z}\Big((1-z)^2\breve{f}_1(z,u)\Big) + \frac{(1-uz)^2(1-z)}{2(u-1)}\breve{f}_3(z,u)\right), \\
&\breve{A}(0,u) = 0, \quad \frac{\partial}{\partial z}\breve{A}(z,u)\bigg|_{z=0} = 0, \quad \frac{\partial^2}{\partial z^2}\breve{A}(z,u)\bigg|_{z=0} = 4u, \quad \frac{\partial^3}{\partial z^3}\breve{A}(z,u)\bigg|_{z=0} = 28u + 12u^2.
\end{aligned}
\tag{1.72}
$$

Der Ausdruck auf der rechten Seite dieses Gleichungssystems, welchen ich hier mit $r(z,u)$ bezeichne, schreibt sich unter Verwendung von (1.61), (1.64) und (1.66) als

$$
\begin{aligned}
r(z,u) = &\left[4\frac{(1-uz)u^2}{u-1}\log\left(\frac{1}{1-z}\right) - \frac{2}{3}\frac{(uz+u-2)u^2}{u-1} + \frac{4}{3}\frac{(u-1)^3}{u(1-z)^3} - \frac{2}{3}\frac{(u-1)^2}{(1-z)^2}\right] \times \\
&\times \int_0^z \frac{1}{(1-t)^2(1-ut)^2}\log\left(\frac{1}{1-t}\right)dt \\
&+ 4\frac{(u+1)(1-uz)u^2}{(u-1)^4}\log\left(\frac{1}{1-z}\right)\log\left(\frac{1}{1-uz}\right) \\
&+ \left(-\frac{2}{3}\frac{(uz+u-2)(u+1)u^2}{(u-1)^4} + \frac{4}{3}\frac{u+1}{u(1-z)^3} - \frac{2}{3}\frac{u+1}{(u-1)(1-z)^2}\right)\log\left(\frac{1}{1-uz}\right) \\
&+ \left(-4\frac{u^2(u+1)(1-uz)}{(u-1)^4} - \frac{4}{3}\frac{1}{(1-z)^3} + \frac{2}{3}\frac{u}{(1-z)^2(u-1)}\right)\log^2\left(\frac{1}{1-z}\right) + \\
&+ \left[-\frac{2}{3}\frac{(11u^2z+5u^2-13uz-11u+8)u^2}{(u-1)^4} - \frac{4}{3}\frac{u-1}{u(1-z)^4}\right. \\
&\quad \left. - \frac{22}{3}\frac{1}{(1-z)^3} + 12\frac{u}{(1-z)^2(u-1)} + 2\frac{(u+1)u}{(1-z)(u-1)^2}\right]\log\left(\frac{1}{1-z}\right) \\
&+ \frac{4}{3}\frac{u-1}{u(1-z)^4} - \frac{2}{3}\frac{17u-2}{u(1-z)^3} + 2\frac{5u-1}{(1-z)^2(u-1)} + 2\frac{u^2}{(1-z)(u-1)^2} \\
&- \frac{2}{3}\frac{(2uz+3u-3)u^2}{(u-1)^3}.
\end{aligned}
\tag{1.73}
$$

Nachfolgend wird nun skizziert, wie man unter Verwendung eines Formelmanipulationsprogrammes die Lösung für die Differentialgleichung

$$
\begin{aligned}
&\frac{2uz-u-1}{(1-z)(1-u)}\breve{A}(z,u) - 2\frac{1-uz}{1-u}\frac{\partial}{\partial z}\breve{A}(z,u) \\
&- \frac{3}{2}\frac{(1-uz)(2uz-u-1)}{u(1-u)}\frac{\partial^2}{\partial z^2}\breve{A}(z,u) - \frac{1}{2}\frac{(1-uz)^2(1-z)}{u(1-u)}\frac{\partial^3}{\partial z^3}\breve{A}(z,u) = r(z,u)
\end{aligned}
\tag{1.74}
$$

findet.

Mit den nacheinander ausgeführten Transformationen

$$\begin{aligned}
\breve{A}(z,u) &= \frac{\breve{E}(z,u)}{(1-z)(1-uz)}\,,\\
\breve{F}(z,u) &= \frac{\partial}{\partial z}\breve{E}(z,u)\,,\\
\breve{F}(z,u) &= (1-z)(1-uz)\breve{G}(z,u)\,,\\
\breve{H}(z,u) &= \frac{\partial}{\partial z}\breve{G}(z,u)\,,
\end{aligned}$$

kommt man nach etlichen Umformungen schließlich zur linearen Differentialgleichung erster Ordnung

$$\frac{\partial}{\partial z}\breve{H}(z,u) - 2\frac{1+u-2uz}{(1-z)(1-uz)}\breve{H}(z,u) = \frac{2u(u-1)}{(1-z)(1-uz)^2}\,r(z,u)\,. \tag{1.75}$$

Durch Variation der Konstanten bekommt man eine Partikulärlösung obiger Differentialgleichung durch

$$\breve{H}(z,u) = \frac{m(z,u)}{(1-z)^2(1-uz)^2}\,, \tag{1.76}$$

wobei $m(z,u)$ gegeben ist durch

$$\begin{aligned}
\frac{\partial}{\partial z}m(z,u) = \Bigg[&8u^3(1-z)(1-uz)\log\left(\frac{1}{1-z}\right) - \frac{4}{3}\frac{u(u-1)^3}{1-z} + \frac{8}{3}\frac{(u-1)^4}{(1-z)^2}\\
&- \frac{4}{3}u^3(1-z)(uz+u-2)\Bigg]\int_0^z \frac{1}{(1-t)^2(1-ut)^2}\log\left(\frac{1}{1-t}\right)\,dt\\
&+ 8\frac{u^3(u+1)(1-z)(1-uz)}{(u-1)^2}\log\left(\frac{1}{1-z}\right)\log\left(\frac{1}{1-uz}\right)\\
&+ \left[-\frac{4}{3}\frac{u^3(u+1)(1-z)(uz+u-2)}{(u-1)^3} + \frac{8}{3}\frac{(u-1)(u+1)}{(1-z)^2} - \frac{4}{3}\frac{(u+1)u}{1-z}\right]\times\\
&\times\log\left(\frac{1}{1-uz}\right)\\
&+ \left(-8\frac{u^3(u+1)(1-z)(1-uz)}{(u-1)^3} - \frac{8}{3}\frac{(u-1)u}{(1-z)^2} + \frac{4}{3}\frac{u^2}{1-z}\right)\log^2\left(\frac{1}{1-z}\right)\\
&+ \Bigg[\frac{4}{3}\frac{u^2}{(u-1)^3}(11u^3z^2 - 6u^3z - 2u^3 - 13u^2z^2 + 2u^2z + 8u^2 + 8uz - 11u + 3)\\
&\qquad - \frac{8}{3}\frac{(u-1)^2}{(1-z)^3} + \frac{44}{3}\frac{(u-1)u}{(1-z)^2} + 24\frac{u^2}{1-z}\Bigg]\log\left(\frac{1}{1-z}\right)\\
&+ \frac{8}{3}\frac{(u-1)^2}{(1-z)^3} - \frac{4}{3}\frac{(17u-2)(u-1)}{(1-z)^2} + 4\frac{(5u-1)u}{1-z} + \frac{4}{3}\frac{u^3z(2uz+u-3)}{(u-1)^2}\,.
\end{aligned} \tag{1.77}$$

Damit ergibt sich durch Rücktransformation eine Partikulärlösung von (1.74) durch

$$\breve{A}_p(z,u) = \\
= \frac{1}{(1-z)(1-uz)}\int_0^z \left[(1-y)(1-uy)\int_0^y \left[\frac{1}{(1-x)^2(1-ux)^2}\int_0^x m(t,u)\,dt\right]dx\right]dy\,. \tag{1.78}$$

Im Hinblick auf ein späteres Ablesen der Koeffizienten empfiehlt es sich, durch partielle Integration diesen Ausdruck weiter umzuformen. Mit den Bezeichnungen

$$g_1(z,u) = (1-z)(1-uz)\,, \quad g_2(z,u) = \frac{1}{(1-z)^2(1-uz)^2}\,,$$
$$g_3(z,u) = \int_0^z g_1(t,u)\,dt\,, \quad g_4(z,u) = \int_0^z g_2(t,u)\,dt\,, \quad g_5(z,u) = \int_0^z g_2(t,u)\,g_3(t,u)\,dt$$

erhält man

$$\begin{aligned}\breve{A}_p(z,u) = {} & \frac{g_3(z,u)\,g_4(z,u)}{(1-z)(1-uz)} \int_0^z m(t,u)\,dt - \frac{g_3(z,u)}{(1-z)(1-uz)} \int_0^z g_4(t,u)\,m(t,u)\,dt \\ & - \frac{g_5(z,u)}{(1-z)(1-uz)} \int_0^z m(t,u)\,dt + \frac{1}{(1-z)(1-uz)} \int_0^z g_5(t,u)\,m(t,u)\,dt\,. \end{aligned} \tag{1.79}$$

Dies führt nach zahlreichen Manipulationen zu folgendem Ausdruck:

$$\begin{aligned}\breve{A}_p(z,u) = {} & \left[\frac{8}{9}\frac{(2uz-1-u)u^3}{(1-u)^3} + \frac{4}{9}\frac{u^3}{(1-uz)(1-u)} + \frac{4}{9}\frac{u^2}{(1-z)(1-u)}\right] \times \\
& \times \int_0^z \frac{1}{1-ut}\log^3\left(\frac{1}{1-t}\right) dt \\
& - \left[-\frac{2}{3}\frac{(8u-1)(2uz-1-u)u^2}{(1-u)^3} - \frac{2}{27}\frac{(70u-3)u^2}{(1-uz)(1-u)} - \frac{4}{27}\frac{(u-3)u}{(1-z)(1-u)}\right] \times \\
& \times \int_0^z \frac{1}{1-ut}\log^2\left(\frac{1}{1-t}\right) dt \\
& + \Bigg[\left(-\frac{4}{9}\frac{(2uz-1-u)u^3}{(1-u)^3} - \frac{4}{9}\frac{u^3}{(1-uz)(1-u)}\right)\log\left(\frac{1}{1-z}\right)^2 \\
& \quad - \left(\frac{4}{27}\frac{(10uz-11u+1)u^3}{(1-u)^3} + \frac{140}{27}\frac{u^3}{(1-uz)(1-u)} - \frac{8}{3}\frac{u^2}{(1-z)(1-u)}\right) \times \\
& \quad \times \log\left(\frac{1}{1-z}\right) + \frac{2}{81}\frac{(106u^2z-131u^2-234uz+142u+117)u^2}{(1-u)^3} \\
& \quad - \frac{70}{27}\frac{u^2}{(1-uz)(1-u)} - \frac{2}{81}\frac{(79u+12)u}{(1-z)(1-u)}\Bigg] \int_0^z \frac{1}{1-ut}\log\left(\frac{1}{1-t}\right) dt \\
& - \left(\frac{2}{27}\frac{(38uz-25u-13)u^2}{(1-u)^3} + \frac{14}{27}\frac{u^2}{(1-uz)(1-u)}\right)\log^3\left(\frac{1}{1-z}\right) \\
& - \left(-\frac{2}{9}\frac{(2uz-1-u)u}{(1-u)^3} - \frac{2}{9}\frac{u}{(1-uz)(1-u)}\right)\log^2\left(\frac{1}{1-z}\right)\log\left(\frac{1}{1-uz}\right) \\
& + \Bigg[-\frac{1}{81}\frac{(34u^2z-479u^2-102uz+532u+15)u}{(1-u)^3} - \frac{1}{81}\frac{(191u+21)u}{(1-uz)(1-u)} \\
& \quad + 4\frac{u}{(1-z)(1-u)}\Bigg]\log^2\left(\frac{1}{1-z}\right) \\
& + \left(-\frac{2}{27}\frac{(10uz-11u+1)u}{(u-1)^3} - \frac{70}{27}\frac{u}{(1-uz)(u-1)} + \frac{4}{3}\frac{1}{(1-z)(u-1)}\right) \times \\
& \times \log\left(\frac{1}{1-z}\right)\log\left(\frac{1}{1-uz}\right)
\end{aligned}$$

$$
\begin{aligned}
&- \Bigg[\frac{(288u^5z-748u^5-1052u^4z+2600u^4+1500u^3z-3056u^3-954u^2z+955u^2+194uz+568u-295)u}{81(1-u)^6} \\
&\quad + \frac{1}{81}\frac{(702u^4-2386u^3+2982u^2-1593u+283)u}{(1-uz)(1-u)^4} + \frac{4}{3}\frac{1+u}{(1-u)(1-z)}\Bigg] \times \\
&\times \log\left(\frac{1}{1-z}\right) \\
&- \Bigg[- \frac{u(168u^5z-84u^5-572u^4z+124u^4+780u^3z+208u^3-474u^2z-621u^2+74uz+512u-115)}{81(1-u)^6} \\
&\quad - \frac{1}{81}\frac{84u^4-418u^3+786u^2-633u+169}{(1-u)^4(1-z)}\Bigg] \log\left(\frac{1}{1-uz}\right) \\
&+ \frac{2(32u^4z+304u^4-150u^3z-819u^3+132u^2z+687u^2+34uz-106u-102)u}{81(1-u)^5} \\
&+ 1486\frac{(984u^4+341u^3-5883u^2+6672u-2118)u}{(1-uz)(1-u)^4} \\
&- \frac{1}{486}\frac{36u^4+887u^3-2265u^2+1878u-540}{(1-u)^4(1-z)} + k(z,u)\,. \qquad (1.80)
\end{aligned}
$$

Es muß hier eine Funktion $k(z,u)$ angegeben werden, da bei den diversen Integrationen, die mit der Hilfe eines Formelmanipulationsprogramms durchgeführt wurden, aus technischen Gründen die Ausdrücke nicht immer von 0 weg integriert wurden, sondern von einer beliebigen Grenze k weg. Dadurch sind Integrationskonstanten dazugekommen, die in $k(z,u)$ zusammengefaßt wurden.

Es läßt sich $k(z,u)$ schreiben als

$$
\begin{aligned}
k(z,u) = &\left(\frac{1}{(1-u)(1-z)} - \frac{u}{(1-u)(1-uz)}\right) g_1(u) \\
&+ \left(\frac{2uz-1-u}{u} - \frac{-3+u}{(1-u)(1-z)} - \frac{3u-1}{u(1-u)(1-uz)}\right) g_2(u) \\
&+ \Bigg[\left(-6u(2uz-1-u) - 6\frac{(u-1)^2}{1-z}\right)\log\left(\frac{1}{1-uz}\right) \\
&\quad + \left(6u(2uz-1-u) + 6\frac{u(u-1)^2}{1-uz}\right)\log\left(\frac{1}{1-z}\right) \\
&\quad - 12u(u-1) - \frac{8u+5+5u^2}{1-z} + \frac{u(8u+5+5u^2)}{1-uz}\Bigg] g_3(u) \\
&+ \Bigg[\left(\frac{2}{9}u(2uz-1-u) + \frac{2}{9}\frac{u(u-1)^2}{1-uz}\right)\log^2\left(\frac{1}{1-z}\right) \\
&\quad + \Bigg(\frac{4(-3u^4+12u^4z+3u^3-42u^3z+15u^2+48u^2z-28u-16uz+11)u}{27(1-u)^3} \\
&\qquad - \frac{4}{27}\frac{(9u^3-24u^2+21u-7)u}{(1-uz)(1-u)} - \frac{4}{3}\frac{(u-1)^2}{1-z}\Bigg)\log\left(\frac{1}{1-z}\right)
\end{aligned}
$$

$$\begin{aligned}
&+ \left(-\frac{2}{27} \frac{u(17u^3 - 57u^2 + 63u - 21)(2uz - 1 - u)}{(1-u)^3} \right. \\
&\qquad \left. - \frac{2}{27} \frac{17u^3 - 57u^2 + 63u - 21}{(1-z)(1-u)} \right) \log \left(\frac{1}{1-uz} \right) \\
&+ \frac{4}{81} \frac{(5u^3 z + 68u^3 - 10u^2 z - 227u^2 + 5uz + 250u - 85)u}{(1-u)^2} \\
&- \frac{1}{81} \frac{(30u^5 - 1669u^2 + 617u - 37 - 534u^4 + 1557u^3)u}{(1-uz)(1-u)^3} \\
&- \left. \frac{1}{81} \frac{112u^5 - 176u^4 - 137u^3 + 249u^2 + 93u - 105}{(1-u)^3(1-z)} \right] g_4(u) ,
\end{aligned} \tag{1.81}$$

mit nur von u abhängigen Funktionen $g_1(u)$, $g_2(u)$, $g_3(u)$ und $g_4(u)$.

Es sei angemerkt, daß nicht jede beliebige Wahl für diese Funktionen $g_i(u)$ eine Partikulärlösung für die Differentialgleichung (1.72) darstellt. Allerdings läßt sich die Lösung $\breve{A}(z,u)$ so darstellen, und wir können die $g_i(u)$ durch Anpassen an die Anfangswerte bestimmen.

Für die Bestimmung der Lösung von (1.72) ergeben sich also die Gleichungen

$$\begin{aligned}
\breve{A}(0,u) = 0 , \quad & \frac{\partial}{\partial z} \breve{A}(z,u) \Big|_{z=0} = 0 , \\
\frac{\partial^2}{\partial z^2} \breve{A}(z,u) \Big|_{z=0} = 4u , \quad & \frac{\partial^3}{\partial z^3} \breve{A}(z,u) \Big|_{z=0} = 28u + 12u^2 .
\end{aligned} \tag{1.82}$$

Dies führt zu folgenden Lösungen für die Funktionen $g_i(u)$:

$$\begin{aligned}
g_1(u) &= \frac{1}{243} \frac{357\,u^6 - 1415\,u^5 + 91\,u^4 + 4300\,u^3 - 5142\,u^2 + 2034\,u - 45}{(1-u)^5} , \\
g_2(u) &= \frac{1}{27} \frac{u^2\,(\,119\,u^3 - 339\,u^2 + 345\,u - 141\,)}{(1-u)^5} , \\
g_3(u) &= -\frac{1}{162} \frac{6\,u^4 - 35\,u^3 + 57\,u^2 - 36\,u + 12}{(1-u)^6} , \\
g_4(u) &= -\frac{1}{(1-u)^2} .
\end{aligned}$$

Damit bekommt man nun

$$\begin{aligned}
\breve{A}(z,u) = & \left[\frac{8}{9} \frac{(2uz - 1 - u)u^3}{(1-u)^3} + \frac{u^3}{(1-uz)(1-u)} + \frac{4}{9} \frac{u^2}{(1-z)(1-u)} \right] \times \\
& \times \int_0^z \frac{1}{1-ut} \log^3 \left(\frac{1}{1-t} \right) dt \\
& - \left[-\frac{2}{3} \frac{(8u-1)(2uz - 1 - u)u^2}{(1-u)^3} - \frac{2}{27} \frac{(70u - 3)u^2}{(1-uz)(1-u)} - \frac{4}{27} \frac{(-3+u)u}{(1-z)(1-u)} \right] \times \\
& \times \int_0^z \frac{1}{1-ut} \log^2 \left(\frac{1}{1-t} \right) dt
\end{aligned}$$

$$
\begin{aligned}
&+\Bigg[\left(-\frac{4}{9}\frac{(2uz-1-u)u^3}{(1-u)^3}-\frac{4}{9}\frac{u^3}{(1-uz)(1-u)}\right)\log^2\left(\frac{1}{1-z}\right)\\
&\quad-\left(\frac{4}{27}\frac{(10uz-11u+1)u^3}{(1-u)^3}+\frac{140}{27}\frac{u^3}{(1-uz)(1-u)}-\frac{8}{3}\frac{u^2}{(1-z)(1-u)}\right)\times\\
&\quad\times\log\left(\frac{1}{1-z}\right)+\frac{2}{81}\frac{106u^2z-131u^2-234uz+142u+117)u^2}{(1-u)^3}\\
&\quad-\frac{70}{27}\frac{u^2}{(1-uz)(1-u)}-\frac{2}{81}\frac{(79u+12)u}{(1-z)(1-u)}\Bigg]\int_0^z\frac{1}{1-ut}\log\left(\frac{1}{1-t}\right)dt\\
&-\left(\frac{2}{27}\frac{(38uz-25u-13)u^2}{(1-u)^3}+\frac{14}{27}\frac{u^2}{(1-uz)(1-u)}\right)\log^3\left(\frac{1}{1-z}\right)\\
&-\left(-\frac{2}{9}\frac{u(2uz-1-u)}{(1-u)^3}-\frac{2}{9}\frac{u}{(1-u)(1-uz)}\right)\log^2\left(\frac{1}{1-z}\right)\log\left(\frac{1}{1-uz}\right)\\
&+\Bigg[\frac{1}{81}\frac{(2u^2z+461u^2+66uz-532u+3)u}{(1-u)^3}-\frac{1}{81}\frac{(173u+39)u}{(1-uz)(1-u)}\\
&\quad+4\frac{u}{(1-z)(1-u)}\Bigg]\log^2\left(\frac{1}{1-z}\right)\\
&+\left[\frac{2}{27}\frac{(10uz-11u+1)u}{(1-u)^3}+\frac{70}{27}\frac{u}{(1-u)(1-uz)}-\frac{4}{3}\frac{1}{(1-u)(1-z)}\right]\times\\
&\times\log\left(\frac{1}{1-z}\right)\log\left(\frac{1}{1-uz}\right)\\
&-\Bigg[-\frac{1}{81}\frac{(180u^2z-730u^2-74uz+425u+199)u}{(1-u)^3}+\frac{8}{3}\frac{u}{(1-z)(1-u)}\\
&\quad-\frac{1}{81}\frac{(828u-403)u}{(1-u)(1-uz)}\Bigg]\log\left(\frac{1}{1-z}\right)\\
&-\left(\frac{1}{81}\frac{u(106uz-131u+25)}{(1-u)^3}-\frac{79}{81}\frac{1}{(1-u)(1-z)}\right)\log\left(\frac{1}{1-uz}\right)\\
&-\frac{(487u+798uz-487)u}{(1-u)^2}+\frac{341}{81}\frac{u}{(1-z)(1-u)}-\frac{92}{9}\frac{u}{(1-u)(1-uz)}\,. \qquad (1.83)
\end{aligned}
$$

Es stellt sich heraus, daß es für das Ablesen der Koeffizienten von Vorteil ist, diesen Ausdruck noch weiter umzuformen zu

$$
\begin{aligned}
\breve{A}(z,u)=&\left(\frac{4}{9}u(2uz-1-u)+\frac{2}{9}\frac{u(u-1)^2}{1-uz}+\frac{2}{9}\frac{(u-1)^2}{1-z}\right)\times\\
&\times\int_0^z\frac{1}{(1-ut)^2(1-t)^2}\log^3\left(\frac{1}{1-t}\right)dt\\
&+\Bigg[4u^2z-2u^2-\frac{2}{3}uz-\frac{5}{3}u+\frac{1}{3}+\frac{1}{27}\frac{(61u-3)(u-1)^2}{1-uz}\\
&\quad-\frac{1}{27}\frac{(7u+6)(u-1)^2}{u(1-z)}\Bigg]\int_0^z\frac{1}{(1-ut)^2(1-t)^2}\log^2\left(\frac{1}{1-t}\right)dt
\end{aligned}
$$

$$
\begin{aligned}
&-\Bigg[\left(\frac{2}{9}u(2uz-1-u)+\frac{2}{9}\frac{u(u-1)^2}{1-uz}\right)\log\left(\frac{1}{1-z}\right)^2\\
&\quad-\left(-\frac{2}{27}u(10uz-11u+1)-\frac{70}{27}\frac{u(u-1)^2}{1-uz}+\frac{4}{3}\frac{(u-1)^2}{1-z}\right)\log\left(\frac{1}{1-z}\right)\\
&\quad+\frac{218}{81}u^2z-\frac{31}{81}u^2+\frac{20}{9}uz-\frac{277}{81}u-\frac{10}{9}+\frac{1}{27}\frac{(61u+32)(u-1)^2}{1-uz}\\
&\quad+\frac{2}{81}\frac{(29u-3)(u-1)^2}{u(1-z)}\Bigg]\int_0^z\frac{1}{(1-ut)^2(1-t)^2}\log\left(\frac{1}{1-t}\right)dt\\
&+\left(-\frac{1}{9}\frac{u^2}{(1-u)(1-uz)}+\frac{1}{9}\frac{u}{(1-u)(1-z)}\right)\log^4\left(\frac{1}{1-z}\right)\\
&-\Bigg[\frac{2}{9}\frac{(1+u)(2uz-1-u)u}{(1-u)^3}+\frac{2}{81}\frac{(65u+3)u}{(1-uz)(1-u)}-\frac{4}{81}\frac{28u-3}{(1-z)(1-u)}\\
&\quad+\frac{2}{9}\frac{1}{(1-z)^2}\Bigg]\log^3\left(\frac{1}{1-z}\right)-\\
&-\left(-\frac{2}{9}\frac{(1+u)(2uz-1-u)u}{(1-u)^3}-\frac{2}{9}\frac{(1+u)u}{(1-uz)(1-u)}\right)\log^2\left(\frac{1}{1-z}\right)\log\left(\frac{1}{1-uz}\right)\\
&+\Bigg[\frac{2}{27}\frac{(2u^2z+11u^2-22uz+10u-1)u}{(1-u)^3}+\frac{1}{81}\frac{407u-12}{(1-z)(1-u)}\\
&\quad-\frac{1}{27}\frac{11u-6}{(1-z)^2u}-\frac{1}{81}\frac{(473u+126)u}{(1-uz)(1-u)}+\frac{1}{9}\frac{(3u+1)u}{(1-uz)^2}\Bigg]\log^2\left(\frac{1}{1-z}\right)\\
&+\Bigg[\frac{2}{27}\frac{(1+u)(10uz-11u+1)u}{(1-u)^3}+\frac{70}{27}\frac{(1+u)u}{(1-uz)(1-u)}\\
&\quad-\frac{4}{3}\frac{1+u}{(1-z)(1-u)}\Bigg]\log\left(\frac{1}{1-z}\right)\log\left(\frac{1}{1-uz}\right)\\
&-\Bigg[\frac{1}{81}\frac{(98u^2z+41u^2+194uz-290u-43)u}{(1-u)^3}-\frac{1}{81}\frac{(177u-271)u}{(1-uz)(1-u)}\\
&\quad-\frac{1}{27}\frac{(61u+32)u}{(1-uz)^2}+\frac{2}{81}\frac{89u+33}{(1-z)(1-u)}-\frac{2}{81}\frac{8u-21}{(1-z)^2u}\Bigg]\log\left(\frac{1}{1-z}\right)\\
&-\Bigg[-\frac{1}{81}\frac{(218u^2z-31u^2+74uz-146u-115)u}{(1-u)^3}-\frac{1}{27}\frac{(61u+32)u}{(1-uz)(1-u)}\\
&\quad-\frac{1}{81}\frac{58u+73}{(1-z)(1-u)}\Bigg]\log\left(\frac{1}{1-uz}\right)\\
&-\frac{4}{81}\frac{91u^2z+48u^2-18uz-57u+9}{(1-u)^2}-\frac{1}{27}\frac{129u-26}{(1-uz)(1-u)}\\
&+\frac{1}{81}\frac{287u^2-92u-42}{(1-z)u(1-u)}+\frac{2}{81}\frac{46u+21}{u(1-z)^2}\,.
\end{aligned}
\tag{1.84}
$$

Für das Ablesen der Koeffizienten von obigen Ausdruck (1.84) benötigt man verschiedene Beziehungen für die Harmonischen Zahlen.

Nachfolgend seien die hier relevanten zusammengestellt.

$$\sum_{k=1}^{n} H_k = (n+1)H_n - n\,, \tag{1.85a}$$

$$\sum_{k=1}^{n} H_k^{(2)} = (n+1)H_n^{(2)} - H_n\,, \tag{1.85b}$$

$$\sum_{k=1}^{n} H_k^{(3)} = (n+1)H_n^{(3)} - H_n^{(2)}\,, \tag{1.85c}$$

$$\sum_{k=1}^{n} H_k^2 = (n+1)H_n^2 - (2n+1)H_n + 2n\,, \tag{1.85d}$$

$$\sum_{k=1}^{n} H_k^3 = (n+1)H_n^3 - \left(3n+\frac{3}{2}\right)H_n^2 + (6n+3)H_n + \frac{1}{2}H_n^{(2)} - 6n\,, \tag{1.85e}$$

$$\sum_{k=1}^{n} H_k H_k^{(2)} = (n+1)H_n H_n^{(2)} - \left(n+\frac{1}{2}\right)H_n^{(2)} - \frac{1}{2}H_n^2 + H_n\,, \tag{1.85f}$$

$$\sum_{k=1}^{n} \frac{H_k}{k} = \frac{1}{2}H_n^2 + \frac{1}{2}H_n^{(2)}\,, \tag{1.85g}$$

$$\sum_{k=1}^{n} kH_k = \frac{1}{2}n^2H_n + \frac{1}{2}nH_n - \frac{1}{4}n^2 + \frac{1}{4}n\,, \tag{1.85h}$$

$$\sum_{k=1}^{n} k^2H_k = \frac{n(n+1)(2n+1)}{6}H_n - \frac{n(n-1)(4n+1)}{36}\,, \tag{1.85i}$$

$$\sum_{k=1}^{n-1} kH_{n-k} = \frac{1}{2}n(n+1)\left(H_{n+1} - \frac{3}{2}\right)\,, \tag{1.85j}$$

$$\sum_{k=1}^{n} \frac{H_{n-k}}{k} = H_n^2 - H_n^{(2)}\,, \tag{1.85k}$$

$$\sum_{k=1}^{n-1} H_kH_{n-k} = (n+1)\left(H_n^2 - H_n^{(2)}\right) - 2nH_n + 2n\,, \tag{1.85l}$$

$$\sum_{k=1}^{n-1} kH_kH_{n-k} = \frac{1}{2}n(n+1)\left(H_n^2 - H_n^{(2)}\right) - n^2H_n + n^2\,, \tag{1.85m}$$

$$\begin{aligned}\sum_{k=1}^{n-1} k^2H_kH_{n-k} &= \frac{1}{6}n(n+1)(2n+1)H_n^2 - \frac{1}{18}n(13n^2+6n-1) - \\ &\quad - \frac{1}{6}n(n+1)(2n+1)H_n^{(2)} + \frac{1}{108}n(71n^2+30n+7)\,,\end{aligned} \tag{1.85n}$$

$$\begin{aligned}\sum_{k=1}^{n-1} H_kH_{n-k}^2 &= (n+1)\sum_{k=1}^{n-1}\frac{H_{n-k}}{k^2} + (n+1)H_n^3 - (3n+1)H_n^2 + (6n+1)H_n - \\ &\quad - 3(n+1)H_nH_n^{(2)} + 2(n+1)H_n^{(3)} + (2n+1)H_n^{(2)} - 6n\,,\end{aligned} \tag{1.85o}$$

$$\sum_{k=1}^{n-1} kH_k H_{n-k}^2 = \frac{1}{2}n(n+1)\sum_{k=1}^{n-1}\frac{H_{n-k}}{k^2} + \frac{1}{2}n(n+1)H_n^3 - \frac{1}{4}(7n^2+5n+2)H_n^2 + \\ + \frac{1}{4}(15n^2+5n-1)H_n - \frac{3}{2}n(n+1)H_nH_n^{(2)} + n(n+1)H_n^{(3)} + \\ + \frac{1}{2}(3n^2+3n+1)H_n^{(2)} - \frac{1}{8}n(31n+7)\,, \tag{1.85p}$$

$$\sum_{k=1}^{n-1} k^2 H_k H_{n-k}^2 = \frac{1}{6}n(n+1)(2n+1)\sum_{k=1}^{n-1}\frac{H_{n-k}}{k^2} + \frac{1}{6}n(n+1)(2n+1)H_n^3 - \\ - \frac{1}{12}(16n^3+21n^2+11n+2)H_n^2 + \\ + \frac{1}{36}(104n^3+81n^2+7n-1)H_n - \frac{1}{2}n(n+1)(2n+1)H_nH_n^{(2)} + \\ + \frac{1}{3}n(n+1)(2n+1)H_n^{(3)} + \frac{1}{18}(2n+1)(11n^2+11n+3)H_n^{(2)} - \\ - \frac{1}{216}n(640n^2+435n+71)\,, \tag{1.85q}$$

$$\sum_{k=1}^{n-1}\frac{H_kH_{n-k}}{k} = \sum_{k=1}^{n-1}\frac{H_{n-k}}{k^2} + \frac{1}{2}H_n^3 - \frac{3}{2}H_nH_n^{(2)} + H_n^{(3)}\,, \tag{1.85r}$$

$$\sum_{k=1}^{n-1}\frac{H_{n-k}^2}{k} = \sum_{k=1}^{n-1}\frac{H_{n-k}}{k^2} + H_n^3 - 3H_nH_n^{(2)} + 2H_n^{(3)}\,. \tag{1.85s}$$

Von besonderer Bedeutung sind auch wieder die Beziehungen (1.19).

Beim Ablesen der Koeffizienten von (1.84) sind einige Terme darunter, für die sich dieses Entwickeln um $z = 0$ und $u = 0$ als aufwendiger erweist. Diese seien nachfolgend angeführt.

$$[z^nu^j]\int\frac{1}{(1-z)^2(1-uz)^2}\log^3\left(\frac{1}{1-z}\right)dz \\ = \frac{(j+1)(n-j)}{n}\left((H_{n-j}-1)^3 - 3\left(H_{n-j}-1\right)\left(H_{n-j}^{(2)}-1\right) + 2\left(H_{n-j}^{(3)}-1\right)\right)\,, \tag{1.86a}$$

$$[z^nu^j]\frac{1}{1-uz}\int\frac{1}{(1-z)^2(1-uz)^2}\log^3\left(\frac{1}{1-z}\right)dz \\ = (n-j)\left(j+1-(n-j-1)\left(H_n-H_{n-j-1}\right)\right)\times \\ \times\left((H_{n-j}-1)^3 - 3\left(H_{n-j}-1\right)\left(H_{n-j}^{(2)}-1\right) + 2\left(H_{n-j}^{(3)}-1\right)\right)\,, \tag{1.86b}$$

$$[z^nu^j]\frac{1}{1-z}\int\frac{1}{(1-z)^2(1-uz)^2}\log^3\left(\frac{1}{1-z}\right)dz \\ = -j(j+1)\sum_{k=j+1}^{n}\frac{H_{k-j}^3}{k} + 3j(j+1)\sum_{k=j+1}^{n}\frac{H_{k-j}^2}{k} + 3j(j+1)\sum_{k=j+1}^{n}\frac{H_{k-j}\,H_{k-j}^{(2)}}{k} - \\ - 6j(j+1)\sum_{k=j+1}^{n}\frac{H_{k-j}}{k} - 3j(j+1)\sum_{k=j+1}^{n}\frac{H_{k-j}^{(2)}}{k} - 2j(j+1)\sum_{k=j+1}^{n}\frac{H_{k-j}^{(3)}}{k} +$$

$$+ (j+1)(n-j+1)H_{n-j}^3 - 3(j+1)(2n-2n+1)H_{n-j}^2 -$$
$$- 3(j+1)(n-j+1)H_{n-j}H_{n-j}^{(2)} + 6(j+1)(3n-3j+1)H_{n-j} +$$
$$+ 3(j+1)(2n-2j+1)H_{n-j}^{(2)} + 2(j+1)(n-j+1)H_{n-j}^{(3)} + 6j(j+1)H_n -$$
$$- 6j(j+1)H_j - 24(j+1)(n-j)\,, \tag{1.86c}$$

$$[z^n u^j] \int \frac{1}{(1-z)^2(1-uz)^2} \log^2\left(\frac{1}{1-z}\right) dz$$
$$= \frac{(j+1)(n-j)}{n} \left((H_{n-j} - 1) - \left(H_{n-j}^{(2)} - 1 \right) \right) , \tag{1.86d}$$

$$[z^n u^j] \frac{1}{1-uz} \int \frac{1}{(1-z)^2(1-uz)^2} \log^2\left(\frac{1}{1-z}\right) dz$$
$$= (n-j)\left(j+1-(n-j-1)\left(H_n - H_{n-j-1}\right)\right) \left((H_{n-j} - 1)^2 - \left(H_{n-j}^{(2)} - 1 \right) \right) , \tag{1.86e}$$

$$[z^n u^j] \frac{1}{1-z} \int \frac{1}{(1-z)^2(1-uz)^2} \log^2\left(\frac{1}{1-z}\right) dz$$
$$= -j(j+1) \sum_{k=j+1}^{n} \frac{H_{k-j}^2}{k} + 2j(j+1) \sum_{k=j+1}^{n} \frac{H_{k-j}}{k} + j(j+1) \sum_{k=j+1}^{n} \frac{H_{k-j}^{(2)}}{k} +$$
$$+ (j+1)(n-j+1)H_{n-j}^2 - 2(j+1)(2n-2j+1)H_{n-j} -$$
$$- (j+1)(n-j+1)H_{n-j}^{(2)} - 2j(j+1)H_n + 2j(j+1)H_j + 6(j+1)(n-j)\,, \tag{1.86f}$$

$$[z^n u^j] \int \frac{1}{(1-z)^2(1-uz)^2} \log\left(\frac{1}{1-z}\right) dz$$
$$= \frac{(j+1)(n-j)}{n} \left(H_{n-j} - 1 \right) , \tag{1.86g}$$

$$[z^n u^j] \frac{1}{1-uz} \int \frac{1}{(1-z)^2(1-uz)^2} \log\left(\frac{1}{1-z}\right) dz$$
$$= (n-j)\left(j+1-(n-j-1)\left(H_n - H_{n-j-1}\right)\right) \left(H_{n-j} - 1 \right) , \tag{1.86h}$$

$$[z^n u^j] \frac{1}{1-z} \int \frac{1}{(1-z)^2(1-uz)^2} \log\left(\frac{1}{1-z}\right) dz$$
$$= -j(j+1) \sum_{k=j+1}^{n} \frac{H_{k-j}}{k} + (j+1)(n-j+1)H_{n-j} + j(j+1)H_n - j(j+1)H_j -$$
$$- 2(j+1)(n-j)\,, \tag{1.86i}$$

$$
\begin{aligned}
&[z^n u^j]\frac{1}{(1-z)(1-uz)}\log\left(\frac{1}{1-z}\right)\int\frac{1}{(1-z)^2(1-uz)^2}\log\left(\frac{1}{1-z}\right)\,dz\\
&\quad= -\sum_{k=1}^{n-j-1} k^2 H_k H_{n-k} H_{n-j-k} + (2n-2j-1)\sum_{k=1}^{n-j-1} k H_k H_{n-k} H_{n-j-k} -\\
&\qquad - (n-j)(n-j-1)\sum_{k=1}^{n-j-1} H_k H_{n-k} H_{n-j-k} + \sum_{k=1}^{n-j-1} k^2 H_k H_{n-k} -\\
&\qquad - (2n-2j-1)\sum_{k=1}^{n-j-1} k H_k H_{n-k} + (n-j)(n-j-1)\sum_{k=1}^{n-j-1} H_k H_{n-k} +\\
&\qquad + \frac{1}{3}(n-j+1)(n-j)(n-j-1)\sum_{k=1}^{n-j-1}\frac{H_{n-j-k}}{k^2} +\\
&\qquad + \frac{1}{3}(n-j+1)(n-j)(n-j-1)H_{n-j}^3 -\\
&\qquad - \frac{1}{6}(7n^3-6n^2-24jn^2+9jn+27j^2n-7n+7j-2-10j^3-3j^2)H_{n-j}^2 -\\
&\qquad - (n-j+1)(n-j)(n-j-1)H_{n-j}H_{n-j}^{(2)} +\\
&\qquad + \frac{1}{18}(38n^3-141jn^2-39n^2+69jn+168j^2n-38n+38j-65j^3-\\
&\qquad - 30j^2-5)H_{n-j} +\\
&\qquad + \frac{1}{18}(10n^3-12n^2-39jn^2+15jn+48j^2n-16n+16j-6-19j^3-3j^2)H_{n-j}^{(2)} +\\
&\qquad + \frac{2}{3}(n-j+1)(n-j)(n-j-1)H_{n-j}^{(3)} -\\
&\qquad - \frac{1}{108}(n-j)(202n^2-593jn-279n+391j^2-187+252j)\,,
\end{aligned}
\tag{1.86j}
$$

$$
\begin{aligned}
&[z^n u^j]\frac{1}{1-uz}\log^2\left(\frac{1}{1-z}\right)\int\frac{1}{(1-z)^2(1-uz)^2}\log\left(\frac{1}{1-z}\right)\,dz\\
&\quad= -2\sum_{k=1}^{n-j-1} k H_k H_{n-k} H_{n-j-k} + 2(2n-2n-1)\sum_{k=1}^{n-j-1} H_k H_{n-k} H_{n-j-k} -\\
&\qquad - 2(n-j)(n-j-1)\sum_{k=1}^{n-j-1}\frac{H_k H_{n-k} H_{n-j-k}}{k} + 2\sum_{k=1}^{n-j-1} k H_k H_{n-k} -\\
&\qquad - 2(2n-2j-1)\sum_{k=1}^{n-j-1} H_k H_{n-k} + 2(n-j)(n-j-1)\sum_{k=1}^{n-j-1}\frac{H_k H_{n-k}}{k} +\\
&\qquad + 2\sum_{k=1}^{n-j-1} H_{n-k} H_{n-j-k} - 2(2n-2j-1)\sum_{k=1}^{n-j-1}\frac{H_{n-k} H_{n-j-k}}{k} +\\
&\qquad + 2(n-j)(n-j-1)\sum_{k=1}^{n-j-1}\frac{H_{n-k} H_{n-j-k}}{k^2} + 2(2n-2j-1)\sum_{k=1}^{n-j-1}\frac{H_{n-k}}{k} -
\end{aligned}
$$

$$
\begin{aligned}
&- 2(n-j)(n-j-1)\sum_{k=1}^{n-j-1}\frac{H_{n-k}}{k^2} - 3(n-j)(n-j-1)\sum_{k=1}^{n-j-1}\frac{H_{n-j-k}}{k^2} + \\
&+ (n-j)(n-j-1)\sum_{k=1}^{n-j-1}\frac{H_{n-j-k}^2}{k^2} - (n-j)(n-j-1)\sum_{k=1}^{n-j-1}\frac{H_{n-j-k}^{(2)}}{k^2} + \\
&+ (n-j)(n-j-1)H_{n-j}^4 - (4n^2-9jn-4n+5j^2+4j-1)H_{n-j}^3 - \\
&-6(n-j)(n-j-1)H_{n-j}^2H_{n-j}^{(2)} + \frac{1}{2}(23n^2-52jn-19n+29j^2+19j-8)H_{n-j}^2 + \\
&+ 8(n-j)(n-j-1)H_{n-j}H_{n-j}^{(2)} + 3(4n^2-9jn-4n+5j^2+4j-1)H_{n-j}H_{n-j}^{(2)} - \\
&- \frac{1}{2}\frac{45n^3-147jn^2-41n^2+82jn+159j^2n-7n-57j^3-41j^2+7j-4}{n-j}H_{n-j} + \\
&+ 3(n-j)(n-j-1)\left(H_{n-j}^{(2)}\right)^2 - (8n^2-19jn-6n+11j^2+6j-4)H_{n-j}^{(2)} - \\
&- 2(4n^2-9jn-4n+5j^2+4j-1)H_{n-j}^{(3)} - 6(n-j)(n-j-1)H_{n-j}^{(4)} - \\
&- 2nH_n + 2(j+1)H_j + \\
&+ \frac{1}{4}\frac{1}{(n-j)^2}(89n^4-380jn^3-87n^3+606j^2n^2+261jn^2-428j^3n-261j^2n + \\
&\qquad + 113j^4+87j^3-8)\,,
\end{aligned}
\tag{1.86k}
$$

$$
\begin{aligned}
&[z^nu^j]\frac{1}{(1-uz)(1-u)}\log\left(\frac{1}{1-uz}\right)\log^2\left(\frac{1}{1-z}\right) \\
&= 2\sum_{k=n-j}^{n}\frac{H_kH_{n-k}}{k} - 2\sum_{k=n-j}^{n}\frac{H_{n-k}}{k^2}\,,
\end{aligned}
\tag{1.86l}
$$

$$
\begin{aligned}
&[z^nu^j]\frac{1}{(1-u)^3}\log\left(\frac{1}{1-uz}\right)\log^2\left(\frac{1}{1-z}\right) \\
&= \frac{(j+2)(j+1)}{n}\sum_{k=1}^{j}\frac{H_{n-k}}{k} - \frac{1}{2}\frac{(n-j-1)(n-j-2)}{n}H_{n-j}^2 + \\
&+ \frac{n^4-2jn^3+j^2n^2-jn^2-3n^2+2j^2n+6jn+4n-j^3-3j^2-2j}{(n-j)n^2}H_{n-j} + \\
&+ \frac{1}{2}\frac{(n-j-1)(n-j-2)}{n}H_{n-j}^{(2)} + \frac{1}{2}\frac{(n-j-1)(n-j-2)}{n}H_n^2 - \\
&- \frac{n^3-2jn-3n+2j^2+6j+4}{n^2}H_n - \frac{1}{2}\frac{(n-j-1)(n-j-2)}{n}H_n^{(2)} - \\
&- \frac{(j+1)(j+2)}{n^2}H_j + \\
&+ \frac{j(n^5-2jn^4+j^2n^3-n^3+4jn^2+6n^2-5j^2n-12jn-6n+2j^3+6j^2+4j)}{n^3(n-j)^2}\,.
\end{aligned}
\tag{1.86m}
$$

Unter Zuhilfenahme dieser Formeln kommt man nach Vereinfachungen schon zu

Hilfssatz 1.4.4. *Das zweite Faktorielle Moment $\breve{A}_{n,j}$ der Höhe des Knotens mit Schlüssel j in einem p-tree der Größe n ist gegeben durch*

$$
\begin{aligned}
&\breve{A}_{n,j} = \\
&\frac{1}{9}H_{n-j}^4 - \frac{4}{9}H_{n-j}^3 H_n + \left(\frac{100}{81} + \frac{8}{9}\frac{(j-1)^2}{n} - \frac{8}{9}\frac{(j-1)(j-2)}{n-1} + \frac{4}{9}\frac{1}{n-j+1}\right) H_{n-j}^3 + \\
&+ \frac{2}{3}H_{n-j}^2 H_{n-j}^{(2)} + \left(-\frac{134}{27} - \frac{4}{3}\frac{1}{n-j+1}\right) H_{n-j}^2 H_n + \frac{4}{3}H_{n-j}H_{n-j}^{(2)}H_n - \frac{4}{9}H_n^3 + \\
&+ \left(\frac{1}{81}\frac{125j-456}{j} + \frac{8}{27}\frac{14j-3}{j(n-j+1)} + \frac{2}{27}\frac{74-103j+23j^2}{n} + \frac{8}{9}\frac{(j-1)^2}{n^2} - \right. \\
&\quad \left. - \frac{2}{27}\frac{112-123j+23j^2}{n-1} - \frac{8}{9}\frac{(j-1)(j-2)}{(n-1)^2}\right) H_{n-j}^2 + \\
&+ \left(\frac{56}{9} - \frac{8}{3}\frac{(j-1)^2}{n} + \frac{8}{3}\frac{(j-1)(j-2)}{n-1} + \frac{4}{3}\frac{1}{n-j+1}\right) H_{n-j}H_{n-j}^{(2)} - \frac{16}{9}H_{n-j}H_{n-j}^{(3)} + \\
&+ \left(\frac{70}{27} - \frac{280}{27}\frac{1}{n-j+1}\right) H_{n-j}H_n + \left(\frac{70}{27} + \frac{4}{9}\frac{1}{n-j} - \frac{8}{9}\frac{1}{n-j+1}\right) H_{n-j}H_j - \\
&- \left(H_{n-j}^{(2)}\right)^2 + \left(\frac{134}{27} + \frac{4}{3}\frac{1}{n-j+1}\right) H_n H_{n-j}^{(2)} - \frac{8}{9}H_n H_{n-j}^{(3)} + \\
&+ \left(-\frac{68}{27} - \frac{8}{9}\frac{(j-1)^2}{n-1} + \frac{4}{9}\frac{2j^2-2j+1}{n}\right) H_n^2 + \frac{4}{3}H_n H_n^{(2)} + \\
&+ \left(-\frac{1}{81}\frac{72+24j+137j^2}{j^2} - \frac{2}{81}\frac{235j^2-212j-137}{n} - \frac{4}{27}\frac{13j^2-11j+1}{n^2} - \right. \\
&\quad - \frac{8}{9}\frac{(j-1)^2}{n^3} + \frac{2}{81}\frac{235j^2-591j+326}{n-1} + \frac{4}{27}\frac{13j^2-33j+26}{(n-1)^2} + \\
&\quad \left. + \frac{8}{9}\frac{(j-1)(j-2)}{(n-1)^3} + \frac{2}{81}\frac{173j-210}{j(n-j+1)}\right) H_{n-j} + \\
&+ \left(-\frac{1}{81}\frac{335j-456}{j} + \frac{4}{27}\frac{7j+6}{j(n-j+1)} - \frac{2}{27}\frac{74-103j+23j^2}{n} - \frac{8}{9}\frac{(j-1)^2}{n^2} + \right. \\
&\quad \left. + \frac{2}{27}\frac{112-123j+23j^2}{n-1} + \frac{8}{9}\frac{(j-1)(j-2)}{(n-1)^2}\right) H_{n-j}^{(2)} + \\
&+ \left(-\frac{604}{81} + \frac{16}{9}\frac{(j-1)^2}{n} - \frac{16}{9}\frac{(j-1)(j-2)}{n-1} - \frac{16}{9}\frac{1}{n-j+1}\right) H_{n-j}^{(3)} + 2H_{n-j}^{(4)} + \\
&+ \left(-\frac{1}{81}\frac{79j^2-24j-72}{j^2} + \frac{4}{27}\frac{j(2j-3)}{n-1} + \frac{8}{9}\frac{j(2j-1)}{(n-1)^2} - \frac{2}{27}\frac{j(4j-3)}{n} - \right. \\
&\quad \left. - \frac{4}{9}\frac{j(4j-3)}{n^2} + \frac{140}{27}\frac{1}{n-j} - \frac{140}{27}\frac{j-1}{j(n-j+1)}\right) H_n + \\
&+ \left(\frac{68}{27} - \frac{4}{9}\frac{2j^2-2j+1}{n} + \frac{8}{9}\frac{(j-1)^2}{n-1}\right) H_n^{(2)} - \frac{8}{9}H_n^{(3)} + \\
&+ \left(\frac{1}{81}\frac{29j^2-24j-72}{j^2} - \frac{4}{27}\frac{41j+35}{j(n-j+1)} - \frac{4}{27}\frac{j(2j-3)}{n-1} - \frac{8}{9}\frac{j(2j-3)}{(n-1)^2} + \right. \\
&\quad \left. + \frac{2}{27}\frac{j(4j-3)}{n} + \frac{4}{9}\frac{j(4j-3)}{n^2} + \frac{8}{9}\frac{1}{n-j} - \frac{4}{9}\frac{1}{(n-j)^2}\right) H_j +
\end{aligned}
\tag{1.87}
$$

$$
\begin{aligned}
&+\frac{4}{9}\frac{j(4j-3)}{n^3}+\frac{2}{27}\frac{(4j-3)(j-6)}{n^2}+\frac{1}{81}\frac{532j^2-9j+18}{n}-\frac{16}{9}\frac{(j-1)^2}{(n-1)^3}-\\
&-\frac{8}{27}\frac{(j-1)(j-10)}{(n-1)^2}-\frac{4}{81}\frac{133j^2-257j+142}{n-1}-\frac{92}{9}\frac{1}{n-j+1}+\frac{341}{81}-\\
&-\frac{4}{9}\sum_{k=j+1}^{n}\frac{H_{k-j}^3}{k}+\frac{8}{27}\sum_{k=j+1}^{n}\frac{H_{k-j}^2}{k}+\frac{4}{3}\sum_{k=j+1}^{n}\frac{H_{k-j}H_{k-j}^{(2)}}{k}+\\
&+\left(\frac{2}{81}\frac{91j+36}{j}-\frac{4}{27}\frac{(j-1)(j-2)}{n-1}-\frac{(j-1)(j-2)}{(n-1)^2}+\frac{4}{27}\frac{(j-1)^2}{n}+\right.\\
&\qquad\left.+\frac{8}{9}\frac{(j-1)^2}{n^2}\right)\sum_{k=j+1}^{n}\frac{H_{k-j}}{k}-\\
&-\frac{8}{27}\sum_{k=j+1}^{n}\frac{H_{k-j}^{(2)}}{k}-\frac{8}{9}\sum_{k=j+1}^{n}\frac{H_{k-j}^{(3)}}{k}-\frac{4}{9}\sum_{k=1}^{n-j-1}\frac{H_{n-j-k}^2}{k^2}+\\
&+\left(-\frac{140}{27}-\frac{8}{9}\frac{1}{n-j+1}+\frac{(j-1)^2}{n}-\frac{8}{9}\frac{(j-1)(j-2)}{n-1}\right)\sum_{k=1}^{n-j-1}\frac{H_{n-j-k}}{k^2}+\\
&+\frac{4}{9}\sum_{k=1}^{n-j-1}\frac{H_{n-j-k}^{(2)}}{k^2}+\left(\frac{8}{9}\frac{(j-1)(j-2)}{n-1}-\frac{8}{9}\frac{(j-1)^2}{n}\right)\sum_{k=1}^{n-j-1}\frac{H_kH_{n-j-k}}{k}+\\
&+\frac{8}{9}\sum_{k=1}^{n-j-1}\frac{H_kH_{n-k}H_{n-j-k}}{k}+\\
&+\left(\frac{68}{27}+\frac{8}{9}\frac{(j-1)(j-2)}{n-1}-\frac{8}{9}\frac{(j-1)^2}{n}\right)\sum_{k=1}^{n-j-1}\frac{H_kH_{n-j-k}}{n-k}+\\
&+\left(\frac{4}{9}+\frac{8}{9}\frac{1}{n-j+1}\right)\sum_{k=1}^{n-j-1}\frac{H_kH_{n-k}}{k}+\frac{140}{27}\frac{1}{n-j+1}\sum_{k=1}^{n-j-1}\frac{H_{n-k}}{k}+\\
&+\frac{140}{27}\frac{1}{n-j+1}\sum_{k=1}^{n-j-1}\frac{H_{n-k}}{n-j-k}+\\
&+\left(\frac{2}{27}\frac{17j+70}{j}-\frac{8}{9}\frac{j^2-j+1}{n-1}+\frac{8}{9}\frac{1}{j(n-j+1)}+\right.\\
&\qquad\left.+\frac{4}{9}\frac{(2j+1)(j-1)}{n}\right)\sum_{k=1}^{n-j-1}\frac{H_k}{n-k}+\\
&+\left(\frac{140}{27}+\frac{8}{9}\frac{1}{n-j+1}\right)\sum_{k=1}^{n-j-1}\frac{H_kH_{n-k}}{n-j-k}-\frac{8}{9}\sum_{k=1}^{n-j-1}\frac{H_{n-k}H_{n-j-k}}{k^2}+\\
&+\left(-\frac{4}{9}-\frac{8}{9}\frac{1}{n-j+1}\right)\sum_{k=1}^{n-j-1}\frac{H_{n-k}}{k^2}+\\
&+\left(-\frac{8}{3}+\frac{8}{9}\frac{(j-1)^2}{n-1}-\frac{4}{9}\frac{2j^2-2j+1}{n}\right)\sum_{k=1}^{j}\frac{H_{n-k}}{k}+\frac{140}{27}\sum_{k=n-j}^{n-1}\frac{H_{n-k}}{k}+
\end{aligned}
$$

$$+\frac{8}{9}\sum_{k=n-j}^{n-1}\frac{H_kH_{n-k}}{k}-\frac{8}{9}\sum_{k=n-j}^{n-1}\frac{H_{n-k}}{k^2} \quad \text{für} \quad 1\le j<n\,,$$

$$\breve{A}_{n,n}=0\,. \tag{1.88}$$

1.4.4 Die Varianz

Bezeichnen wir nun mit $\hat{A}_{n,j}$ die Varianz der Höhe des Knotens mit Schlüssel j in einem p-tree der Größe n, so ist diese mittels der Beziehung

$$\hat{A}_{n,j}=\breve{A}_{n,j}+A_{n,j}-(A_{n,j})^2 \tag{1.89}$$

aus den in den vorigen Unterabschnitten 1.4.1 und 1.4.2 errechneten Größen einfach zu erhalten.

Damit erhalten wir schon

Satz 1.4.5. *Die Varianz $\hat{A}_{n,j}$ der Höhe des Knotens mit Schlüssel j in einem p-tree der Größe n ist gegeben durch*

$$
\begin{aligned}
\hat{A}_{n,j} &= \\
&-\frac{4}{9}H_{n-j}^3H_n+\left(\frac{10}{81}+\frac{4}{9}\frac{(j-1)^2}{n}-\frac{4}{9}\frac{(j-1)(j-2)}{n-1}\right)H_{n-j}^3+\frac{8}{9}H_{n-j}^2H_{n-j}^{(2)}+\\
&+\left(-\frac{4}{3}\frac{1}{n-j+1}-\frac{140}{27}\right)H_{n-j}^2H_n+\frac{2}{9}H_{n-j}^2H_j+\frac{4}{3}H_{n-j}H_{n-j}^{(2)}H_n-\frac{4}{9}H_n^3+\\
&+\left(-\frac{1}{81}\frac{37j+456}{j}+\frac{2}{27}\frac{12j^4-60j^3+103j^2-39j-28}{n-1}-\right.\\
&\qquad-\frac{4}{9}\frac{(j-1)(j-2)(j^2-3j+4)}{(n-1)^2}+\frac{8}{27}\frac{2j-3}{j(n-j+1)}-\frac{4}{9}\frac{1}{(n-j+1)^2}-\\
&\qquad\left.-\frac{2}{27}\frac{12j^4-60j^3+103j^2-44j-8}{n}-\frac{4}{9}\frac{(j^2-2j-1)(j-1)^2}{n^2}\right)H_{n-j}^2+\\
&+\left(\frac{22}{3}+\frac{20}{9}\frac{(j-1)(j-2)}{n-1}+\frac{16}{9}\frac{1}{n-j+1}-\frac{22}{9}\frac{(j-1)^2}{n}\right)H_{n-j}H_{n-j}^{(2)}-\\
&-\frac{16}{9}H_{n-j}H_{n-j}^{(3)}+\\
&+\left(\frac{40}{27}+\frac{4}{9}\frac{(j-1)(j-2)}{n-1}-\frac{292}{27}\frac{1}{n-j+1}-\frac{4}{9}\frac{(j-1)^2}{n}\right)H_{n-j}H_n+\\
&+\left(\frac{100}{27}-\frac{4}{9}\frac{1}{n-j+1}-\frac{4}{9}\frac{(j-1)(j-2)}{n-1}+\frac{4}{9}\frac{1}{n-j}+\frac{4}{9}\frac{(j-1)^2}{n}\right)H_{n-j}H_j-\\
&-\frac{10}{9}\left(H_{n-j}^{(2)}\right)^2+\left(\frac{4}{3}\frac{1}{n-j+1}+\frac{140}{27}\right)H_{n-j}^{(2)}H_n-\frac{2}{9}H_{n-j}^{(2)}H_j-\frac{8}{9}H_{n-j}^{(3)}H_n+\\
&+\left(-\frac{71}{27}+\frac{4}{9}\frac{2j^2-2j+1}{n}-\frac{8}{9}\frac{(j-1)^2}{n-1}\right)H_n^2+\frac{4}{3}H_nH_n^{(2)}+\frac{2}{9}H_nH_j-\frac{1}{9}H_j^2+
\end{aligned}
$$

$$
\begin{aligned}
&+\Bigg(\frac{2}{81}\frac{1}{n^3(n-1)^3}(89n^6-91n^5j-65n^5+82n^4j^2-99n^4j+11n^4-162n^3j^3+\\
&\quad+460n^3j^2-279n^3j+159n^3+72n^2j^4-36n^2j^3-332n^2j^2+229n^2j-\\
&\quad-56n^2-72nj^4+198nj^3-210nj^2+204nj-102n+36j^2-72j+36)-\\
&\quad-\frac{4}{81}\frac{9-77n+132n^2+8n^3}{(n+1)n(n-1)(n-j+1)}-\frac{8}{3}\frac{1}{(n-j+1)^2}-\\
&\quad-\frac{4}{27}\frac{2n+37}{j(n+1)}-\frac{8}{9}\frac{1}{j^2}\Bigg)H_{n-j}+\\
&+\Bigg(-\frac{2}{81}\frac{199j-228}{j}-\frac{2}{27}\frac{35j^2-112j+74}{n}-\frac{8}{9}\frac{(j-1)^2}{n^2}+\frac{8}{27}\frac{8j+3}{j(n-j+1)}+\\
&\quad+\frac{2}{27}\frac{35j^2-147j+124}{n-1}+\frac{8}{9}\frac{(j-1)(j-2)}{(n-1)^2}\Bigg)H_{n-j}^{(2)}+\\
&+\Bigg(-\frac{604}{81}+\frac{16}{9}\frac{(j-1)^2}{n}-\frac{16}{9}\frac{(j-1)(j-2)}{n-1}-\frac{16}{9}\frac{1}{n-j+1}\Bigg)H_{n-j}^{(3)}+2H_{n-j}^{(4)}+\\
&+\Bigg(-\frac{8}{81}\frac{(2j+3)(j-3)}{j^2}+\frac{4}{27}\frac{j(4j-3)}{n}-\frac{4}{9}\frac{j(4j-3)}{n^2}-\frac{4}{27}\frac{44j-35}{j(n-j+1)}+\\
&\quad+\frac{140}{27}\frac{1}{n-j}-\frac{4}{27}\frac{4j^2-9j+6}{n-1}+\frac{8}{9}\frac{j(2j-3)}{(n-1)^2}\Bigg)H_n+\\
&+\Bigg(\frac{68}{27}-\frac{4}{9}\frac{2j^2-2j+1}{n}+\frac{8}{9}\frac{(j-1)^2}{n-1}\Bigg)H_n^{(2)}-\frac{8}{9}H_n^{(3)}+\\
&+\Bigg(\frac{8}{81}\frac{(2j+3)(j-3)}{j^2}-\frac{4}{27}\frac{j(4j-3)}{n}+\frac{4}{9}\frac{j(4j-3)}{n^2}-\frac{4}{27}\frac{32j+35}{j(n-j+1)}+\\
&\quad+\frac{8}{9}\frac{1}{n-j}-\frac{4}{9}\frac{1}{(n-j)^2}+\frac{4}{27}\frac{4j^2-9j+6}{n-1}-\frac{8}{9}\frac{j(2j-3)}{(n-1)^2}\Bigg)H_j-\\
&-4\frac{1}{(n-j+1)^2}-\frac{1}{27}\frac{(4j-3)(12j^3-9j^2-2j+12)}{n^2}-\frac{2}{9}\frac{19n^2-n+6}{(n+1)n(n-j+1)}+\\
&+\frac{2}{81}\frac{352nj-203n+144j^4-396j^3+220j^2+18j+45}{n(n-1)}-\frac{16}{9}\frac{(j-1)^2}{(n-1)^3}+\\
&+\frac{4}{9}\frac{j(4j-3)}{n^3}-\frac{8}{27}\frac{(j-1)(6j^3-18j^2+19j-16)}{(n-1)^2}+\frac{251}{81}-\\
&-\frac{4}{9}\sum_{k=j+1}^{n}\frac{H_{k-j}^3}{k}+\frac{8}{27}\sum_{k=j+1}^{n}\frac{H_{k-j}^2}{k}+\frac{4}{3}\sum_{k=j+1}^{n}\frac{H_{k-j}H_{k-j}^{(2)}}{k}-\frac{4}{9}\left(\sum_{k=j+1}^{n}\frac{H_{k-j}}{k}\right)^2+\\
&+\Bigg(\frac{4}{9}H_{n-j}^2+\Bigg(\frac{20}{9}+\frac{8}{9}\frac{(j-1)^2}{n}-\frac{8}{9}\frac{(j-1)(j-2)}{n-1}+\frac{8}{9}\frac{1}{n-j+1}\Bigg)H_{n-j}-\\
&\quad-\frac{4}{9}H_{n-j}^{(2)}+\frac{4}{9}H_n-\frac{4}{9}H_j+\frac{8}{81}\frac{7j+9}{j}+\frac{4}{27}\frac{(11j-10)(j-1)}{n-1}-\\
&\quad-\frac{8}{9}\frac{(j-1)(j-2)}{(n-1)^2}+\frac{8}{3}\frac{1}{n-j+1}-\frac{4}{27}\frac{11j^2-7j-1}{n}+\\
&\quad+\frac{8}{9}\frac{(j-1)^2}{n^2}\Bigg)\sum_{k=j+1}^{n}\frac{H_{k-j}}{k}-
\end{aligned}
$$

$$
\begin{aligned}
&-\frac{8}{27}\sum_{k=j+1}^{n}\frac{H_{k-j}^{(2)}}{k}-\frac{8}{9}\sum_{k=j+1}^{n}\frac{H_{k-j}^{(3)}}{k}-\frac{4}{9}\sum_{k=1}^{n-j-1}\frac{H_{n-j-k}^{2}}{k^2}+\\
&+\left(-\frac{140}{27}-\frac{8}{9}\frac{1}{n-j+1}+\frac{(j-1)^2}{n}-\frac{8}{9}\frac{(j-1)(j-2)}{n-1}\right)\sum_{k=1}^{n-j-1}\frac{H_{n-j-k}}{k^2}+\\
&+\frac{4}{9}\sum_{k=1}^{n-j-1}\frac{H_{n-j-k}^{(2)}}{k^2}+\left(\frac{8}{9}\frac{(j-1)(j-2)}{n-1}-\frac{8}{9}\frac{(j-1)^2}{n}\right)\sum_{k=1}^{n-j-1}\frac{H_kH_{n-j-k}}{k}+\\
&+\frac{8}{9}\sum_{k=1}^{n-j-1}\frac{H_kH_{n-k}H_{n-j-k}}{k}+\\
&+\left(\frac{68}{27}+\frac{8}{9}\frac{(j-1)(j-2)}{n-1}-\frac{8}{9}\frac{(j-1)^2}{n}\right)\sum_{k=1}^{n-j-1}\frac{H_kH_{n-j-k}}{n-k}+\\
&+\left(\frac{4}{9}+\frac{8}{9}\frac{1}{n-j+1}\right)\sum_{k=1}^{n-j-1}\frac{H_kH_{n-k}}{k}+\frac{140}{27}\frac{1}{n-j+1}\sum_{k=1}^{n-j-1}\frac{H_{n-k}}{k}+\\
&+\frac{140}{27}\frac{1}{n-j+1}\sum_{k=1}^{n-j-1}\frac{H_{n-k}}{n-j-k}+\\
&+\left(\frac{2}{27}\frac{17j+70}{j}-\frac{8}{9}\frac{j^2-j+1}{n-1}+\frac{8}{9}\frac{1}{j(n-j+1)}+\right.\\
&\qquad\left.+\frac{4}{9}\frac{(2j+1)(j-1)}{n}\right)\sum_{k=1}^{n-j-1}\frac{H_k}{n-k}+\\
&+\left(\frac{140}{27}+\frac{8}{9}\frac{1}{n-j+1}\right)\sum_{k=1}^{n-j-1}\frac{H_kH_{n-k}}{n-j-k}-\frac{8}{9}\sum_{k=1}^{n-j-1}\frac{H_{n-k}H_{n-j-k}}{k^2}+\\
&+\left(-\frac{4}{9}-\frac{8}{9}\frac{1}{n-j+1}\right)\sum_{k=1}^{n-j-1}\frac{H_{n-k}}{k^2}+\\
&+\left(-\frac{8}{3}+\frac{8}{9}\frac{(j-1)^2}{n-1}-\frac{4}{9}\frac{2j^2-2j+1}{n}\right)\sum_{k=1}^{j}\frac{H_{n-k}}{k}+\frac{140}{27}\sum_{k=n-j}^{n-1}\frac{H_{n-k}}{k}+\\
&+\frac{8}{9}\sum_{k=n-j}^{n-1}\frac{H_kH_{n-k}}{k}-\frac{8}{9}\sum_{k=n-j}^{n-1}\frac{H_{n-k}}{k^2}\quad \text{für}\quad 1\le j<n\,,
\end{aligned}
$$

$$\hat{A}_{n,n}=0\,. \tag{1.90}$$

1.4.5 Asymptotische Betrachtungen und weitere Folgerungen

Da man aufgrund der Komplexität der gewonnenen Ergebnisse das Verhalten der betrachteten Größen nicht mehr unmittelbar ersehen kann, ist es hilfreich, asymptotische Betrachtungen anzustellen. Insbesondere betrachtet man die Fälle j fix (also am 'linken Rand' des Baumes), $n-j$ fix (also am 'rechten Rand' des Baumes) und $j \sim \rho n$ für $0 < \rho < 1$ (also ein fixes

Verhältnis von j/n).

Um zu diesen asymptotischen Aussagen zu gelangen benötigt man zunächst einmal die asymptotischen Entwicklungen der Harmonischen Zahlen. Da diese Zahlen durch die Beziehungen

$$H_n = \Psi(n+1) + \gamma\,, \quad H_n^{(2)} = \frac{\pi^2}{6} - \Psi'(n+1) \tag{1.91}$$

mit der Psi-Funktion $\Psi(z) = \frac{d}{dz}\log(\Gamma(z))$ bzw. ihrer Ableitung $\Psi'(z) = \frac{d^2}{dz^2}\log(\Gamma(z))$ in engem Zusammenhang stehen, findet man die gesuchten asymptotischen Entwicklungen in Handbüchern über spezielle mathematische Funktionen (siehe beispielsweise [1]).

Es gilt jedenfalls

$$H_n = \log n + \gamma + \frac{1}{2n} - \frac{1}{12n^2} + \mathcal{O}\left(\frac{1}{n^4}\right)\,, \tag{1.92a}$$

$$H_n^{(2)} = \frac{\pi^2}{6} - \frac{1}{n} + \frac{1}{2n^2} + \mathcal{O}\left(\frac{1}{n^3}\right)\,. \tag{1.92b}$$

Des weiteren benötigt man noch asymptotische Aussagen über das Verhalten der bei verschiedenen Resultaten auftretenden Summe $\sum_{k=j+1}^{n} \frac{H_{k-j}}{k}$. Da diese Summe durch ein im Unterabschnitt 1.6 gezeigtes Reziprokgesetz (1.176a) mit der Summe $\sum_{k=1}^{j} \frac{H_{n-k}}{k}$ in enger Verbindung steht, können wir wahlweise auch diese Summe entwickeln, wenn sich dies als vorteilhaft herausstellt.

Hier und auch in weiterer Folge begegnen uns einerseits der Dilogarithmus $\operatorname{dilog} z$ (siehe [1]), welcher als Integral definiert ist durch

$$\operatorname{dilog} z = \int_1^z \frac{\log t}{1-t}\,dt\,, \tag{1.93}$$

bzw. die sogenannten Polylogarithmen $\mathrm{L}_m(z)$ (siehe [26]), welche für $|z| < 1$ gegeben sind durch

$$\mathrm{L}_m(z) = \sum_{n\ge 1} \frac{z^n}{n^m}\,. \tag{1.94}$$

Wichtig sind auch folgende Beziehungen

$$\mathrm{L}_2(z) = \operatorname{dilog}(1-z) \text{ für } 0 < z < 1\,, \tag{1.95a}$$

$$\operatorname{dilog} z + \operatorname{dilog}(1-z) = -\log z \log(1-z) + \frac{\pi^2}{6} \text{ für } 0 < z < 1\,, \tag{1.95b}$$

$$\operatorname{dilog} z + \operatorname{dilog}\left(\frac{1}{z}\right) = -\frac{1}{2}\log^2 z \text{ für } 0 < z < 1\,. \tag{1.95c}$$

Für ein Studium des Verhaltens oben erwähnter Summen für ein fixes Verhältnis $\rho = j/n$ empfiehlt es sich, die Summe $\sum_{k=1}^{j} \frac{H_{n-k}}{k}$ zu betrachten.

Setzen wir also $j = \rho n$ mit $0 < \rho < 1$ und entwickeln wir H_{n-k} nach obiger Gleichung (1.92a), dann erhalten wir zunächst

$$\begin{aligned}\sum_{k=1}^{\rho n} \frac{H_{n-k}}{k} &= \sum_{k=1}^{\rho n} \frac{\log(n-k) + \gamma + \frac{1}{2(n-k)} + \mathcal{O}(1/n^2)}{k} \\ &= \sum_{k=1}^{\rho n} \frac{\log(n-k)}{k} + \sum_{k=1}^{\rho n} \frac{\gamma}{k} + \sum_{k=1}^{\rho n} \frac{1}{2k(n-k)} + \mathcal{O}(1/n)\,. \end{aligned} \tag{1.96}$$

Sofort finden wir

$$\sum_{k=1}^{\rho n} \frac{\gamma}{k} = \gamma H_{\rho n} = \gamma\left(\log(\rho n) + \gamma + \mathcal{O}(1/n)\right) = \gamma \log n + \gamma \log \rho + \gamma^2 + o(1) \tag{1.97}$$

und

$$\sum_{k=1}^{\rho n} \frac{1}{2k(n-k)} = \frac{1}{2\,n} \sum_{k=1}^{\rho n} \left(\frac{1}{k} + \frac{1}{n-k}\right) = \frac{1}{2n}\left(H_{\rho n} + H_{n-1} - H_{n(1-\rho)-1}\right) = o(1)\,. \tag{1.98}$$

Erstere Summe in (1.96) entwickeln wir weiter in

$$\sum_{k=1}^{\rho n} \frac{\log(n-k)}{k} = \sum_{k=1}^{\rho n} \frac{\log n + \log(1-\frac{k}{n})}{k} = \log n \sum_{k=1}^{\rho n} \frac{1}{k} + \sum_{k=1}^{\rho n} \frac{\log(1-\frac{k}{n})}{k}\,. \tag{1.99}$$

Für den einen Ausdruck bekommen wir weiter

$$\log n \sum_{k=1}^{\rho n} \frac{1}{k} = \log n H_{\rho n} = \log n(\log(\rho n) + \gamma + \mathcal{O}(1/n)) = \log^2 n + (\log \rho + \gamma)\log n + o(1)\,, \tag{1.100}$$

während wir den anderen Ausdruck mittels

$$\sum_{k=1}^{\rho n} \frac{\log(1-\frac{k}{n})}{k} = \sum_{\frac{k}{n}=\frac{1}{n}}^{\rho} \frac{\log\left(1-\frac{k}{n}\right)}{n\cdot\frac{k}{n}} = \sum_{x=\frac{1}{n}}^{\rho} \frac{\log(1-x)}{nx} = \int_{\frac{1}{n}}^{\rho} \frac{\log(1-x)}{x}\,dx + \mathcal{O}\left(1/n\right) \tag{1.101}$$

als Riemannsche Zwischensumme des angegebenen Integrals auffassen. Der dabei auftretende Fehlerterm läßt sich durch die Summanden am Rand nach oben abschätzen, was das angegebene Zwischenresultat liefert.

Dieses Integral läßt sich durch die Variablensubstitution $t = \frac{1}{1-x}$ in Beziehung zu dem oben eingeführten Dilogarithmus bringen:

$$\begin{aligned}\int_{\frac{1}{n}}^{\rho} \frac{\log(1-x)}{x}\,dx + \mathcal{O}\left(1/n\right) &= \int_{\frac{n}{n-1}}^{\frac{1}{1-\rho}} \frac{\log t}{1-t}dt + \int_{\frac{n}{n-1}}^{\frac{1}{1-\rho}} \frac{\log t}{t}dt + o(1) \\ &= \operatorname{dilog}\left(\frac{1}{1-\rho}\right) - \operatorname{dilog}\left(\frac{n}{n-1}\right) + \frac{1}{2}\log^2\left(\frac{1}{1-\rho}\right) - \frac{1}{2}\log^2\left(\frac{n}{n-1}\right)\,. \end{aligned} \tag{1.102}$$

Unter Verwendung von (1.95) und der Entwicklung $\operatorname{dilog}\left(\frac{n}{1-n}\right) = \mathcal{O}(1/n)$ für $n \to \infty$ erhalten wir nun

$$\sum_{k=1}^{\rho n} \frac{\log(1-\frac{k}{n})}{k} = -\operatorname{dilog}(1-\rho) + o(1) = -\operatorname{L}_2(\rho) + o(1)\,. \tag{1.103}$$

Faßt man nun (1.97), (1.98), (1.100) und (1.103) zusammen, folgert man aus (1.96) schon

Hilfssatz 1.4.6. *Für $n \to \infty$ und $j = \rho n$ mit $0 < \rho < 1$ gilt*

$$\sum_{k=1}^{j} \frac{H_{n-k}}{k} = \log^2(n) + (\log\rho + 2\gamma)\log n + \gamma^2 + \gamma\log\rho - \operatorname{dilog}(1-\rho) + o(1)\,. \tag{1.104}$$

Diese Beziehung findet sich auch in [22].

Einsetzen von (1.104) in (1.176) und Entwickeln führt sofort zu

Hilfssatz 1.4.7. *Für $n \to \infty$ und $j = \rho n$ mit $0 < \rho < 1$ gilt*

$$\sum_{k=j+1}^{n} \frac{H_{k-j}}{k} = -\log\rho\log n + \mathcal{O}(1)\,. \tag{1.105}$$

Nun wollen wir das Verhalten der Summe $\sum_{k=1}^{j} \frac{H_{n-k}}{k}$ für ein fixes j und $n \to \infty$ untersuchen. Wir bekommen zunächst

$$\sum_{k=1}^{j} \frac{H_{n-k}}{k} = \sum_{k=1}^{j} \frac{\log(n-k) + \gamma + \mathcal{O}(1/n)}{k} = \sum_{k=1}^{j} \frac{\log(n-k)}{k} + \sum_{k=1}^{j} \frac{\gamma}{k} + \mathcal{O}(1/n)\,. \tag{1.106}$$

Die auftretenden Summen lassen sich leicht entwickeln, und wir erhalten für diese

$$\sum_{k=1}^{j} \frac{\log(n-k)}{k} = \sum_{k=1}^{j} \frac{\log(n) + \log(1-\frac{k}{n})}{k} = H_j \log n + \mathcal{O}(1/n) \tag{1.107}$$

und

$$\sum_{k=1}^{j} \frac{\gamma}{k} = \gamma H_j\,. \tag{1.108}$$

Zusammengefaßt liefert dies

Hilfssatz 1.4.8. *Für fixes j und $n \to \infty$ gilt*

$$\sum_{k=1}^{j} \frac{H_{n-k}}{k} = H_j \log n + \gamma H_j + \mathcal{O}(1/n)\,. \tag{1.109}$$

Einsetzen von (1.109) in (1.176a) und Entwickeln führt dann zu

Hilfssatz 1.4.9. *Für fixes j und $n \to \infty$ gilt*

$$\sum_{k=j+1}^{n} \frac{H_{k-j}}{k} = \frac{1}{2}\log^2 n + \gamma \log n - \frac{\pi^2}{12} + \frac{\gamma^2}{2} - \frac{H_j^2}{2} + \frac{H_j}{j} - \frac{H_j^{(2)}}{2} + \mathcal{O}(1/n)\,. \tag{1.110}$$

Nun wollen wir noch das Verhalten der Summe $\sum_{k=j+1}^{n} \frac{H_{k-j}}{k}$ für festes $l = n - j$ für $n \to \infty$ betrachten.

Dies liefert

$$\sum_{k=j+1}^{n} \frac{H_{k-j}}{k} = \sum_{k-j=1}^{n-j} \frac{H_{k-j}}{k-j-(n-j)+n} = \sum_{i=1}^{l} \frac{H_i}{n+i-l} = \mathcal{O}(1/n)\,. \tag{1.111}$$

Also erhalten wir

Hilfssatz 1.4.10. *Für fixes $l = n - j$ und $n \to \infty$ gilt*

$$\sum_{k=j+1}^{n} \frac{H_{k-j}}{k} = \mathcal{O}(1/n)\,. \tag{1.112}$$

Einsetzen von (1.112) in (1.176a) und Entwickeln führt unmittelbar zu

Hilfssatz 1.4.11. *Für fixes $l = n - j$ und $n \to \infty$ gilt*

$$\sum_{k=1}^{j} \frac{H_{n-k}}{k} = \log^2 n + 2\gamma \log n + \gamma^2 - \frac{\pi^2}{6} + \mathcal{O}(1/n)\,. \tag{1.113}$$

Mit den Beziehungen (1.92), (1.105), (1.110) und (1.112) erhalten wir aus Gleichung (1.63)

Korollar 1.4.12. *Der Erwartungswert $A_{n,j}$ der Höhe des Knotens mit Schlüssel j in einem p-tree der Größe n, also einem Objekt der Größe n aus der Familie $\mathcal{A}$, ist asymptotisch für $j = \rho n$ und $0 < \rho < 1$ gegeben durch*

$$\boxed{A_{n,j} = \frac{1}{3}\log^2 n + \left(-\frac{2}{3}\rho^2 + \frac{2}{3}\rho + \frac{2}{3}\gamma + \frac{5}{3} + \frac{2}{3}\log\rho + \frac{2}{3}\log(1-\rho)\right)\log n + \mathcal{O}(1)\,.} \tag{1.114}$$

Korollar 1.4.13. *Der Erwartungswert $A_{n,j}$ der Höhe des Knotens mit Schlüssel j in einem p-tree der Größe n, also einem Objekt der Größe n aus der Familie $\mathcal{A}$, ist asymptotisch für fixes j gegeben durch*

$$\boxed{A_{n,j} = 2\log n + \mathcal{O}(1)\,.} \tag{1.115}$$

Korollar 1.4.14. *Der Erwartungswert $A_{n,j}$ der Höhe des Knotens mit Schlüssel j in einem p-tree der Größe n, also einem Objekt der Größe n aus der Familie $\mathcal{A}$, ist asymptotisch für fixes $l = n - j$ gegeben durch*

$$\boxed{A_{n,j} = \frac{1}{3}H_{l+1}^2 + \frac{5}{3}H_{l+1} - \frac{1}{3}H_{l+1}^{(2)} - 1 + \frac{1}{3(l+1)} + \mathcal{O}(1/n)\,.} \tag{1.116}$$

Von besonderem Interesse sind im Falle der Erwartungswerte $A_{n,j}$ der betrachteten Größen in diesem und den darauffolgenden Unterabschnitten auch die arithmetischen Mittel

$$A_n = \frac{1}{n} \sum_{j=1(0)}^{n} A_{n,j} .$$

Dieser Mittelwert ist dann der Erwartungswert der betrachteten Größe für einen zufällig ausgewählten Knoten.

Die zum Erhalt dieser Größen A_n auftretenden Summen können großteils durch die Beziehungen (1.85) vereinfacht werden. Zusätzlich benötigt man nur noch

$$\sum_{j=1}^{n} \sum_{k=j+1}^{n} \frac{H_{k-j}}{k} = \sum_{k=1}^{n} \frac{1}{k} \sum_{j=1}^{k-1} H_{k-j} = \sum_{k=1}^{n} \frac{1}{k}(kH_k - k) = \sum_{k=1}^{n} H_k - n = (n+1)H_n - 2n , \tag{1.117a}$$

$$\begin{aligned} \sum_{j=1}^{n} \sum_{k=1}^{j} \frac{H_{n-k}}{k} &= \sum_{k=1}^{n} \frac{H_{n-k}}{k} \sum_{j=k}^{n} 1 = \sum_{k=1}^{n} \frac{H_{n-k}}{k}(n-k+1) = (n+1) \sum_{k=1}^{n} \frac{H_{n-k}}{k} - \sum_{k=1}^{n} H_{n-k} \\ &= (n+1)(H_n^2 - H_n^{(2)}) - nH_n + n . \end{aligned} \tag{1.117b}$$

Damit erhält man aus (1.63) durch Aufsummieren

Korollar 1.4.15. *Der Erwartungswert A_n der Höhe eines zufällig gewählten Knotens in einem p-tree der Größe n, also einem Objekt der Größe n aus der Familie $\mathcal{A}$, ist gegeben durch*

$$\boxed{\begin{aligned} &A_n = \frac{n+1}{3n} H_n^2 + \frac{2(2n+5)}{9n} H_n - \frac{n+1}{3n} H_n^{(2)} - \frac{13n+1}{27n} \quad \text{für } n \geq 2 , \\ &A_1 = 1 , \end{aligned}} \tag{1.118}$$

bzw. asymptotisch durch

$$\boxed{A_n = \frac{1}{3} \log^2 n + \left(\frac{2}{3}\gamma + \frac{4}{9}\right) \log n + \mathcal{O}(1) \ \text{ für } n \to \infty .} \tag{1.119}$$

Obwohl man diese asymptotischen Betrachtungen auch für die exakt berechnete Varianz durchführen könnte, werden sie hier nicht ausgeführt, da dies ein sehr aufwendiges Unterfangen darstellt und keine wesentlichen zusätzlichen Einsichten bringt.

Von Interesse ist allerdings die sogenannte *Varianz der Mittelwerte*, also die Varianz $\hat{A}_n$ der Höhe eines zufällig gewählten Knotens in einem p-tree der Größe n. Um diesen Parameter zu erhalten, definieren wir die Hilfsgröße $\breve{A}_n$, welche das zweite Faktorielle Moment der Höhe eines zufällig gewählten Knotens bezeichnen soll.

Aus der durch Gleichung (1.84) gegebenen Funktion $\breve{A}(z,u)$ erhalten wir $\breve{A}_n$ durch die Beziehung

$$\breve{A}_n = \frac{1}{n} [z^n] \lim_{u \to 1} \breve{A}(z,u) . \tag{1.120}$$

Es ergibt sich aus (1.84) zunächst

$$\begin{aligned}\lim_{u\to 1}\breve{A}(z,u) &= \frac{1}{9\,(1-z)^2}\log^4\left(\frac{1}{1-z}\right)+\frac{70}{81\,(1-z)^2}\log^3\left(\frac{1}{1-z}\right)\\ &+\frac{2z^3-6z^2+6z+118}{81\,(1-z)^2}\log^2\left(\frac{1}{1-z}\right)+\frac{10z^3-30z^2+30z+230}{243\,(1-z)^2}\log\left(\frac{1}{1-z}\right)\\ &+\frac{13z^3-39z^2-690z}{729\,(1-z)^2}\,,\end{aligned} \tag{1.121}$$

woraus man mit den Beziehungen (1.19) und (1.120)

$$\begin{aligned}\breve{A}_n &= \frac{1}{n}\Big[-\frac{1112}{729}n+\frac{13}{729}-\frac{4}{81\,(n-1)^2}-\frac{2}{243\,(n-1)}+\frac{4}{81\,n^2}+\frac{2}{243\,n}+\frac{1}{9}(n+1)H_n^4\\ &+\frac{1}{81}(34n+70)H_n^3+\frac{1}{27}(6n+40)H_n^2+\frac{4}{81}\left(11n+20+\frac{1}{n-1}-\frac{1}{n}\right)H_n\\ &+\frac{1}{3}(n+1)\left(H_n^{(2)}\right)^2-\frac{2}{27}(3n+20)H_n^{(2)}+\frac{4}{81}(17n+35)H_n^{(3)}-\frac{2}{3}(n+1)H_n^{(4)}\\ &-\frac{2}{3}(n+1)H_n^2H_n^{(2)}+\frac{8}{9}(n+1)H_nH_n^{(3)}-\frac{2}{27}(17n+35)H_nH_n^{(2)}\Big]\quad\text{für } n\geq 2\,,\\ \breve{A}_1 &= 0\end{aligned} \tag{1.122}$$

erhält.

Die gesuchte Varianz $\hat{A}_n$ bekommt man nun leicht mittels

$$\hat{A}_n=\breve{A}_n+A_n-A_n^2\,, \tag{1.123}$$

wobei A_n den durch (1.118) gegebenen Erwartungswert der Höhe eines zufällig gewählten Knotens bezeichnet.

Also erhält man

Korollar 1.4.16. *Die Varianz $\hat{A}_n$ der Höhe eines zufällig gewählten Knotens in einem p-tree der Größe n, also einem Objekt der Größe n aus der Familie $\mathcal{A}$, ist gegeben durch*

$$\boxed{\begin{aligned}\hat{A}_n &= \frac{1}{n^2}\Big[-\frac{1}{9}(n+1)H_n^4+\frac{2}{81}(5n^2-7n-30)H_n^3+\frac{1}{81}(55n^2+95n-98)H_n^2\\ &+\frac{2}{243}\left(172n^2+389n+10+\frac{6}{n-1}\right)H_n+\frac{1}{9}(2n^2+n-1)\left(H_n^{(2)}\right)^2\\ &-\frac{1}{81}(71n^2+175n+2)H_n^{(2)}+\frac{4}{81}(17n^2+35n)H_n^{(3)}-\frac{2}{3}(n^2+n)H_n^{(4)}\\ &+\frac{2}{9}(-2n^2-n+1)H_n^2H_n^{(2)}+\frac{2}{27}(-13n^2-21n+10)H_nH_n^{(2)}+\frac{8}{9}(n^2+n)H_nH_n^{(3)}\\ &-\frac{544}{243}n^2-\frac{40}{729}n-\frac{1}{729}-\frac{4}{81\,(n-1)^2}-\frac{14}{243}\frac{14}{243\,(n-1)}+\frac{4}{81\,n}\Big]\quad\text{für } n\geq 2\,,\\ \hat{A}_1 &= 0\,,\end{aligned}} \tag{1.124}$$

bzw. asymptotisch durch

$$\boxed{\begin{aligned}\hat{A}_n = \frac{10}{81}\log^3 n + \left(\frac{10}{27}\gamma + \frac{55}{81} - \frac{2}{27}\pi^2\right)\log^2 n \\ + \left(\frac{10}{27}\gamma^2 + \frac{110}{81}\gamma - \frac{4}{27}\pi^2\gamma - \frac{13}{81}\pi^2 + \frac{8}{9}\zeta(3) + \frac{344}{243}\right)\log n + \mathcal{O}(1) \quad \text{für } n \to \infty\,.\end{aligned}} \tag{1.125}$$

1.5 Die Anzahl der linksseitigen Kanten auf dem Weg zu einem Knoten

1.5.1 Die Erwartungswerte

Im Gegensatz zum vorangegangenen Kapitel betrachten wir nun die auf dem direkten Weg von der Wurzel zu einem Knoten führenden Kanten, zählen hier aber nur jene, die nach links führen. So bekommen wir einen Eindruck, wie 'schief' der Baum eigentlich ist. Diese Linkslastigkeit ist natürlich gewollt, denn die Bedeutung der p-trees als Datenstruktur für Prioritätswarteschlangen liegt ja gerade in seiner Funktion als Mittelding zwischen der linearen Liste auf der einen Seite und dem Binären Suchbaum auf der anderen Seite.

Bezeichnen wir nun die Erwartungswerte der Anzahl der nach links führenden Kanten auf dem Weg von der Wurzel zum Knoten mit Schlüssel j in Objekten der Größe n aus den Familien $\mathcal{A}$, $\mathcal{B}$ und $\mathcal{C}$ mit $A^{[L]}_{n,j}$, $B^{[L]}_{n,j}$ und $C^{[L]}_{n,j}$.

Für die Erstellung einer Rekursionsformel obiger Größen ist es wieder nützlich, den Erwartungswert für die Länge des linken Pfades in Objekten der Familie $\mathcal{C}$ zu kennen. Wie in 1.3.2 errechnet, ergibt sich als Erwartungswert dafür in einem Objekt der Größe n der Wert H_n.

Nun können wir daran gehen, Rekursionsformeln für die gesuchten Erwartungswerte zu ermitteln. Teilen wir die Familien $\mathcal{A}$, $\mathcal{B}$ und $\mathcal{C}$ aufgrund des ersten Elementes einer Zufallspermutation wie in den Abbildungen 1.2, 1.3 und 1.4 gezeigt auf, so erhält man unter Angabe der Anfangswerte folgende Rekursionen:

$$A^{[L]}_{n,j} = \frac{1}{n}\left(\sum_{k=1}^{j-1} C^{[L]}_{n-k,j-k} + H_{n-j} + \sum_{k=j+1}^{n}\left(B^{[L]}_{k-1,j} + H_{n-k}\right)\right) \quad \text{für } n \ge j \ge 1\,, \tag{1.126a}$$

$$B^{[L]}_{n,j} = \frac{1}{n}\left(\sum_{k=1}^{j-1} A^{[L]}_{n-k,j-k} + 1 + \sum_{k=j+1}^{n}\left(B^{[L]}_{k-1,j} + 1\right)\right) \quad \text{für } n \ge j \ge 1\,, \tag{1.126b}$$

$$C^{[L]}_{n,j} = \frac{1}{n}\left(\sum_{k=1}^{j-1} C^{[L]}_{n-k,j-k} + H_{n-j} + \sum_{k=j+1}^{n}\left(A^{[L]}_{k-1,j} + H_{n-k}\right)\right) \quad \text{für } n \ge j \ge 1\,, \tag{1.126c}$$

$$A^{[L]}_{0,0} = 0\,, \quad B^{[L]}_{0,0} = 0\,, \quad C^{[L]}_{0,0} = 0\,.$$

Führt man durch

$$A^{[L]}(z,u)=\sum_{j\geq 1}\sum_{n\geq j}A^{[L]}_{n,j}z^n u^j\,,\quad B^{[L]}(z,u)=\sum_{j\geq 1}\sum_{n\geq j}B^{[L]}_{n,j}z^n u^j\,,$$
$$C^{[L]}(z,u)=\sum_{j\geq 1}\sum_{n\geq j}C^{[L]}_{n,j}z^n u^j\,.$$

doppelte erzeugende Funktionen in z und u ein, so übersetzen sich obige Rekursionsformeln (1.126) nach Multiplikation mit $nz^{n-1}u^j$ und Aufsummieren über alle $n \geq j \geq 1$ in folgendes System von Differentialgleichungen:

$$\frac{\partial}{\partial z}A^{[L]}(z,u)=\frac{u}{1-uz}C^{[L]}(z,u)+\frac{u}{(1-z)(1-uz)}\log\left(\frac{1}{1-z}\right)$$
$$+\frac{uz}{(1-z)^2(1-uz)}\log\left(\frac{1}{1-z}\right)+\frac{1}{1-z}B^{[L]}(z,u)\,, \tag{1.127a}$$
$$\frac{\partial}{\partial z}B^{[L]}(z,u)=\frac{u}{1-uz}A^{[L]}(z,u)+\frac{u}{(1-z)(1-uz)}+\frac{uz}{(1-z)^2(1-uz)}+\frac{1}{1-z}B^{[L]}(z,u)\,, \tag{1.127b}$$
$$\frac{\partial}{\partial z}C^{[L]}(z,u)=\frac{u}{1-uz}C^{[L]}(z,u)+\frac{u}{(1-z)(1-uz)}\log\left(\frac{1}{1-z}\right)$$
$$+\frac{uz}{(1-z)^2(1-uz)}\log\left(\frac{1}{1-z}\right)+\frac{1}{1-z}A^{[L]}(z,u)\,. \tag{1.127c}$$

Mit den Abkürzungen

$$f_1(z,u)=\frac{u}{(1-z)(1-uz)}\log\left(\frac{1}{1-z}\right)+\frac{uz}{(1-z)^2(1-uz)}\log\left(\frac{1}{1-z}\right)\,,$$
$$f_2(z,u)=\frac{u}{(1-z)(1-uz)}+\frac{uz}{(1-z)^2(1-uz)}\,,$$
$$f_3(z,u)=\frac{u}{(1-z)(1-uz)}\log\left(\frac{1}{1-z}\right)+\frac{uz}{(1-z)^2(1-uz)}\log\left(\frac{1}{1-z}\right)$$

läßt sich das Differentialgleichungssystem (1.127) schreiben als

$$\frac{\partial}{\partial z}A^{[L]}(z,u)=\frac{u}{1-uz}C^{[L]}(z,u)+\frac{1}{1-z}B^{[L]}(z,u)+f_1(z,u)\,, \tag{1.128a}$$
$$\frac{\partial}{\partial z}B^{[L]}(z,u)=\frac{u}{1-uz}A^{[L]}(z,u)+\frac{1}{1-z}B^{[L]}(z,u)+f_2(z,u)\,, \tag{1.128b}$$
$$\frac{\partial}{\partial z}C^{[L]}(z,u)=\frac{u}{1-uz}C^{[L]}(z,u)+\frac{1}{1-z}A^{[L]}(z,u)+f_3(z,u)\,. \tag{1.128c}$$

Dieses System von Differentialgleichungen erlaubt ein analoges Vorgehen zum Auffinden einer Lösung, wie dies für (1.34) durchgeführt wurde. Man erhält so als lineare Differentialgleichung dritter Ordnung für $A^{[L]}(z,u)$:

$$\begin{aligned}
&\frac{2uz-u-1}{(1-z)(1-u)}A^{[L]}(z,u)-2\frac{1-uz}{1-u}\frac{\partial}{\partial z}A^{[L]}(z,u)\\
&-\frac{3}{2}\frac{(1-uz)(2uz-u-1)}{u(1-u)}\frac{\partial^2}{\partial z^2}A^{[L]}(z,u)-\frac{1}{2}\frac{(1-uz)^2(1-z)}{u(1-u)}\frac{\partial^3}{\partial z^3}A^{[L]}(z,u)\\
&=f_3(z,u)+\frac{1}{1-uz}\frac{\partial}{\partial z}\left(\frac{(1-uz)^3}{2u(u-1)}f_2(z,u)+\right.\\
&\quad\left.+\frac{(1-uz)^3}{2u(u-1)(1-z)}\frac{\partial}{\partial z}\left((1-z)^2f_1(z,u)\right)+\frac{(1-uz)^2(1-z)}{2(u-1)}f_3(z,u)\right).
\end{aligned}\tag{1.129}$$

Nach dem Einsetzen für $f_1(z,u)$, $f_2(z,u)$ und $f_3(z,u)$ in diese Gleichung erhält man unter Berücksichtigung der Anfangswerte $A^{[L]}_{0,0}=0$, $A^{[L]}_{1,1}=0$, $A^{[L]}_{2,1}=1$ und $A^{[L]}_{2,2}=0$ nach Vereinfachungen schließlich

$$\begin{aligned}
&\frac{2uz-u-1}{(1-z)(1-u)}A^{[L]}(z,u)-2\frac{1-uz}{1-u}\frac{\partial}{\partial z}A^{[L]}(z,u)\\
&-\frac{3}{2}\frac{(1-uz)(2uz-u-1)}{u(1-u)}\frac{\partial^2}{\partial z^2}A^{[L]}(z,u)-\frac{1}{2}\frac{(1-uz)^2(1-z)}{u(1-u)}\frac{\partial^3}{\partial z^3}A^{[L]}(z,u)\\
&=\frac{uz+u-2}{(1-z)^3(1-u)},\\
&A^{[L]}(0,u)=0,\quad \frac{\partial}{\partial z}A^{[L]}(z,u)\Big|_{z=0}=0,\quad \frac{\partial^2}{\partial z^2}A^{[L]}(z,u)\Big|_{z=0}=2u.
\end{aligned}\tag{1.130}$$

Mit den wie in Unterabschnitt 1.3.2 durchgeführten Transformationen

$$\begin{aligned}
A^{[L]}(z,u)&=\frac{E^{[L]}(z,u)}{(1-z)(1-uz)},\\
F^{[L]}(z,u)&=\frac{\partial}{\partial z}E^{[L]}(z,u),\\
F^{[L]}(z,u)&=(1-z)(1-uz)G^{[L]}(z,u),\\
H^{[L]}(z,u)&=\frac{\partial}{\partial z}G^{[L]}(z,u),
\end{aligned}$$

ergibt sich daraus die lineare Differentialgleichung erster Ordnung

$$\frac{\partial}{\partial z}H^{[L]}(z,u)-2\frac{1+u-2uz}{(1-z)(1-uz)}H^{[L]}(z,u)=-2\frac{(uz+u-2)u}{(1-z)^2}.\tag{1.131}$$

Mittels Variation der Konstanten findet man eine Partikulärlösung von (1.131) durch

$$H_p^{[L]}(z,u)=2\frac{u^2}{(1-z)^2(1-uz)^2}\log\left(\frac{1}{1-z}\right)-4\frac{u(u-1)}{(1-z)^3(1-uz)^2}.\tag{1.132}$$

Substituiert man nun wieder $H^{[L]}(z,u)$ durch $A^{[L]}(z,u)$, so erhält man als Partikulärlösung für (1.130)

$$A_p^{[L]}(z,u)=\frac{1}{(1-z)(1-uz)}\int_0^z\left[(1-x)(1-ux)\int_0^x H_p^{[L]}(t,u)\,dt\right]dx.\tag{1.133}$$

Durch partielle Integration kann man diesen Ausdruck für ein einfacheres Ablesen der Koeffizienten weiter abwandeln.

Hierzu definiert man

$$g(z,u) = \int_0^z (1-x)(1-ux)\,dx$$

und findet so die Partikulärlösung gegeben als

$$A_p^{[L]}(z,u) = \frac{g(z,u)}{(1-z)(1-uz)} \int_0^z H_p(t,u)\,dt - \frac{1}{(1-z)(1-uz)} \int_0^z g(t,u) H_p^{[L]}(t,u)\,dt\,. \tag{1.134}$$

Führt man diese Integrationen aus, ergibt sich

$$\begin{aligned} A_p^{[L]}(z,u) = &\left(\frac{2}{3}\frac{(2uz-u-1)u^3}{(1-u)^3} + \frac{2}{3}\frac{u^2}{(1-z)(1-u)}\right) \int_0^z \frac{1}{1-ut} \log\left(\frac{1}{1-t}\right) dt \\ &+ \left(-\frac{1}{3}\frac{(2uz-u-1)u^2}{(1-u)^3} - \frac{1}{3}\frac{u^2}{(1-uz)(1-u)}\right) \log^2\left(\frac{1}{1-z}\right) \\ &+ \left(\frac{1}{3}\frac{(10uz-u-9)u^2}{(1-u)^3} - \frac{1}{3}\frac{u^2}{(1-uz)(1-u)} + 2\frac{u}{(1-z)(1-u)}\right) \log\left(\frac{1}{1-z}\right) \\ &+ \left(-\frac{5}{3}\frac{(2uz-u-1)u^2}{(1-u)^3} - \frac{5}{3}\frac{u}{(1-z)(1-u)}\right) \log\left(\frac{1}{1-uz}\right) \\ &+ \frac{14}{3}\frac{u^2}{(1-u)^2} + 2\frac{u^2}{(1-uz)(1-u)} - \frac{5}{3}\frac{u}{(1-z)(1-u)} \end{aligned} \tag{1.135}$$

und nach Umformungen, die für ein leichteres Ablesen der Koeffizienten dienlich sind, schließlich

$$\begin{aligned} A_p^{[L]}(z,u) = &\left(\frac{1}{3}u(2uz-u-1) + \frac{1}{3}\frac{(u-1)^2}{1-z}\right) \int_0^z \frac{1}{(1-t)^2(1-ut)^2} \log\left(\frac{1}{1-t}\right) dt \\ &+ \left(-\frac{1}{3}\frac{u^2}{(1-uz)(1-u)} + \frac{1}{3}\frac{u}{(1-z)(1-u)}\right) \log^2\left(\frac{1}{1-z}\right) \end{aligned} \tag{1.136}$$

$$\begin{aligned} &+ \left[2\frac{(2uz-u-1)u^2}{(1-u)^3} + \frac{7}{3}\frac{u}{(1-z)(1-u)} - \frac{1}{3}\frac{1}{(1-z)^2} - \right. \\ &\qquad \left. - \frac{1}{3}\frac{u^2}{(1-uz)(1-u)}\right] \log\left(\frac{1}{1-z}\right) \\ &+ \left(-2\frac{(2uz-u-1)u^2}{(1-u)^3} - 2\frac{u}{(1-z)(1-u)}\right) \log\left(\frac{1}{1-uz}\right) \\ &+ 4\frac{u^2}{(1-u)^2} - 2\frac{u}{(1-z)(1-u)} + \frac{1}{3}\frac{1}{(1-z)^2} + 2\frac{u^2}{(1-uz)(1-u)}\,. \end{aligned} \tag{1.137}$$

Die der Differentialgleichung (1.130) entsprechende homogene Differentialgleichung ist die gleiche wie die homogene Differentialgleichung von (1.34). Als Lösung für diese wurde dort berechnet

$$\begin{aligned} A_h^{[L]}(z,u) = &\frac{c_1(u)}{(1-z)(1-uz)} + c_2(u)\frac{z(6-3uz-3z+2uz^2)}{(1-z)(1-uz)} \\ &+ \frac{c_3(u)}{(1-z)(1-uz)} \int_0^z \left[(1-x)(1-ux) \int_0^x \frac{1}{(1-t)^2(1-ut)^2}\,dt\right] dx\,, \end{aligned} \tag{1.138}$$

mit drei nur von u abhängigen Funktionen $c_1(u)$, $c_2(u)$ und $c_3(u)$.

Durch Anpassen der Anfangswerte erhalten wir nun für $A^{[L]}(z,u)$ die Gleichungen

$$\begin{aligned}
&A^{[L]}(z,u) = A_p^{[L]}(z,u) + A_h^{[L]}(z,u)\,,\\
&A^{[L]}(0,u) = 0\,, \quad \frac{\partial}{\partial z}A^{[L]}(z,u)\Big|_{z=0} = 0\,, \quad \frac{\partial^2}{\partial z^2}A^{[L]}(z,u)\Big|_{z=0} = 2u\,,
\end{aligned}$$

woraus sich die Funktionen $c_1(u)$, $c_2(u)$ und $c_3(u)$ ergeben als

$$c_1(u) = \frac{2}{9}\frac{4u^3-31u^2+5u+1}{(1-u)^2}\,, \quad c_2(u) = \frac{4}{3}\frac{u^2}{(1-u)^2} \text{ und } c_3(u) = \frac{1}{9}\frac{1-2u}{(1-u)^3}\,.$$

Faßt man diese Resultate zusammen, erhält man für die bivariate erzeugende Funktion $A^{[L]}(z,u)$ für die Erwartungswerte $A_{n,j}^{[L]}$ der Anzahl der in einem p-tree der Größe n auf dem Weg von der Wurzel zum Knoten j nach links führenden Kanten

$$\begin{aligned}
A^{[L]}(z,u) = &\left(\frac{1}{3}u(2uz-u-1) + \frac{1}{3}\frac{(u-1)^2}{1-z}\right)\int_0^z \frac{1}{(1-t)^2(1-ut)^2}\log\left(\frac{1}{1-t}\right)dt\\
&+\left(-\frac{1}{3}\frac{u^2}{(1-uz)(1-u)} + \frac{1}{3}\frac{u}{(1-z)(1-u)}\right)\log^2\left(\frac{1}{1-z}\right)\\
&+\Bigg[\frac{2}{3}\frac{(u+1)(2uz-u-1)u}{(1-u)^3} + \frac{7}{3}\frac{u}{(1-z)(1-u)} - \frac{1}{3}\frac{1}{(1-z)^2} -\\
&\qquad - \frac{1}{3}\frac{(5u-2)u}{(1-uz)(1-u)}\Bigg]\log\left(\frac{1}{1-z}\right)\\
&+\left(-\frac{2}{3}\frac{(u+1)(2uz-u-1)u}{(1-u)^3} - \frac{2}{3}\frac{u+1}{(1-z)(1-u)}\right)\log\left(\frac{1}{1-uz}\right)\\
&+\frac{8}{3}\frac{u^2z}{(1-u)^2} - \frac{1}{3}\frac{4u+1}{(1-z)(1-u)} + \frac{1}{3}\frac{1}{(1-z)^2} + \frac{5}{3}\frac{u}{(1-uz)(1-u)}\,.
\end{aligned} \tag{1.139}$$

Für das Ablesen der Koeffizienten von $A^{[L]}(z,u)$ kann man wieder die schon im vorigen Abschnitt verwendeten Beziehungen (1.19) und (1.62) verwenden und gewinnt daraus

Satz 1.5.1. *Der Erwartungswert $A_{n,j}^{[L]}$ der Anzahl der nach links führenden Kanten auf dem Weg von der Wurzel zum Knoten mit Schlüssel j in einem p-tree der Größe n, das heißt also in einem Objekt aus der Familie $\mathcal{A}$, ist gegeben durch*

$$\boxed{\begin{aligned}
A_{n,j}^{[L]} = &-\frac{2}{3}\sum_{k=j+1}^{n}\frac{H_{k-j}}{k} + \frac{1}{3}H_{n-j}^2 + H_n\\
&+\left(1 + \frac{2}{3}\frac{(j-1)^2}{n} - \frac{2}{3}\frac{(j-1)(j-2)}{n-1} + \frac{2}{3}\frac{1}{n-j+1}\right)H_{n-j}\\
&- H_j - \frac{1}{3}H_{n-j}^{(2)} - \frac{1}{3}\frac{(3j-1)(2j-1)}{n} + 2\frac{(j-1)^2}{n-1}\\
&+\frac{5}{3}\frac{1}{n-j+1} + \frac{1}{3j} - \frac{4}{3} \quad \text{für } 1 \le j \le n\,.
\end{aligned}} \tag{1.140}$$

Um zu den bivariaten erzeugenden Funktionen $C^{[L]}(z,u)$ für die Erwartungswerte $C^{[L]}_{n,j}$ der Anzahl der in einem Objekt der Größe n aus der Familie $\mathcal{C}$ auf dem Weg von der Wurzel zum Knoten j nach links führenden Kanten zu gelangen, können wir das eben errechnete $A^{[L]}(z,u)$ in Gleichung (1.39) einsetzen und bekommen so mit Beziehung (1.35)

$$
\begin{aligned}
C^{[L]}(z,u) = &\left(-\frac{1}{3}u(uz+u-2)+\frac{1}{3}\frac{(u-1)^2}{1-z}\right)\int_0^z \frac{1}{(1-t)^2(1-ut)^2}\log\left(\frac{1}{1-t}\right)dt \\
&+\left(-\frac{1}{3}\frac{u^2}{(1-uz)(1-u)}+\frac{1}{3}\frac{u}{(1-z)(1-u)}\right)\log^2\left(\frac{1}{1-z}\right) \\
&+\left[-\frac{2}{3}\frac{(u+1)(uz+u-2)u}{(1-u)^3}-\frac{1}{3}\frac{1}{(1-z)^2}+\frac{7}{3}\frac{u}{(1-z)(1-u)}\right. \\
&\qquad\left.-\frac{1}{3}\frac{(5u-2)u}{(1-uz)(1-u)}\right]\log\left(\frac{1}{1-z}\right)+ \\
&+\left(\frac{2}{3}\frac{(u+1)(uz+u-2)u}{(1-u)^3}-\frac{2}{3}\frac{u+1}{(1-z)(1-u)}\right)\log\left(\frac{1}{1-uz}\right) \\
&-\frac{2}{3}\frac{(2uz+3u-3)u}{(1-u)^2}+\frac{5}{3}\frac{u}{(1-uz)(1-u)}-\frac{1}{3}\frac{10u+1}{(1-z)(1-u)}+\frac{1}{3}\frac{1}{(1-z)^2}\,.
\end{aligned}
\tag{1.141}
$$

Das Ablesen der Koeffizienten gelingt wiederum leicht unter Benützung der Beziehungen (1.62) und man folgert

Hilfssatz 1.5.2. *Der Erwartungswert $C^{[L]}_{n,j}$ der Anzahl der nach links führenden Kanten auf dem Weg von der Wurzel zum Knoten mit Schlüssel j in einem Objekt aus der Familie $\mathcal{C}$ ist gegeben durch*

$$
\boxed{
\begin{aligned}
C^{[L]}_{n,j} = &-\frac{2}{3}\sum_{k=j+1}^{n}\frac{H_{k-j}}{k}+\frac{1}{3}H^2_{n-j}+H_n \\
&+\left(2-\frac{1}{3}\frac{(j+2)(j-1)}{n}+\frac{1}{3}\frac{(j-1)(j-2)}{n-1}+\frac{2}{3}\frac{1}{n-j+1}\right)H_{n-j} \\
&-H_j-\frac{1}{3}H^{(2)}_{n-j}+\frac{1}{3}\frac{3j^2+5j-1}{n}-\frac{(j-1)^2}{n-1} \\
&+\frac{5}{3}\frac{1}{n-j+1}+\frac{1}{3j}-\frac{13}{3}\quad \text{für } 1\le j\le n\,.
\end{aligned}
}
\tag{1.142}
$$

Um nun schließlich auch $B^{[L]}(z,u)$ zu erhalten, setzt man die beiden eben erhaltenen Resultate (1.139) und (1.141) in Gleichung (1.37) ein und wir erhalten mit Beziehung (1.35)

$$
\begin{aligned}
B^{[L]}(z,u) = &\left(-\frac{1}{3}u(uz-2u+1)+\frac{1}{3}\frac{(u-1)^2}{1-z}\right)\int_0^z \frac{1}{(1-t)^2(1-ut)^2}\log\left(\frac{1}{1-t}\right)dt \\
&+\left(-\frac{1}{3}\frac{u^2}{(1-uz)(1-u)}+\frac{1}{3}\frac{u}{(1-z)(1-u)}\right)\log^2\left(\frac{1}{1-z}\right) \\
&+\left[-\frac{2}{3}\frac{(u+1)(uz-2u+1)u}{(1-u)^3}+\frac{4}{3}\frac{u}{(1-z)(1-u)}-\frac{1}{3}\frac{1}{(1-z)^2}+\right. \\
&\qquad\left.+\frac{2}{3}\frac{u}{1-uz}\right]\log\left(\frac{1}{1-z}\right)
\end{aligned}
$$

$$+\left(\frac{2}{3}\frac{(u+1)(uz-2u+1)u}{(1-u)^3}-\frac{2}{3}\frac{u+1}{(1-z)(1-u)}\right)\log\left(\frac{1}{1-uz}\right)$$
$$-\frac{2}{3}\frac{(2uz-3u+3)u}{(1-u)^2}+\frac{5}{3}\frac{u}{(1-uz)(1-u)}+\frac{1}{3}\frac{2u-1}{(1-z)(1-u)}+\frac{1}{3}\frac{1}{(1-z)^2}. \tag{1.143}$$

Liest man nun aus obigem Ausdruck unter Zuhilfenahme von (1.62) die Koeffizienten ab, ergibt dies

Hilfssatz 1.5.3. *Der Erwartungswert $B^{[L]}_{n,j}$ der Anzahl der nach links führenden Kanten auf dem Weg von der Wurzel zum Knoten mit Schlüssel j in einem Objekt aus der Familie $\mathcal{B}$ ist gegeben durch*

$$\boxed{\begin{aligned}B^{[L]}_{n,j} &= -\frac{2}{3}\sum_{k=j+1}^{n}\frac{H_{k-j}}{k}+\frac{1}{3}H^2_{n-j}+H_n\\ &+\left(-\frac{1}{3}\frac{(j-1)(j-4)}{n}+\frac{1}{3}\frac{(j-1)(j-2)}{n-1}+\frac{2}{3}\frac{1}{n-j+1}\right)H_{n-j}\\ &-H_j-\frac{1}{3}H^{(2)}_{n-j}+\frac{1}{3}\frac{3j^2-10j+2}{n}-\frac{(j-1)^2}{n-1}\\ &+\frac{5}{3}\frac{1}{n-j+1}+\frac{1}{3j}+\frac{2}{3}\quad \text{für } 1\le j\le n\,.\end{aligned}} \tag{1.144}$$

1.5.2 Die Erwartungswerte der Anzahl an rechtsseitigen Kanten

Nun berechnen wir zum Vergleich die durchschnittliche Anzahl an rechtsseitigen Kanten die zu einem speziellen Knoten führen. Bezeichnen wir also hier die Erwartungswerte der Anzahl der rechten Kanten auf dem Weg von der Wurzel zum Knoten mit Schlüssel j in Objekten der Größe n aus den Familien $\mathcal{A}$, $\mathcal{B}$ und $\mathcal{C}$ mit $A^{[R]}_{n,j}$, $B^{[R]}_{n,j}$ und $C^{[R]}_{n,j}$.

Für die Berechnung dieser Größen betrachten wir zunächst die Erwartungswerte $A^{[E]}_{n,j}$, $B^{[E]}_{n,j}$ und $C^{[E]}_{n,j}$ der Gesamtanzahl an Kanten auf dem Weg von der Wurzel zum Knoten mit Schlüssel j in Objekten der Größe n aus den Familien $\mathcal{A}$, $\mathcal{B}$ und $\mathcal{C}$. Da ja für die in 1.4.2 berechneten Erwartungswerte $A_{n,j}$, $B_{n,j}$ und $C_{n,j}$ der Anzahl an Knoten auf dem Weg von der Wurzel zu einem speziellen Knoten die Beziehungen

$$A_{n,j}=A^{[E]}_{n,j}+1\ ,\quad B_{n,j}=B^{[E]}_{n,j}+1 \text{ und } \quad C_{n,j}=C^{[E]}_{n,j}+1$$

gelten, gewinnt man

$$\begin{aligned}A^{[E]}_{n,j} &= -\frac{2}{3}\sum_{k=j+1}^{n}\frac{H_{k-j}}{k}+\frac{1}{3}H^2_{n-j}\\ &+\left(\frac{5}{3}+\frac{2}{3}\frac{(j-1)^2}{n}-\frac{2}{3}\frac{(j-1)(j-2)}{n-1}+\frac{2}{3}\frac{1}{n-j+1}\right)H_{n-j}+\frac{1}{3}H_n-\frac{1}{3}H_j\\ &-\frac{1}{3}H^{(2)}_{n-j}-\frac{1}{3}\frac{j(4j-3)}{n}+\frac{4}{3}\frac{(j-1)^2}{n-1}+2\frac{1}{n-j+1}-\frac{5}{3},\end{aligned} \tag{1.145a}$$

$$
\begin{aligned}
B_{n,j}^{[E]} = &-\frac{2}{3}\sum_{k=j+1}^{n}\frac{H_{k-j}}{k}+\frac{1}{3}H_{n-j}^{2}\\
&+\left(\frac{2}{3}-\frac{1}{3}\frac{(j-1)(j-4)}{n}+\frac{1}{3}\frac{(j-1)(j-2)}{n-1}+\frac{2}{3}\frac{1}{n-j+1}\right)H_{n-j}+\frac{1}{3}H_n-\frac{1}{3}H_j\\
&-\frac{1}{3}H_{n-j}^{(2)}+\frac{2}{3}\frac{j(j-3)}{n}-\frac{2}{3}\frac{(j-1)^2}{n-1}+2\frac{1}{n-j+1}+\frac{1}{3}\,,
\end{aligned}
\tag{1.145b}
$$

$$
\begin{aligned}
C_{n,j}^{[E]} = &-\frac{2}{3}\sum_{k=j+1}^{n}\frac{H_{k-j}}{k}+\frac{1}{3}H_{n-j}^{2}\\
&+\left(\frac{8}{3}-\frac{1}{3}\frac{(j+2)(j-1)}{n}+\frac{1}{3}\frac{(j-1)(j-2)}{n-1}+\frac{2}{3}\frac{1}{n-j+1}\right)H_{n-j}+\frac{1}{3}H_n-\frac{1}{3}H_j\\
&-\frac{1}{3}H_{n-j}^{(2)}+\frac{1}{3}\frac{j(2j+3)}{n}-\frac{2}{3}\frac{(j-1)^2}{n-1}+2\frac{1}{n-j+1}-\frac{11}{3}\,.
\end{aligned}
\tag{1.145c}
$$

Nun bekommt man die gesuchten Erwartungswerte an rechtsseitigen Kanten leicht mit den Beziehungen

$$
A_{n,j}^{[R]} = A_{n,j}^{[E]} - A_{n,j}^{[L]}\,, \quad B_{n,j}^{[R]} = B_{n,j}^{[E]} - B_{n,j}^{[L]} \text{ und } C_{n,j}^{[R]} = C_{n,j}^{[E]} - C_{n,j}^{[L]}\,. \tag{1.146}
$$

Damit folgert man unmittelbar

Satz 1.5.4. *Die Erwartungswerte $A_{n,j}^{[R]}$, $B_{n,j}^{[R]}$ und $C_{n,j}^{[R]}$ der Anzahl der nach rechts führenden Kanten auf dem Weg von der Wurzel zum Knoten mit Schlüssel j in Objekten der Größe n aus den Familien $\mathcal{A}$, $\mathcal{B}$ und $\mathcal{C}$ sind gegeben durch*

$$
\boxed{\begin{aligned}
A_{n,j}^{[R]} = &\frac{2}{3}H_{n-j}-\frac{2}{3}H_n+\frac{2}{3}H_j+\frac{1}{3}\frac{2j^2-2j+1}{n}-\frac{2}{3}\frac{(j-1)^2}{n-1}\\
&+\frac{1}{3}\frac{1}{n-j+1}-\frac{1}{3j}-\frac{1}{3} \quad \text{für } 1 \le j \le n\,.
\end{aligned}}
\tag{1.147a}
$$

$$
\boxed{\begin{aligned}
B_{n,j}^{[R]} = &\frac{2}{3}H_{n-j}-\frac{2}{3}H_n+\frac{2}{3}H_j-\frac{1}{3}\frac{j^2-4j+2}{n}+\frac{1}{3}\frac{(j-1)^2}{n-1}\\
&+\frac{1}{3}\frac{1}{n-j+1}-\frac{1}{3j}-\frac{1}{3} \quad \text{für } 1 \le j \le n\,.
\end{aligned}}
\tag{1.147b}
$$

$$
\boxed{\begin{aligned}
C_{n,j}^{[R]} = &\frac{2}{3}H_{n-j}-\frac{2}{3}H_n+\frac{2}{3}H_j-\frac{1}{3}\frac{j^2+2j-1}{n}+\frac{1}{3}\frac{(j-1)^2}{n-1}\\
&+\frac{1}{3}\frac{1}{n-j+1}-\frac{1}{3j}+\frac{2}{3} \quad \text{für } 1 \le j \le n\,.
\end{aligned}}
\tag{1.147c}
$$

1.5.3 Asymptotische Betrachtungen und weitere Folgerungen

Mit den Beziehungen (1.92), (1.105), (1.110) und (1.112) erhalten wir aus Gleichung (1.140) sofort die asymptotischen Entwicklungen der linksseitigen Länge des Pfades von der Wurzel zu einem Knoten j.

Korollar 1.5.5. *Der Erwartungswert $A^{[L]}_{n,j}$ der Anzahl der nach links führenden Kanten auf dem Weg von der Wurzel zum Knoten mit Schlüssel j in einem p-tree der Größe n, das heißt also in einem Objekt aus der Familie $\mathcal{A}$, ist asymptotisch für $j = \rho n$ und $0 < \rho < 1$ gegeben durch*

$$\boxed{A^{[L]}_{n,j} = \frac{1}{3}\log^2 n + \left(-\frac{2}{3}\rho^2 + \frac{2}{3}\rho + \frac{2}{3}\gamma + 1 + \frac{2}{3}\log\rho + \frac{2}{3}\log(1-\rho)\right)\log n + \mathcal{O}(1)\,.} \tag{1.148}$$

Korollar 1.5.6. *Der Erwartungswert $A^{[L]}_{n,j}$ der Anzahl der nach links führenden Kanten auf dem Weg von der Wurzel zum Knoten mit Schlüssel j in einem p-tree der Größe n, das heißt also in einem Objekt aus der Familie $\mathcal{A}$, ist asymptotisch für fixes j gegeben durch*

$$\boxed{A^{[L]}_{n,j} = 2\log n + \mathcal{O}(1)\,.} \tag{1.149}$$

Korollar 1.5.7. *Der Erwartungswert $A^{[L]}_{n,j}$ der Anzahl der nach links führenden Kanten auf dem Weg von der Wurzel zum Knoten mit Schlüssel j in einem p-tree der Größe n, das heißt also in einem Objekt aus der Familie $\mathcal{A}$, ist asymptotisch für fixes $l = n - j$ gegeben durch*

$$\boxed{A^{[L]}_{n,j} = \frac{1}{3}H_l^2 + H_l - \frac{5}{3} - \frac{1}{3}H_l^{(2)} + \frac{2}{3}\frac{H_l}{l+1} + \frac{5}{3}\frac{1}{l+1} + \mathcal{O}(1/n)\,.} \tag{1.150}$$

Weiters ergibt sich für das arithmetische Mittel $A^{[L]}_n = \frac{1}{n}\sum_{j=1}^{n} A^{[L]}_{n,j}$ der Länge des linksseitigen Pfades von der Wurzel zu einem zufällig gewählten Knoten unter Verwendung der Beziehungen (1.85) und (1.117) durch Aufsummieren von (1.140)

Korollar 1.5.8. *Der Erwartungswert $A^{[L]}_n$ der Anzahl der nach links führenden Kanten auf dem Weg von der Wurzel zu einem zufällig gewählten Knoten in einem p-tree der Größe n, das heißt also in einem Objekt aus der Familie $\mathcal{A}$, ist gegeben durch*

$$\boxed{\begin{aligned} &A^{[L]}_n = \frac{n+1}{3n}H_n^2 - \frac{2(n-2)}{9n}H_n - \frac{n+1}{3n}H_n^{(2)} + \frac{2(n-2)}{27n} \quad \text{für } n \geq 2\,,\\ &A^{[L]}_1 = 0\,. \end{aligned}} \tag{1.151}$$

Für die rechtsseitige Länge des Pfades von der Wurzel zu einem Knoten j bekommen wir aus Gleichung (1.147a) mit den asymptotischen Entwicklungen (1.92) sofort

Korollar 1.5.9. *Der Erwartungswert $A^{[R]}_{n,j}$ der Anzahl der nach rechts führenden Kanten auf dem Weg von der Wurzel zum Knoten mit Schlüssel j in einem p-tree der Größe n, das heißt also in einem Objekt aus der Familie $\mathcal{A}$, ist asymptotisch für $j = \rho n$ und $0 < \rho < 1$ gegeben durch*

$$\boxed{A^{[R]}_{n,j} = \frac{2}{3}\log n - \frac{2}{3}\rho^2 + \frac{2}{3}\rho - \frac{1}{3} + \frac{2}{3}\gamma + \frac{2}{3}\log\rho + \frac{2}{3}\log(1-\rho) + \mathcal{O}(1/n)\,.} \tag{1.152}$$

Korollar 1.5.10. *Der Erwartungswert $A^{[R]}_{n,j}$ der Anzahl der nach rechts führenden Kanten auf dem Weg von der Wurzel zum Knoten mit Schlüssel j in einem p-tree der Größe n, das heißt also in einem Objekt aus der Familie $\mathcal{A}$, ist asymptotisch für fixes j gegeben durch*

$$\boxed{A^{[R]}_{n,j} = \frac{2}{3}H_j - \frac{1}{3} - \frac{1}{3j} + \mathcal{O}(1/n)\,.} \tag{1.153}$$

Korollar 1.5.11. *Der Erwartungswert $A_{n,j}^{[R]}$ der Anzahl der nach rechts führenden Kanten auf dem Weg von der Wurzel zum Knoten mit Schlüssel j in einem p-tree der Größe n, das heißt also in einem Objekt aus der Familie $\mathcal{A}$, ist asymptotisch für fixes $l = n - j$ gegeben durch*

$$\boxed{A_{n,j}^{[R]} = \frac{2}{3}H_l - \frac{1}{3} + \frac{1}{3}\frac{1}{l+1} + \mathcal{O}(1/n)\,.} \tag{1.154}$$

Für das arithmetische Mittel $A_n^{[R]} = \frac{1}{n}\sum_{j=1}^{n} A_{n,j}^{[L]}$ der Länge des rechtsseitigen Pfades zu einem zufällig gewählten Knoten unter Verwendung der Beziehungen (1.85) und (1.117) erhält man durch Aufsummieren von (1.147a)

Korollar 1.5.12. *Der Erwartungswert $A_n^{[R]}$ der Anzahl der nach rechts führenden Kanten auf dem Weg von der Wurzel zu einem zufällig gewählten Knoten in einem p-tree der Größe n, das heißt also in einem Objekt aus der Familie $\mathcal{A}$, ist gegeben durch*

$$\boxed{\begin{aligned} &A_n^{[R]} = \frac{2(n+1)}{3n}H_n - \frac{14n-1}{9n} \qquad \text{für } n \geq 2\,,\\ &A_1^{[R]} = 0\,. \end{aligned}} \tag{1.155}$$

1.6 Die Anzahl der Nachfolger eines Knotens

1.6.1 Die Erwartungswerte

Im folgenden betrachten wir die Anzahl der Nachfolger der Knoten in Objekten aus den Familien $\mathcal{A}$, $\mathcal{B}$ und $\mathcal{C}$. Unter der Anzahl der Nachfolger (= Nachkommen) eines Knotens p verstehen wir die Größe des entsprechenden Unterbaums mit Knoten p als Wurzel. Das heißt also der Knoten p wird quasi als sein eigener Nachkomme gezählt.

Seien nun also $A_{n,j}$, $B_{n,j}$ und $C_{n,j}$ die Erwartungswerte der Anzahl der Nachkommen des Knotens mit Schlüssel j in einem Objekt der Größe n aus den Familien $\mathcal{A}$, $\mathcal{B}$ und $\mathcal{C}$.

Um zu einer Rekursionsformel für $A_{n,j}$, $B_{n,j}$ und $C_{n,j}$ zu gelangen ist es nützlich, wenn man für die Objektfamilie $\mathcal{C}$ noch die Hilfsgrößen $q_{n,j}$, $C_{n,j}^{[Q]}$ und $C_{n,j}^{[\neg Q]}$ definiert.

$q_{n,j}$ soll hier die Wahrscheinlichkeit sein, daß sich der Knoten mit Schlüssel j im Objekt $\mathcal{C}$ der Größe n auf dem linken Pfad, das heißt auf dem Weg von der Wurzel bis zum linken Blatt, befindet. Das ist gleichbedeutend damit, daß der Knoten j ein links-rechts-Maximum in einer Zufallspermutation der Größe n darstellt.

Um nun $q_{n,j}$ zu berechnen, genügt die Überlegung, daß bei einer Anordnung der $n-j+1$ Zahlen von j bis n für ein links-rechts-Maximum der Schlüssel j vor den größeren kommen muß, während die Zahlen von 1 bis $j-1$ beliebig angeordnet werden können. Die Wahrscheinlichkeit, daß in einer Zufallspermutation der Schlüssel j vor den größeren Zahlen angeordnet ist, ist $\frac{1}{n-j+1}$. Das heißt also

$$q_{n,j} = \frac{1}{n-j+1}\,.$$

$C^{[Q]}_{n,j}$ ist der bedingte Erwartungswert der Anzahl der Nachfolger des Knotens mit Schlüssel j in einem Objekt der Größe n aus der Familie $\mathcal{C}$, unter der Bedingung, daß sich dieser Knoten j auf dem linken Pfad befindet.

$C^{[\neg Q]}_{n,j}$ ist der bedingte Erwartungswert der Anzahl der Nachfolger des Knotens mit Schlüssel j in einem Objekt der Größe n aus der Familie $\mathcal{C}$, unter der Bedingung, daß sich dieser Knoten j nicht auf dem linken Pfad befindet.

Teilt man nun die Familien $\mathcal{A}$, $\mathcal{B}$ und $\mathcal{C}$ aufgrund des ersten Elements k einer Zufallspermutation auf, so ergeben sich die Rekursionen

$$A_{n,j} = \frac{1}{n}\left(\sum_{k=1}^{j-1}\left(q_{n-k,j-k}\left(C^{[Q]}_{n-k,j-k}+k\right)+(1-q_{n-k,j-k})\,C^{[\neg Q]}_{n-k,j-k}\right)+j+\sum_{k=j+1}^{n}B_{k-1,j}\right)$$
$$\text{für } n \geq j \geq 1\,, \tag{1.156a}$$

$$B_{n,j} = \frac{1}{n}\left(\sum_{k=1}^{j-1}A_{n-k,j-k}+j+\sum_{k=j+1}^{n}B_{k-1,j}\right) \quad \text{für } n \geq j \geq 1\,, \tag{1.156b}$$

$$C_{n,j} = \frac{1}{n}\left(\sum_{k=1}^{j-1}\left(q_{n-k,j-k}\left(C^{[Q]}_{n-k,j-k}+k\right)+(1-q_{n-k,j-k})\,C^{[\neg Q]}_{n-k,j-k}\right)+j+\sum_{k=j+1}^{n}A_{k-1,j}\right)$$
$$\text{für } n \geq j \geq 1\,. \tag{1.156c}$$

Wegen des Satzes von der vollständigen Erwartung gilt natürlich

$$C_{n,j} = q_{n,j}C^{[Q]}_{n,j}+(1-q_{n,j})\,C^{[\neg Q]}_{n,j}\,. \tag{1.157}$$

Unter Berücksichtigung dieser Beziehung erhält man nach dem Einsetzen von $q_{n,j}$ und der Angabe der Anfangswerte folgende Rekursionsformeln

$$A_{n,j} = \frac{1}{n}\left(\sum_{k=1}^{j-1}\left(C_{n-k,j-k}+\frac{k}{n-j+1}\right)+j+\sum_{k=j+1}^{n}B_{k-1,j}\right) \quad \text{für } n \geq j \geq 1\,, \tag{1.158a}$$

$$B_{n,j} = \frac{1}{n}\left(\sum_{k=1}^{j-1}A_{n-k,j-k}+j+\sum_{k=j+1}^{n}B_{k-1,j}\right) \quad \text{für } n \geq j \geq 1\,, \tag{1.158b}$$

$$C_{n,j} = \frac{1}{n}\left(\sum_{k=1}^{j-1}\left(C_{n-k,j-k}+\frac{k}{n-j+1}\right)+j+\sum_{k=j+1}^{n}A_{k-1,j}\right) \quad \text{für } n \geq j \geq 1\,, \tag{1.158c}$$

$$A_{0,0} = 0\,, \quad B_{0,0} = 0\,, \quad C_{0,0} = 0\,.$$

Führt man die bivariaten erzeugenden Funktionen in z und u

$$A(z,u) = \sum_{j\geq 1}\sum_{n\geq j}A_{n,j}z^n u^j\,, \quad B(z,u) = \sum_{j\geq 1}\sum_{n\geq j}B_{n,j}z^n u^j\,, \quad C(z,u) = \sum_{j\geq 1}\sum_{n\geq j}C_{n,j}z^n u^j$$

ein, so transformieren sich obige Rekursionsformeln (1.158) durch Multiplikation mit $nz^{n-1}u^j$ und Aufsummieren über $n \geq j \geq 1$ zu einem System von Differentialgleichungen:

$$\begin{aligned}
\frac{\partial}{\partial z}A(z,u) &= \frac{u}{1-uz}C(z,u) + \frac{u^2}{(1-uz)^3}\log\left(\frac{1}{1-z}\right) + \frac{u}{(1-z)(1-uz)^2} \\
&\quad + \frac{1}{1-z}B(z,u)\,, &(1.159a)\\
\frac{\partial}{\partial z}B(z,u) &= \frac{u}{1-uz}A(z,u) + \frac{u}{(1-z)(1-uz)^2} + \frac{1}{1-z}B(z,u)\,, &(1.159b)\\
\frac{\partial}{\partial z}C(z,u) &= \frac{u}{1-uz}C(z,u) + \frac{u^2}{(1-uz)^3}\log\left(\frac{1}{1-z}\right) + \frac{u}{(1-z)(1-uz)^2} \\
&\quad + \frac{1}{1-z}A(z,u)\,. &(1.159c)
\end{aligned}$$

Mit den Bezeichnungen

$$\begin{aligned}
f_1(z,u) &= \frac{u^2}{(1-uz)^3}\log\left(\frac{1}{1-z}\right) + \frac{u}{(1-z)(1-uz)^2}\,,\\
f_2(z,u) &= \frac{u}{(1-z)(1-uz)^2}\,,\\
f_3(z,u) &= \frac{u^2}{(1-uz)^3}\log\left(\frac{1}{1-z}\right) + \frac{u}{(1-z)(1-uz)^2}
\end{aligned}$$

schreibt sich dieses als

$$\begin{aligned}
\frac{\partial}{\partial z}A(z,u) &= \frac{u}{1-uz}C(z,u) + \frac{1}{1-z}B(z,u) + f_1(z,u)\,, &(1.160a)\\
\frac{\partial}{\partial z}B(z,u) &= \frac{u}{1-uz}A(z,u) + \frac{1}{1-z}B(z,u) + f_2(z,u)\,, &(1.160b)\\
\frac{\partial}{\partial z}C(z,u) &= \frac{u}{1-uz}C(z,u) + \frac{1}{1-z}A(z,u) + f_3(z,u)\,. &(1.160c)
\end{aligned}$$

Da dieses System von linearen Differentialgleichungen bis auf sich unterscheidende Störfunktionen $f_1(z,u)$, $f_2(z,u)$ und $f_3(z,u)$ ident ist mit (1.34), kann man zur Lösung von (1.160) analog wie dort vorgehen.

Tut man dies, kommt man zur folgenden linearen Differentialgleichung dritter Ordnung für $A(z,u)$:

$$\begin{aligned}
&\frac{2uz-u-1}{(1-z)(1-u)}A(z,u) - 2\frac{1-uz}{1-u}\frac{\partial}{\partial z}A(z,u)\\
&- \frac{3}{2}\frac{(1-uz)(2uz-u-1)}{u(1-u)}\frac{\partial^2}{\partial z^2}A(z,u) - \frac{1}{2}\frac{(1-uz)^2(1-z)}{u(1-u)}\frac{\partial^3}{\partial z^3}A(z,u)\\
&= f_3(z,u) + \frac{1}{1-uz}\frac{\partial}{\partial z}\Bigg[\frac{(1-uz)^3}{2u(u-1)}f_2(z,u)\\
&\quad + \frac{(1-uz)^3}{2u(u-1)(1-z)}\frac{\partial}{\partial z}\left((1-z)^2 f_1(z,u)\right) + \frac{(1-uz)^2(1-z)}{2(u-1)}f_3(z,u)\Bigg]\,.
\end{aligned} \tag{1.161}$$

Setzt man nun für $f_1(z,u)$, $f_2(z,u)$ und $f_3(z,u)$ ein und vereinfacht die Ausdrücke, so folgt unter Berücksichtigung der Anfangswerte $A_{0,0}=0$, $A_{1,1}=1$, $A_{2,1}=1$ und $A_{2,2}=2$ als zu lösendes Differentialgleichungssystem

$$\begin{aligned}
&\frac{2uz-u-1}{(1-z)(1-u)}A(z,u)-2\frac{1-uz}{1-u}\frac{\partial}{\partial z}A(z,u)\\
&-\frac{3}{2}\frac{(1-uz)(2uz-u-1)}{u(1-u)}\frac{\partial^2}{\partial z^2}A(z,u)-\frac{1}{2}\frac{(1-uz)^2(1-z)}{u(1-u)}\frac{\partial^3}{\partial z^3}A(z,u)\\
&\quad=3\frac{u^2}{(1-uz)^3}\log\left(\frac{1}{1-z}\right)+\frac{u(uz-3u+2)}{(1-u)(1-z)(1-uz)^2}\,,\\
&A(0,u)=0\,,\quad \frac{\partial}{\partial z}A(z,u)\bigg|_{z=0}=u\,,\quad \frac{\partial^2}{\partial z^2}A(z,u)\bigg|_{z=0}=2\left(u+2u^2\right)\,.
\end{aligned}\tag{1.162}$$

Mit den Transformationen

$$\begin{aligned}
A(z,u)&=\frac{E(z,u)}{(1-z)(1-uz)}\,,\\
F(z,u)&=\frac{\partial}{\partial z}E(z,u)\,,\\
F(z,u)&=(1-z)(1-uz)G(z,u)\,,\\
H(z,u)&=\frac{\partial}{\partial z}G(z,u)\,,
\end{aligned}$$

ergibt sich daraus die lineare Differentialgleichung erster Ordnung

$$\begin{aligned}
\frac{\partial}{\partial z}H(z,u)-2\frac{1+u-2uz}{(1-z)(1-uz)}H(z,u)&=6\frac{u^3(u-1)}{(1-z)(1-uz)^5}\log\left(\frac{1}{1-z}\right)\\
&\quad-2\frac{u^2(uz-3u+2)}{(1-z)^2(1-uz)^4}\,.
\end{aligned}\tag{1.163}$$

Als eine Partikulärlösung obiger Differentialgleichung findet man durch Variation der Konstanten

$$H_p(z,u)=\frac{1}{(1-z)^2(1-uz)^2}\int_0^z m(t,u)\,dt\,,\tag{1.164}$$

wobei $m(z,u)$ gegeben ist durch

$$m(z,u)=\left(6\frac{u^2(u-1)^2}{(1-uz)^3}+6\frac{u^2(u-1)}{(1-uz)^2}\right)\log\left(\frac{1}{1-z}\right)+6\frac{u^2(u-1)}{(1-uz)^2}+2\frac{u^2}{1-uz}\,.\tag{1.165}$$

Führt man diese Integrationen aus, schreibt sich diese Partikulärlösung für (1.163) als

$$\begin{aligned}
H_p(z,u)&=3\frac{u^3}{(1-uz)^4}\log\left(\frac{1}{1-z}\right)-\frac{u}{(1-z)^2(1-uz)^2}\log\left(\frac{1}{1-uz}\right)\\
&\quad+3\frac{u(u-1)}{(1-z)^2(1-uz)^3}\,.
\end{aligned}\tag{1.166}$$

Durch Rücksubstitution erhält man daraus als Partikulärlösung für (1.162)

$$A_p(z,u) = \frac{1}{(1-z)(1-uz)} \int_0^z \left[(1-x)(1-ux) \int_0^x H_p(t,u)\,dt\right] dx\,. \tag{1.167}$$

Für ein leichteres Ablesen der Koeffizienten wendet man bei obigem Ausdruck partielle Integration an. Sei

$$g(z,u) = \int_0^z (1-x)(1-ux)\,dx\,,$$

dann bekommen wir zunächst

$$A_p(z,u) = \frac{g(z,u)}{(1-z)(1-uz)} \int_0^z H_p(t,u)\,dt - \frac{1}{(1-z)(1-uz)} \int_0^z g(t,u)H_p(t,u)\,dt\,. \tag{1.168}$$

Daraus ergibt sich

$$\begin{aligned}
A_p(z,u) = \Bigg[&-\frac{1}{3}\frac{u^2(2uz-u-1)}{(1-u)^3} - \frac{4}{3}\frac{u^2}{(1-uz)(1-u)} \\
&+ \frac{u}{(1-z)(1-u)}\Bigg] \int_0^z \frac{1}{1-ut}\log\left(\frac{1}{1-t}\right) dt \qquad (1.169)\\
&+ \left(\frac{1}{3}\frac{u(2uz-u-1)}{(1-u)^3} + \frac{1}{3}\frac{u}{(1-uz)(1-u)}\right)\log\left(\frac{1}{1-z}\right)\log\left(\frac{1}{1-uz}\right) \\
&+ \left(-\frac{1}{6}\frac{u(2uz-u-1)}{(1-u)^3} - \frac{1}{6}\frac{1}{(1-z)(1-u)}\right)\log^2\left(\frac{1}{1-uz}\right) \\
&- \left(-\frac{3}{2}\frac{u(2uz-u-1)}{(1-u)^3} - \frac{u}{(1-uz)^2} - \frac{3}{2}\frac{u}{(1-uz)(1-u)}\right)\log\left(\frac{1}{1-z}\right) \\
&- \left(\frac{1}{6}\frac{u(18uz-5u-13)}{(1-u)^3} + \frac{3}{2}\frac{1}{(1-z)(1-u)}\right)\log\left(\frac{1}{1-uz}\right) \\
&+ \frac{7}{3}\frac{u}{(1-u)^2} + \frac{u}{(1-uz)(1-u)} - \frac{5}{6}\frac{1}{(1-z)(1-u)} \qquad (1.170)
\end{aligned}$$

und weiter

$$\begin{aligned}
A_p(z,u) = \Bigg[&-\frac{1}{3}uz + \frac{1}{6}u + \frac{1}{6} + \frac{1}{2}\frac{(u-1)^2}{u(1-z)} \\
&- \frac{2}{3}\frac{(u-1)^2}{1-uz}\Bigg] \int_0^z \frac{1}{(1-t)^2(1-ut)^2}\log\left(\frac{1}{1-t}\right) dt \\
&+ \left(-\frac{1}{6}\frac{(2uz-u-1)u}{(1-u)^3} - \frac{2}{3}\frac{u}{(1-uz)(1-u)} + \frac{1}{2}\frac{1}{(1-z)(1-u)}\right)\log^2\left(\frac{1}{1-z}\right) \\
&+ \left(\frac{1}{3}\frac{(2uz-u-1)u}{(1-u)^3} + \frac{1}{3}\frac{u}{(1-uz)(1-u)}\right)\log\left(\frac{1}{1-z}\right)\log\left(\frac{1}{1-uz}\right) \\
&+ \left(-\frac{1}{6}\frac{(2uz-u-1)u}{(1-u)^3} - \frac{1}{6}\frac{1}{(1-z)(1-u)}\right)\log^2\left(\frac{1}{1-uz}\right) \\
&- \Bigg[-\frac{2}{3}\frac{(4uz-u-3)u}{(1-u)^3} - \frac{1}{2}\frac{1}{(1-z)(1-u)} + \frac{1}{2}\frac{1}{(1-z)^2u} - \frac{5}{6}\frac{u}{(1-uz)(1-u)} \\
&\quad - \frac{5}{3}\frac{u}{(1-uz)^2}\Bigg]\log\left(\frac{1}{1-z}\right)
\end{aligned}$$

$$
\begin{aligned}
&-\left(\frac{2}{3}\frac{(4uz-u-3)u}{(1-u)^3}+\frac{2}{3}\frac{u}{(1-uz)(1-u)}+2\frac{1}{(1-z)(1-u)}\right)\log\left(\frac{1}{1-uz}\right)\\
&+\frac{8}{3}\frac{u}{(1-u)^2}-\frac{4}{3}\frac{1}{(1-z)(1-u)}+\frac{1}{2}\frac{1}{(1-z)^2u}+\frac{5}{3}\frac{u}{(1-uz)(1-u)}\,.
\end{aligned}
\tag{1.171}
$$

Zum Anpassen an die Anfangswerte benötigt man nun noch die allgemeine Lösung der (1.160) entsprechenden homogenen Differentialgleichung. Da diese homogene Differentialgleichung ident ist mit (1.42), übernimmt man von dort

$$
\begin{aligned}
A_h(z,u)=&\frac{c_1(u)}{(1-z)(1-uz)}+c_2(u)\frac{z(6-3uz-3z+2uz^2)}{(1-z)(1-uz)}\\
&+\frac{c_3(u)}{(1-z)(1-uz)}\int_0^z\left[(1-x)(1-ux)\int_0^x\frac{1}{(1-t)^2(1-ut)^2}\,dt\right]dx\,,
\end{aligned}
\tag{1.172}
$$

mit drei nur von u abhängigen Funktionen $c_1(u)$, $c_2(u)$ und $c_3(u)$.

Die sich aus den Anfangsbedingungen ergebenden Gleichungen

$$
\begin{aligned}
&A(z,u)=A_p(z,u)+A_h(z,u)\,,\\
&A(0,u)=0\,,\quad \frac{\partial}{\partial z}A(z,u)\bigg|_{z=0}=u\,,\quad \frac{\partial^2}{\partial z^2}A(z,u)\bigg|_{z=0}=2\left(u+2u^2\right)
\end{aligned}
$$

führen zu folgenden Lösungen für die Funktionen $c_i(u)$, $1\le i\le 3$:

$$
c_1(u)=\frac{1}{18}\frac{10u^3-95u^2+22u-9}{(1-u)^2u}\,,\quad c_2(u)=\frac{1}{6}\frac{u}{1-u}\ \text{ und }\ c_3(u)=-\frac{2}{9}\frac{1}{(1-u)^3}\,.
$$

Damit erhält man nach Vereinfachungen als Lösung der bivariaten erzeugenden Funktion $A(z,u)$ für die Erwartungswerte $A_{n,j}$ der Anzahl der Nachkommen des Knotens mit Schlüssel j in einem p-tree der Größe n:

$$
\begin{aligned}
A(z,u)=&\left[-\frac{1}{3}uz+\frac{1}{6}u+\frac{1}{6}+\frac{1}{2}\frac{(u-1)^2}{u(1-z)}\right.\\
&\left.-\frac{2}{3}\frac{(u-1)^2}{1-uz}\right]\int_0^z\frac{1}{(1-t)^2(1-ut)^2}\log\left(\frac{1}{1-t}\right)\,dt\\
&+\left(-\frac{1}{6}\frac{(2uz-u-1)u}{(1-u)^3}+\frac{1}{2}\frac{1}{(1-z)(1-u)}-\frac{2}{3}\frac{u}{(1-uz)(1-u)}\right)\log^2\left(\frac{1}{1-z}\right)\\
&+\left(+\frac{1}{3}\frac{(2uz-u-1)u}{(1-u)^3}+\frac{1}{3}\frac{u}{(1-uz)(1-u)}\right)\log\left(\frac{1}{1-z}\right)\log\left(\frac{1}{1-uz}\right)\\
&+\left(-\frac{1}{6}\frac{(2uz-u-1)u}{(1-u)^3}-\frac{1}{6}\frac{1}{(1-z)(1-u)}\right)\log^2\left(\frac{1}{1-uz}\right)\\
&+\left[-\frac{2}{3}\frac{u}{(1-u)^2}+\frac{1}{2}\frac{1}{(1-z)(1-u)}-\frac{1}{2}\frac{1}{(1-z)^2u}-\frac{1}{2}\frac{u}{(1-uz)(1-u)}\right.\\
&\qquad\left.+\frac{5}{3}\frac{u}{(1-uz)^2}\right]\log\left(\frac{1}{1-z}\right)\\
&+\left(\frac{2}{3}\frac{u}{(1-u)^2}-\frac{2}{3}\frac{1}{(1-z)(1-u)}+\frac{2}{3}\frac{u}{(1-uz)(1-u)}\right)\log\left(\frac{1}{1-uz}\right)
\end{aligned}
$$

$$+\frac{1}{6}\frac{2uz-u-1}{1-u}+\frac{1}{6}\frac{u^2-3}{(1-z)(1-u)u}+\frac{1}{2}\frac{1}{(1-z)^2u}+\frac{2}{3}\frac{1}{(1-uz)(1-u)}\,. \quad (1.173)$$

Für das Ablesen der Koeffizienten von (1.173) benötigt man außer den Beziehungen (1.19), (1.62) und (1.86) noch

$$[z^nu^j]\frac{1}{(1-u)^3}\log\left(\frac{1}{1-z}\right)\log\left(\frac{1}{1-uz}\right)$$
$$= -\frac{1}{2}j+\frac{1}{2}\frac{j^2+3j+2}{n}H_j+\frac{1}{2}\frac{n^2-2jn-3n+j^2+3j+2}{n}\left(H_{n-1}-H_{n-j-1}\right)\,, \quad (1.174a)$$

$$[z^nu^j]\frac{1}{(1-u)(1-uz)}\log\left(\frac{1}{1-z}\right)\log\left(\frac{1}{1-uz}\right)=\sum_{k=1}^{j}\frac{H_k}{n-k}\,, \quad (1.174b)$$

$$[z^nu^j]\frac{1}{(1-z)(1-u)}\log^2\left(\frac{1}{1-uz}\right)=H_j^2-H_j^{(2)}\,, \quad (1.174c)$$

$$[z^nu^j]\frac{1}{(1-u)^2}\log\left(\frac{1}{1-z}\right)=\frac{j+1}{n}\,, \quad (1.174d)$$

$$[z^nu^j]\frac{1}{(1-uz)^2}\log\left(\frac{1}{1-z}\right)=\frac{j+1}{n-j}\,. \quad (1.174e)$$

Damit ergibt sich nach dem Koeffizientenablesen und dem Vereinfachen der gewonnenen Ausdrücke für die gesuchten Erwartungswerte

$$\begin{aligned}A_{n,j} &= \frac{1}{3}\sum_{k=1}^{j-1}\frac{H_k}{n-k}-\sum_{k=j+1}^{n}\frac{H_{k-j}}{k}-\frac{1}{6}H_n^2-\frac{2}{3}H_{n-j}^2+\frac{4}{3}H_{n-j}H_n-\frac{1}{6}H_j^2\\ &+\frac{1}{6}\frac{j-6}{j}H_n-\frac{1}{6}\frac{j-6}{j}H_{n-j}-\frac{1}{6}\frac{n^2j-6n^2+2nj^2-nj+6n-2j^3}{n(n-1)j}H_j\\ &+\frac{1}{6}H_n^{(2)}-\frac{2}{3}H_{n-j}^{(2)}+\frac{1}{6}H_j^{(2)}+\frac{1}{6}\frac{n^3+6n^2j-3nj^2-3nj-n+2j^3-3j^2+j}{n(n-1)(n-j+1)}\,.\end{aligned} \quad (1.175)$$

Dieses Ergebnis kann noch etwas vereinfacht werden, indem man nachfolgende Formeln anwendet. Es gilt nämlich

$$\sum_{k=j+1}^{n}\frac{H_{k-j}}{k}+\sum_{k=1}^{j-1}\frac{H_{n-k}}{k}=\frac{1}{2}\left(H_n^2-H_n^{(2)}\right)-\frac{1}{2}\left(H_{j-1}^2+H_{j-1}^{(2)}\right)+H_nH_{j-1}\,, \quad (1.176a)$$

$$\sum_{k=1}^{j}\frac{H_{n-k}}{k}-\sum_{k=1}^{j}\frac{H_k}{n-k}=H_jH_{n-j-1}\,. \quad (1.176b)$$

Zum Beweis dieser Identitäten setzt man doppelte erzeugende Funktionen an, das heißt man multipliziert beide Seiten obiger Gleichungen (1.176) mit z^nu^j und summiert über alle $n \geq j \geq 1$ auf.

Dann übersetzt sich (1.176a) in

$$\begin{aligned}
&\frac{u}{1-z}\int_0^z \frac{1}{(1-t)(1-ut)}\log\left(\frac{1}{1-t}\right)dt + \frac{u}{(1-u)(1-z)}\log\left(\frac{1}{1-z}\right)\log\left(\frac{1}{1-uz}\right)\\
&-\frac{u}{(1-u)(1-uz)}\log^2\left(\frac{1}{1-uz}\right)\\
&\quad=\frac{1}{2}\frac{u}{(1-u)(1-z)}\log^2\left(\frac{1}{1-z}\right)-\frac{1}{2}\frac{u}{(1-u)(1-uz)}\log^2\left(\frac{1}{1-uz}\right)\\
&\qquad-\frac{uz}{(1-z)(1-uz)}\int_0^z\frac{1}{t(1-ut)}\log\left(\frac{1}{1-ut}\right)dt\\
&\qquad+\frac{u}{1-z}\int_0^z\frac{1}{(1-t)(1-ut)}\log\left(\frac{1}{1-ut}\right)dt+\frac{uz}{(1-z)(1-uz)}\log^2\left(\frac{1}{1-uz}\right)\\
&\qquad+\frac{uz}{(1-z)(1-uz)}\int_0^z\frac{1}{t}\log\left(\frac{1}{1-ut}\right)dt
\end{aligned}$$

bzw. (1.176b) in

$$\begin{aligned}
&\frac{1}{(1-u)(1-z)}\log\left(\frac{1}{1-z}\right)\log\left(\frac{1}{1-uz}\right)-\frac{u}{(1-u)(1-uz)}\log^2\left(\frac{1}{1-uz}\right)\\
&-\frac{1}{(1-u)(1-uz)}\log\left(\frac{1}{1-z}\right)\log\left(\frac{1}{1-uz}\right)+\frac{u}{(1-u)(1-uz)}\log^2\left(\frac{1}{1-uz}\right)\\
&\quad=\frac{z}{(1-z)(1-uz)}\log\left(\frac{1}{1-z}\right)\log\left(\frac{1}{1-uz}\right).
\end{aligned}$$

Die Korrektheit dieser Identitäten läßt sich durch Nachrechnen leicht zeigen.

Damit erhält man schließlich

Satz 1.6.1. *Der Erwartungswert $A_{n,j}$ der Anzahl der Nachkommen des Knotens mit Schlüssel j in einem p-tree der Größe n, das heißt also in einem Objekt der Größe n aus der Familie $\mathcal{A}$, ist gegeben durch*

$$\boxed{\begin{aligned}
A_{n,j} = {}&\frac{4}{3}\sum_{k=1}^{j}\frac{H_{n-k}}{k}-\frac{2}{3}H_{n-j}^2+\frac{4}{3}H_{n-j}H_n-\frac{1}{3}H_{n-j}H_j-\frac{2}{3}H_n^2-H_nH_j\\
&+\frac{1}{3}H_j^2-\frac{1}{6}H_{n-j}+\frac{1}{6}H_n+\left(-\frac{1}{6}+\frac{1}{3}\frac{j(j-1)}{n-1}-\frac{1}{3}\frac{j^2}{n}\right)H_j\\
&-\frac{2}{3}H_{n-j}^{(2)}+\frac{2}{3}H_n^{(2)}+\frac{2}{3}H_j^{(2)}\\
&+\frac{j}{n-j+1}-\frac{1}{3}\frac{j(j-1)}{n-1}+\frac{1}{6}\frac{j(2j-1)}{n}+\frac{1}{6}\quad \text{für } 1\le j\le n\,.
\end{aligned}} \tag{1.177}$$

Um zu den Erwartungswerten $C_{n,j}$ der Anzahl der Nachkommen des Knotens mit Schlüssel j in Objekten der Größe n aus der Familie $\mathcal{C}$ zu kommen, ermitteln wir zuerst deren erzeugende Funktion $C(z,u)$. Da wir zur Lösung unseres Gleichungssystems (1.160) analog wie in Unterabschnitt 1.4.2 vorgegangen sind, können wir zum Erhalt von $C(z,u)$ das eben errechnete

$A(z,u)$ in Gleichung (1.39) einsetzen und erhalten so unter Verwendung von (1.35)

$$\begin{aligned}
C(z,u) = \Bigg[&\frac{1}{6}uz + \frac{1}{6}u - \frac{1}{3} - \frac{2}{3}\frac{(u-1)^2}{1-uz} \\
&+ \frac{1}{2}\frac{(u-1)^2}{u(1-z)}\Bigg] \int_0^z \frac{1}{(1-t)^2(1-ut)^2} \log\left(\frac{1}{1-t}\right) dt \\
&+ \left(\frac{1}{6}\frac{(uz+u-2)u}{(1-u)^3} - \frac{2}{3}\frac{u}{(1-uz)(1-u)} + \frac{1}{2}\frac{1}{(1-z)(1-u)}\right) \log^2\left(\frac{1}{1-z}\right) \\
&+ \left(-\frac{1}{3}\frac{(uz+u-2)u}{(1-u)^3} + \frac{1}{3}\frac{u}{(1-uz)(1-u)}\right) \log\left(\frac{1}{1-z}\right) \log\left(\frac{1}{1-uz}\right) \\
&+ \left(\frac{1}{6}\frac{(uz+u-2)u}{(1-u)^3} - \frac{1}{6}\frac{1}{(1-z)(1-u)}\right) \log^2\left(\frac{1}{1-uz}\right) \\
&+ \Bigg[\frac{1}{3}\frac{u}{(1-u)^2} + \frac{1}{(1-z)(1-u)} - \frac{1}{2}\frac{1}{(1-z)^2u} - \frac{u}{(1-uz)(1-u)} \\
&\qquad + \frac{5}{3}\frac{u}{(1-uz)^2}\Bigg] \log\left(\frac{1}{1-z}\right) \\
&+ \left(-\frac{1}{3}\frac{u}{(1-u)^2} - \frac{2}{3}\frac{1}{(1-z)(1-u)} + \frac{2}{3}\frac{u}{(1-uz)(1-u)}\right) \log\left(\frac{1}{1-uz}\right) \\
&- \frac{1}{6}\frac{uz+u-2}{1-u} + \frac{1}{6}\frac{u^2-3u-3}{(1-z)(1-u)u} + \frac{1}{2}\frac{1}{(1-z)^2u} + \frac{2}{3}\frac{1}{(1-uz)(1-u)}\,. \qquad (1.178)
\end{aligned}$$

Zum Ablesen der Koeffizienten dieses Ausdrucks verwenden wir wieder (1.62) und (1.174) und gewinnen dadurch

$$\begin{aligned}
C_{n,j} = &\frac{1}{3}\sum_{k=1}^{j-1}\frac{H_k}{n-k} - \sum_{k=j+1}^{n}\frac{H_{k-j}}{k} - \frac{1}{6}H_n^2 - \frac{2}{3}H_{n-j}^2 + \frac{4}{3}H_{n-j}H_n - \frac{1}{6}H_j^2 \\
&- \frac{1}{3}\frac{j+3}{j}H_n + \frac{1}{3}\frac{j+3}{j}H_{n-j} - \frac{1}{6}\frac{n^2j-6n^2-4nj^2-nj+6n+j^3+3j^2}{n(n-1)j}H_j \\
&+ \frac{1}{6}H_n^{(2)} - \frac{2}{3}H_{n-j}^{(2)} + \frac{1}{6}H_j^{(2)} + \frac{1}{6}\frac{n^3+3n^2j+3nj^2-6nj-n-j^3+j}{n(n-1)(n-j+1)}\,. \qquad (1.179)
\end{aligned}$$

Mit den Beziehungen (1.176) ergibt sich daraus

Hilfssatz 1.6.2. *Der Erwartungswert $C_{n,j}$ der Anzahl der Nachkommen des Knotens mit Schlüssel j in einem Objekt der Größe n aus der Familie $\mathcal{C}$ ist gegeben durch*

$$\boxed{\begin{aligned}
C_{n,j} = &\frac{4}{3}\sum_{k=1}^{j}\frac{H_{n-k}}{k} - \frac{2}{3}H_{n-j}^2 + \frac{4}{3}H_{n-j}H_n - \frac{1}{3}H_{n-j}H_j - \frac{2}{3}H_n^2 - H_nH_j \\
&+ \frac{1}{3}H_j^2 + \frac{1}{3}H_{n-j} - \frac{1}{3}H_n + \left(-\frac{1}{6} - \frac{1}{6}\frac{j(j-1)}{n-1} + \frac{1}{6}\frac{j(j+3)}{n}\right)H_j \\
&- \frac{2}{3}H_{n-j}^{(2)} + \frac{2}{3}H_n^{(2)} + \frac{2}{3}H_j^{(2)} \\
&+ \frac{j}{n-j+1} + \frac{1}{6}\frac{j(j-1)}{n-1} - \frac{1}{6}\frac{j(j+1)}{n} + \frac{1}{6} \quad \text{für } 1 \le j \le n\,.
\end{aligned}} \qquad (1.180)$$

Für die Berechnung von $B(z,u)$ können wir wiederum die Beziehungen (1.37) und (1.35) verwenden. Damit erhalten wir

$$\begin{aligned}
B(z,u) = \Bigg[&\frac{1}{6}uz - \frac{1}{3}u + \frac{1}{6} - \frac{2}{3}\frac{(u-1)^2}{1-uz} \\
&+ \frac{1}{2}\frac{(u-1)^2}{u(1-z)}\Bigg] \int_0^z \frac{1}{(1-t)^2(1-ut)^2}\log\left(\frac{1}{1-t}\right)dt \\
&+ \left(\frac{1}{6}\frac{(uz-2u+1)u}{(1-u)^3} - \frac{2}{3}\frac{u}{(1-uz)(1-u)} + \frac{1}{2}\frac{1}{(1-z)(1-u)}\right)\log^2\left(\frac{1}{1-z}\right) \\
&+ \left(-\frac{1}{3}\frac{(uz-2u+1)u}{(1-u)^3} + \frac{1}{3}\frac{u}{(1-uz)(1-u)}\right)\log\left(\frac{1}{1-z}\right)\log\left(\frac{1}{1-uz}\right) \\
&+ \left(\frac{1}{6}\frac{(uz-2u+1)u}{(1-u)^3} - \frac{1}{6}\frac{1}{(1-z)(1-u)}\right)\log^2\left(\frac{1}{1-uz}\right) \\
&+ \Bigg[\frac{1}{3}\frac{u}{(1-u)^2} + \frac{1}{2}\frac{1}{(1-z)(1-u)} - \frac{1}{2}\frac{1}{(1-z)^2u} - \frac{1}{2}\frac{u}{(1-uz)(1-u)} \\
&\quad + \frac{7}{6}\frac{u}{(1-uz)^2}\Bigg]\log\left(\frac{1}{1-z}\right) \\
&+ \left(-\frac{1}{3}\frac{u}{(1-u)^2} - \frac{2}{3}\frac{1}{(1-z)(1-u)} + \frac{2}{3}\frac{u}{(1-uz)(1-u)}\right)\log\left(\frac{1}{1-uz}\right) \\
&- \frac{1}{6}\frac{uz-2u+1}{1-u} + \frac{1}{6}\frac{u^2-3}{(1-z)(1-u)u} + \frac{1}{2}\frac{1}{(1-z)^2u} - \frac{1}{6}\frac{3u-4}{(1-uz)(1-u)}\,. \qquad (1.181)
\end{aligned}$$

Für das Ablesen der Koeffizienten obigen Ausdrucks können wir wieder die Beziehungen (1.62) und (1.174) verwenden und bekommen daraus

$$\begin{aligned}
B_{n,j} = &\frac{1}{3}\sum_{k=1}^{j-1}\frac{H_k}{n-k} - \sum_{k=j+1}^{n}\frac{H_{k-j}}{k} - \frac{1}{6}H_n^2 - \frac{2}{3}H_{n-j}^2 + \frac{4}{3}H_{n-j}H_n - \frac{1}{6}H_j^2 \\
&+ \frac{1}{3}\frac{2j-3}{j}H_n - \frac{1}{3}\frac{2j-3}{j}H_{n-j} - \frac{1}{6}\frac{n^2j-6n^2+2nj^2-nj+6n+j^3-3j^2}{n(n-1)j}H_j \\
&+ \frac{1}{6}H_n^{(2)} - \frac{2}{3}H_{n-j}^{(2)} + \frac{1}{6}H_j^{(2)} + \frac{1}{6}\frac{n^3+3n^2j-3nj-n-j^3+3j^2-2j}{n(n-1)(n-j+1)}\,. \qquad (1.182)
\end{aligned}$$

Mit den Beziehungen (1.176) ergibt sich schließlich

Hilfssatz 1.6.3. *Der Erwartungswert $B_{n,j}$ der Anzahl der Nachkommen des Knotens mit Schlüssel j in einem Objekt der Größe n aus der Familie $\mathcal{B}$ ist gegeben durch*

$$\boxed{\begin{aligned}
B_{n,j} = &\frac{4}{3}\sum_{k=1}^{j}\frac{H_{n-k}}{k} - \frac{2}{3}H_{n-j}^2 + \frac{4}{3}H_{n-j}H_n - \frac{1}{3}H_{n-j}H_j - \frac{2}{3}H_n^2 - H_nH_j \\
&+ \frac{1}{3}H_j^2 - \frac{2}{3}H_{n-j} + \frac{2}{3}H_n + \left(-\frac{1}{6} - \frac{1}{6}\frac{j(j-1)}{n-1} + \frac{1}{6}\frac{j(j-3)}{n}\right)H_j \\
&- \frac{2}{3}H_{n-j}^{(2)} + \frac{2}{3}H_n^{(2)} + \frac{2}{3}H_j^{(2)} \\
&+ \frac{1}{2}\frac{j}{n-j+1} + \frac{1}{6}\frac{j(j-1)}{n-1} - \frac{1}{6}\frac{j(j-2)}{n} + \frac{1}{6} \quad \text{für } 1 \le j \le n\,.
\end{aligned}} \qquad (1.183)$$

1.6.2 Asymptotische Betrachtungen und weitere Folgerungen

Aus Gleichung (1.177) erhalten wir mit den Beziehungen (1.92), (1.104), (1.109) und (1.113) sofort die asymptotischen Entwicklungen der Anzahl der Nachkommen des Knotens j.

Korollar 1.6.4. *Der Erwartungswert $A_{n,j}$ der Anzahl der Nachkommen des Knotens mit Schlüssel j in einem p-tree der Größe n, das heißt also in einem Objekt der Größe n aus der Familie $\mathcal{A}$, ist asymptotisch für $j = \rho n$ und $0 < \rho < 1$ gegeben durch*

$$\boxed{A_{n,j} = \frac{1}{3}\log^2 n + \left(\frac{1}{3}\rho^2 - \frac{1}{3}\rho - \frac{1}{6} + \frac{2}{3}\gamma + \frac{2}{3}\log\rho - \frac{1}{3}\log(1-\rho)\right)\log n + \mathcal{O}(1)\,.} \tag{1.184}$$

Korollar 1.6.5. *Der Erwartungswert $A_{n,j}$ der Anzahl der Nachkommen des Knotens mit Schlüssel j in einem p-tree der Größe n, das heißt also in einem Objekt der Größe n aus der Familie $\mathcal{A}$, ist asymptotisch für fixes j gegeben durch*

$$\boxed{A_{n,j} = \frac{1}{3}H_j^2 - \frac{1}{6}H_j + \frac{2}{3}H_j^{(2)} + \frac{1}{6} + \mathcal{O}(1/n)\,.} \tag{1.185}$$

Korollar 1.6.6. *Der Erwartungswert $A_{n,j}$ der Anzahl der Nachkommen des Knotens mit Schlüssel j in einem p-tree der Größe n, das heißt also in einem Objekt der Größe n aus der Familie $\mathcal{A}$, ist asymptotisch für fixes $l = n - j$ gegeben durch*

$$\boxed{A_{n,j} = \frac{n}{l+1} + H_l\log n + \mathcal{O}(1)\,.} \tag{1.186}$$

Weiters ergibt sich für das arithmetische Mittel $A_n = \frac{1}{n}\sum_{j=1}^{n} A_{n,j}$ der Anzahl der Nachkommen eines zufällig gewählten Knotens unter Verwendung der Beziehungen (1.85) und (1.117) durch Aufsummieren von (1.177)

Korollar 1.6.7. *Der Erwartungswert A_n der Anzahl der Nachkommen eines zufällig gewählten Knotens in einem p-tree der Größe n, das heißt also in einem Objekt der Größe n aus der Familie $\mathcal{A}$, ist gegeben durch*

$$\boxed{\begin{aligned} &A_n = \frac{n+1}{3n}H_n^2 + \frac{2(2n+5)}{9n}H_n - \frac{n+1}{3n}H_n^{(2)} - \frac{13n+1}{27n} \quad \text{für } n \geq 2\,, \\ &A_1 = 1\,. \end{aligned}} \tag{1.187}$$

Das Auffällige am letzterhaltenen Resultat (1.187) des Erwartungswertes der Anzahl der Nachkommen eines zufällig gewählten Knotens ist die Gleichheit mit dem Ergebnis (1.118) des Erwartungswertes der Höhe eines zufällig gewählten Knotens.

Daß dies kein Zufall ist sondern eine zwangsläufige Folgerung, beweist

Hilfssatz 1.6.8. *In einem beliebigen Wurzelbaum T der Größe n ist der Erwartungswert $E_T^{[H]}$ der Höhe eines zufällig gewählten Knotens gleich dem Erwartungswert $E_T^{[D]}$ der Anzahl der Nachfolger eines zufällig gewählten Knotens.*

Beweis.

$$
\begin{aligned}
E_T^{[H]} &= \frac{1}{n}\sum_{j=1}^{n} (\text{Höhe des Knotens } j) \\
&= \frac{1}{n}\sum_{j=1}^{n} [\text{Anzahl der Knoten auf dem Weg von der Wurzel zum Knoten } j] \\
&= \frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n} \left[\begin{array}{l} \text{Knoten } k \text{ ist ein Vorgänger von } j \\ \text{auf dem Weg von der Wurzel zum Knoten } j \end{array}\right] \\
&= \frac{1}{n}\sum_{k=1}^{n}\sum_{j=1}^{n} [\text{Knoten } j \text{ ist ein Nachkomme von } k] \\
&= \frac{1}{n}\sum_{k=1}^{n} (\text{Anzahl der Nachkommen von } k) = E_T^{[D]}
\end{aligned}
$$

□

Dabei sollen die im Beweis verwendeten in eckige Klammern eingeschlossenen Aussagen den Wert 1 bekommen, wenn die Aussage wahr ist und sonst den Wert 0. Ist also A eine mathematische Aussage, dann gilt

$$
[A] = \begin{cases} 1 & \text{wenn } A \text{ wahr ist}\,, \\ 0 & \text{sonst}\,. \end{cases}
$$

1.7 Die Anzahl der Schlüsselvergleiche beim Einfügen

1.7.1 Die Erwartungswerte

Wie schon in Abschnitt 1.1 besprochen, bedient man sich beim Einfügen eines neuen Elements in einen p-tree des rekursiven Algorithmus `Insert`. Wir wollen nun das durchschnittliche Verhalten dieses Algorithmus studieren. Es interessiert uns dabei besonders, wie viele Schlüsselvergleiche beim Einfügen eines speziellen Elementes in einen zufällig gewachsenen p-tree gemacht werden.

Gehen wir nun also von durch Zufallspermutationen erzeugten Objekten der Familien $\mathcal{A}$, $\mathcal{B}$ und $\mathcal{C}$ aus. Dann bezeichnen wir hier mit $A_{n,j}^{[I]}$, $B_{n,j}^{[I]}$ und $C_{n,j}^{[I]}$ die Erwartungswerte der Anzahl an Schlüsselvergleichen, die beim Einfügen eines Elements $j+\frac{1}{2}$ mit $j \in \mathbb{Z}$ und $0 \le j \le n$ in ein entsprechendes Objekt der Größe n aus den Familien $\mathcal{A}$, $\mathcal{B}$ bzw. $\mathcal{C}$ gemacht werden.

Für die Erstellung einer Rekursionsformel der gesuchten Größen, kann man die Objektfamilien wieder aufgrund des ersten Elements der Zufallspermutation aufteilen und erhält so unter Berücksichtigung der Anfangswerte folgende Beziehungen:

$$
A_{n,j}^{[I]} = \frac{1}{n}\left(\sum_{k=1}^{j} C_{n-k,j-k}^{[I]} + \sum_{k=j+1}^{n}\left(B_{k-1,j}^{[I]} + H_{n-k}\right)\right) \quad \text{für } n \ge 1 \text{ und } 0 \le j \le n\,, \tag{1.188a}
$$

$$B_{n,j}^{[I]} = \frac{1}{n}\left(\sum_{k=1}^{j}\left(A_{n-k,j-k}^{[I]}+2\right) + \sum_{k=j+1}^{n}\left(B_{k-1,j}^{[I]}+1\right)\right) \quad \text{für } n \geq 1 \text{ und } 0 \leq j \leq n\,, \tag{1.188b}$$

$$C_{n,j}^{[I]} = \frac{1}{n}\left(\sum_{k=1}^{j}C_{n-k,j-k}^{[I]} + \sum_{k=j+1}^{n}\left(A_{k-1,j}^{[I]}+H_{n-k}+2\right)\right) \quad \text{für } n \geq 1 \text{ und } 0 \leq j \leq n\,, \tag{1.188c}$$

$$A_{0,0}^{[I]} = 0\,, \quad B_{0,0}^{[I]} = 1\,, \quad C_{0,0}^{[I]} = 1\,.$$

Zum Lösen dieser Rekursion wendet man das nun schon mehrmals verwendete Verfahren an. Zunächst definiert man sich also durch

$$A^{[I]}(z,u) = \sum_{j\geq 0}\sum_{n\geq j} A_{n,j}^{[I]} z^n u^j\,, \quad B^{[I]}(z,u) = \sum_{j\geq 0}\sum_{n\geq j} B_{n,j}^{[I]} z^n u^j\,, \quad C^{[I]}(z,u) = \sum_{j\geq 1}\sum_{n\geq j} C_{n,j}^{[I]} z^n u^j$$

die den gesuchten Erwartungswerten entsprechenden bivariaten erzeugenden Funktionen in z und u.

Multipliziert man nun (1.188) mit $nz^{n-1}u^j$ und summiert die entstehenden Gleichungen über alle $n \geq j \geq 0$ auf, entsteht daraus folgendes System von Differentialgleichungen:

$$\frac{\partial}{\partial z}A^{[I]}(z,u) = \frac{u}{1-uz}C^{[I]}(z,u) + \frac{1}{(1-z)^2(1-uz)}\log\left(\frac{1}{1-z}\right) + \frac{1}{1-z}B^{[I]}(z,u)\,, \tag{1.189a}$$

$$\frac{\partial}{\partial z}B^{[I]}(z,u) = \frac{u}{1-uz}A^{[I]}(z,u) + \frac{2u}{(1-z)(1-uz)^2} + \frac{1}{(1-z)^2(1-uz)} + \frac{1}{1-z}B^{[I]}(z,u)\,, \tag{1.189b}$$

$$\frac{\partial}{\partial z}C^{[I]}(z,u) = \frac{u}{1-uz}C^{[I]}(z,u) + \frac{1}{(1-z)^2(1-uz)}\log\left(\frac{1}{1-z}\right) + \frac{2}{(1-z)^2(1-uz)} + \frac{1}{1-z}A^{[I]}(z,u)\,. \tag{1.189c}$$

Bezeichnet man die auftretenden Störfunktionen mit

$$f_1(z,u) = \frac{1}{(1-z)^2(1-uz)}\log\left(\frac{1}{1-z}\right)\,,$$
$$f_2(z,u) = \frac{2u}{(1-z)(1-uz)^2} + \frac{1}{(1-z)^2(1-uz)}\,,$$
$$f_3(z,u) = \frac{1}{(1-z)^2(1-uz)}\log\left(\frac{1}{1-z}\right) + \frac{2}{(1-z)^2(1-uz)}\,,$$

so lautet das zu lösende Differentialgleichungssystem

$$\frac{\partial}{\partial z}A^{[I]}(z,u) = \frac{u}{1-uz}C^{[I]}(z,u) + \frac{1}{1-z}B^{[I]}(z,u) + f_1(z,u)\,, \tag{1.190a}$$

$$\frac{\partial}{\partial z}B^{[I]}(z,u) = \frac{u}{1-uz}A^{[I]}(z,u) + \frac{1}{1-z}B^{[I]}(z,u) + f_2(z,u)\,, \tag{1.190b}$$

$$\frac{\partial}{\partial z}C^{[I]}(z,u) = \frac{u}{1-uz}C^{[I]}(z,u) + \frac{1}{1-z}A^{[I]}(z,u) + f_3(z,u)\,. \tag{1.190c}$$

Wie gehabt kann man dieses System von Differentialgleichungen transformieren in eine lineare Differentialgleichung dritter Ordnung für $A^{[I]}(z,u)$. Diese lautet

$$\begin{aligned}
&\frac{2uz-u-1}{(1-z)(1-u)}A^{[I]}(z,u) - 2\frac{1-uz}{1-u}\frac{\partial}{\partial z}A^{[I]}(z,u) \\
&- \frac{3}{2}\frac{(1-uz)(2uz-u-1)}{u(1-u)}\frac{\partial^2}{\partial z^2}A^{[I]}(z,u) - \frac{1}{2}\frac{(1-uz)^2(1-z)}{u(1-u)}\frac{\partial^3}{\partial z^3}A^{[I]}(z,u) \\
&= f_3(z,u) + \frac{1}{1-uz}\frac{\partial}{\partial z}\left[\frac{(1-uz)^3}{2u(u-1)}f_2(z,u)\right. \\
&\quad \left. + \frac{(1-uz)^3}{2u(u-1)(1-z)}\frac{\partial}{\partial z}\left((1-z)^2 f_1(z,u)\right) + \frac{(1-uz)^2(1-z)}{2(u-1)}f_3(z,u)\right]\,.
\end{aligned}$$

Setzt man für $f_1(z,u)$, $f_2(z,u)$ und $f_3(z,u)$ wieder ein und berücksichtigt die Anfangswerte $A^{[I]}_{0,0}=0$, $A^{[I]}_{1,0}=1$, $A^{[I]}_{1,1}=1$, $A^{[I]}_{2,0}=2$, $A^{[I]}_{2,1}=2$ und $A^{[I]}_{2,2}=1$, so bekommt man

$$\begin{aligned}
&\frac{2uz-u-1}{(1-z)(1-u)}A^{[I]}(z,u) - 2\frac{1-uz}{1-u}\frac{\partial}{\partial z}A^{[I]}(z,u) \\
&- \frac{3}{2}\frac{(1-uz)(2uz-u-1)}{u(1-u)}\frac{\partial^2}{\partial z^2}A^{[I]}(z,u) - \frac{1}{2}\frac{(1-uz)^2(1-z)}{u(1-u)}\frac{\partial^3}{\partial z^3}A^{[I]}(z,u) \\
&= \frac{uz+u-2}{(1-z)^3(1-u)u}\,, \\
&A^{[I]}(0,u)=0\,,\quad \left.\frac{\partial}{\partial z}A^{[I]}(z,u)\right|_{z=0} = 1+u\,,\quad \left.\frac{\partial^2}{\partial z^2}A^{[I]}(z,u)\right|_{z=0} = 2\left(2+2u+u^2\right)\,.
\end{aligned} \tag{1.191}$$

Mit den Transformationen

$$\begin{aligned}
A^{[I]}(z,u) &= \frac{E^{[I]}(z,u)}{(1-z)(1-uz)}\,, \\
F^{[I]}(z,u) &= \frac{\partial}{\partial z}E^{[I]}(z,u)\,, \\
F^{[I]}(z,u) &= (1-z)(1-uz)G^{[I]}(z,u)\,, \\
H^{[I]}(z,u) &= \frac{\partial}{\partial z}G^{[I]}(z,u)\,,
\end{aligned}$$

ergibt sich daraus die lineare Differentialgleichung erster Ordnung

$$\frac{\partial}{\partial z}H^{[I]}(z,u) - 2\frac{1+u-2uz}{(1-z)(1-uz)}H^{[I]}(z,u) = -2\frac{uz+u-2}{(1-z)^4(1-uz)^2}\,. \tag{1.192}$$

Wie man durch Variation der Konstanten findet, ist eine Partikulärlösung dieser Differentialgleichung gegeben durch

$$H_p^{[I]}(z,u) = 2\frac{u}{(1-z)^2(1-uz)^2}\log\left(\frac{1}{1-z}\right) - 4\frac{u-1}{(1-z)^3(1-uz)^2}\,. \tag{1.193}$$

Durch Rücksubstitutionen ergibt sich daraus folgende Partikulärlösung für (1.191):

$$A_p^{[I]}(z,u) = \frac{1}{(1-z)(1-uz)} \int_0^z \left[(1-x)(1-ux) \int_0^x H_p^{[I]}(t,u)\,dt\right] dx\,. \tag{1.194}$$

Für ein leichteres Ablesen der Koeffizienten vereinfacht man durch partielle Integration diesen Ausdruck weiter zu

$$\begin{aligned} A_p^{[I]}(z,u) &= \left(\frac{2}{3}\frac{(2uz-u-1)u^2}{(1-u)^3} + \frac{2}{3}\frac{u}{(1-z)(1-u)}\right) \int_0^z \frac{1}{1-ut} \log\left(\frac{1}{1-t}\right) dt \\ &+ \left(-\frac{1}{3}\frac{(2uz-u-1)u}{(1-u)^3} - \frac{1}{3}\frac{u}{(1-uz)(1-u)}\right) \log^2\left(\frac{1}{1-z}\right) \\ &+ \left(\frac{1}{3}\frac{(10uz-u-9)u}{(1-u)^3} - \frac{1}{3}\frac{u}{(1-uz)(1-u)} + 2\frac{1}{(1-z)(1-u)}\right) \log\left(\frac{1}{1-z}\right) \\ &+ \left(-\frac{5}{3}\frac{(2uz-u-1)u}{(1-u)^3} - \frac{5}{3}\frac{1}{(1-z)(1-u)}\right) \log\left(\frac{1}{1-uz}\right) \\ &+ \frac{14}{3}\frac{u}{(1-u)^2} + 2\frac{u}{(1-uz)(1-u)} - \frac{5}{3}\frac{1}{(1-z)(1-u)} \end{aligned} \tag{1.195}$$

und schließlich zu

$$\begin{aligned} A_p^{[I]}(z,u) &= \left(\frac{2}{3}uz - \frac{1}{3}u - \frac{1}{3} + \frac{1}{3}\frac{(u-1)^2}{(1-z)u}\right) \int_0^z \frac{1}{(1-t)^2(1-ut)^2} \log\left(\frac{1}{1-t}\right) dt \\ &+ \left(-\frac{1}{3}\frac{u}{(1-uz)(1-u)} + \frac{1}{3}\frac{1}{(1-z)(1-u)}\right) \log^2\left(\frac{1}{1-z}\right) \\ &+ \left[2\frac{(2uz-u-1)u}{(1-u)^3} + \frac{7}{3}\frac{1}{(1-z)(1-u)} - \frac{1}{3}\frac{1}{(1-z)^2u} \right. \\ &\qquad \left. - \frac{1}{3}\frac{u}{(1-uz)(1-u)}\right] \log\left(\frac{1}{1-z}\right) \\ &+ \left(-2\frac{(2uz-u-1)u}{(1-u)^3} - 2\frac{1}{(1-z)(1-u)}\right) \log\left(\frac{1}{1-uz}\right) \\ &+ 4\frac{u}{(1-u)^2} - 2\frac{1}{(1-z)(1-u)} + \frac{1}{3}\frac{1}{(1-z)^2u} + 2\frac{u}{(1-uz)(1-u)}\,. \end{aligned} \tag{1.196}$$

Die Lösung der Gleichung (1.191) entsprechenden homogenen Differentialgleichung lautet wieder

$$\begin{aligned} A_h^{[I]}(z,u) &= \frac{c_1(u)}{(1-z)(1-uz)} + c_2(u)\frac{z(6-3uz-3z+2uz^2)}{(1-z)(1-uz)} \\ &+ \frac{c_3(u)}{(1-z)(1-uz)} \int_0^z \left[(1-x)(1-ux) \int_0^x \frac{1}{(1-t)^2(1-ut)^2}\,dt\right] dx\,. \end{aligned}$$

Durch Anpassen der allgemeinen Lösung der Differentialgleichung (1.191) an die Anfangsbedingungen, bekommen wir für die nur von u abhängigen Funktionen $c_1(u)$, $c_2(u)$ und $c_3(u)$ folgende Werte:

$$\begin{aligned} c_1(u) &= -\frac{1}{18}\frac{5u^4-10u^3+126u^2-14u+1}{(1-u)^2u}\,, \quad c_2(u) = \frac{2}{3}\frac{u}{(1-u)}\,, \\ c_3(u) &= -\frac{1}{18}\frac{u^2+6u-1}{(1-u)^3u}\,. \end{aligned}$$

Diese Ergebnisse faßt man zusammen und folgert daraus als Lösung der doppelt erzeugenden Funktion $A^{[I]}(z,u)$ für die Erwartungswerte $A_{n,j}^{[I]}$ der Anzahl der Key-Comparisons beim Einfügen eines Knotens mit Schlüssel $j+\frac{1}{2}$ mit der Prozedur `Insert` in einen p-tree der Größe n

$$\begin{aligned}
A^{[I]}(z,u) &= \left(\frac{2}{3}uz - \frac{1}{3}u - \frac{1}{3} + \frac{1}{3}\frac{(u-1)^2}{(1-z)u}\right)\int_0^z \frac{1}{(1-t)^2(1-ut)^2}\log\left(\frac{1}{1-t}\right)dt \\
&\quad + \left(-\frac{1}{3}\frac{u}{(1-uz)(1-u)} + \frac{1}{3}\frac{1}{(1-z)(1-u)}\right)\log^2\left(\frac{1}{1-z}\right) \\
&\quad + \left[\frac{1}{3}\frac{(u+1)(2uz-u-1)}{(1-u)^2} + \frac{7}{3}\frac{1}{(1-z)(1-u)} - \frac{1}{3}\frac{1}{(1-z)^2u}\right. \\
&\qquad \left. - \frac{1}{3}\frac{u^2+7u-1}{(1-uz)(1-u)}\right]\log\left(\frac{1}{1-z}\right) \\
&\quad + \left(-\frac{1}{3}\frac{(u+1)(2uz-u-1)}{(1-u)^2} - \frac{1}{3}\frac{u+1}{(1-z)u}\right)\log\left(\frac{1}{1-uz}\right) \\
&\quad + \frac{4}{3}\frac{uz}{(1-u)} + \frac{1}{3}\frac{2u^2-2u-1}{(1-z)(1-u)u} + \frac{1}{3}\frac{1}{(1-z)^2u} - \frac{1}{3}\frac{2u-3}{(1-uz)(1-u)}\,.
\end{aligned} \tag{1.197}$$

Für das Ablesen der Koeffizienten von $A^{[I]}(z,u)$ verwendet man wieder die Beziehungen (1.19) und (1.62) aus den vorangegangenen Abschnitten und bekommt daraus

Satz 1.7.1. *Der Erwartungswert $A_{n,j}^{[I]}$ für die Anzahl der Schlüsselvergleiche, die beim Einfügen eines Elements mit Schlüssel $j+\frac{1}{2}$ mit der Prozedur* `Insert` *in einen p-tree der Größe n, das heißt also in ein Objekt der Größe n aus der Familie $\mathcal{A}$, gemacht werden, ist gegeben durch*

$$\boxed{\begin{aligned}
A_{n,j}^{[I]} &= -\frac{2}{3}\sum_{k=j+1}^{n}\frac{H_{k-j}}{k} + \frac{1}{3}H_{n-j}^2 + \left(-\frac{1}{3} - \frac{2}{3}\frac{1}{j}\right)H_n \\
&\quad + \left(\frac{7}{3} + \frac{2}{3j} + \frac{2}{3}\frac{j^2}{n} - \frac{2}{3}\frac{j(j-1)}{n-1}\right)H_{n-j} \\
&\quad + \left(\frac{1}{3} + \frac{2}{3j}\right)H_j - \frac{1}{3}H_{n-j}^{(2)} - \frac{1}{3}\frac{j(2j+5)}{n} + \frac{2}{3}\frac{j^2+j-1}{n-1} \\
&\quad + \frac{1}{3}\frac{1}{n-j+1} + \frac{1}{3}\frac{1}{n-j} - \frac{1}{3}\frac{1}{j+1} - \frac{1}{3j} \quad \text{für } 1 \le j \le n-1\,,
\end{aligned}}$$

$$\boxed{\begin{aligned}
&A_{n,n}^{[I]} = 1 \quad \text{für } n \ge 1\,, \\
&A_{0,0}^{[I]} = 0\,,
\end{aligned}}$$

$$\boxed{A_{n,0}^{[I]} = 2H_n - 1 \quad \text{für } n \ge 1\,.} \tag{1.198}$$

Um nun zu den bivariaten erzeugenden Funktionen $C^{[I]}(z,u)$ zu gelangen, gehen wir wie in den vorigen Abschnitten beschrieben vor, und erhalten so

$$C^{[I]}(z,u) = \left(-\frac{1}{3}uz - \frac{1}{3}u + \frac{2}{3} + \frac{1}{3}\frac{(u-1)^2}{1-z}\right)\int_0^z \frac{1}{(1-t)^2(1-ut)^2}\log\left(\frac{1}{1-t}\right)dt$$

$$
\begin{aligned}
&+\left(-\frac{1}{3}\frac{u}{(1-uz)(1-u)}+\frac{1}{3}\frac{1}{(1-z)(1-u)}\right)\log^2\left(\frac{1}{1-z}\right)\\
&+\left[-\frac{1}{3}\frac{(u+1)(uz+u-2)}{(1-u)^2}-\frac{1}{3}\frac{1}{(1-z)^2u}+\frac{7}{3}\frac{1}{(1-z)(1-u)}\right.\\
&\qquad\left.-\frac{1}{3}\frac{u^2+7u-1}{(1-uz)(1-u)}\right]\log\left(\frac{1}{1-z}\right)\\
&+\left(\frac{1}{3}\frac{(u+1)(uz+u-2)}{(1-u)^2}-\frac{1}{3}\frac{u+1}{(1-z)u}\right)\log\left(\frac{1}{1-uz}\right)\\
&-\frac{1}{3}\frac{(2uz+3u-3)}{1-u}-\frac{1}{3}\frac{2u-3}{(1-uz)(1-u)}+\frac{1}{3}\frac{2u^2-2u-1}{(1-z)(1-u)u}+\frac{1}{3}\frac{1}{(1-z)^2u}\,.
\end{aligned}
\tag{1.199}
$$

Das Ablesen der Koeffizienten führt unmittelbar zu

Hilfssatz 1.7.2. *Der Erwartungswert $C_{n,j}^{[I]}$ für die Anzahl der Schlüsselvergleiche, die beim Einfügen eines Elements mit Schlüssel $j+\frac{1}{2}$ mit der Prozedur* `Insert` *in ein Objekt der Größe n aus der Familie $\mathcal{C}$ gemacht werden, ist gegeben durch*

$$
\boxed{
\begin{aligned}
C_{n,j}^{[I]}=&-\frac{2}{3}\sum_{k=j+1}^{n}\frac{H_{k-j}}{k}+\frac{1}{3}H_{n-j}^2+\left(-\frac{1}{3}-\frac{2}{3j}\right)H_n\\
&+\left(\frac{10}{3}+\frac{2}{3j}-\frac{1}{3}\frac{j(j+3)}{n}+\frac{1}{3}\frac{j(j-1)}{n-1}\right)H_{n-j}\\
&+\left(\frac{1}{3}+\frac{2}{3}\frac{1}{j}\right)H_j-\frac{1}{3}H_{n-j}^{(2)}+\frac{1}{3}\frac{(j+1)(j+3)}{n}-\frac{1}{3}\frac{j^2+j-1}{n-1}\\
&+\frac{1}{3}\frac{1}{n-j+1}+\frac{1}{3}\frac{1}{n-j}-\frac{1}{3}\frac{1}{j+1}-\frac{1}{3j}-1\quad\text{für } 1\le j\le n-1\,,
\end{aligned}
}
$$

$$
\boxed{C_{n,n}^{[I]}=1\quad\text{für } n\ge 0\,,}
$$

$$
\boxed{C_{n,0}^{[I]}=3H_n-2+\frac{1}{n}\quad\text{für } n\ge 1\,.}
\tag{1.200}
$$

Weiters ergibt sich für $B^{[I]}(z,u)$

$$
\begin{aligned}
B^{[I]}(z,u)=&\left(-\frac{1}{3}uz+\frac{2}{3}u-\frac{1}{3}+\frac{1}{3}\frac{(u-1)^2}{(1-z)u}\right)\int_0^z\frac{1}{(1-t)^2(1-ut)^2}\log\left(\frac{1}{1-t}\right)dt\\
&+\left(-\frac{1}{3}\frac{u}{(1-uz)(1-u)}+\frac{1}{3}\frac{1}{(1-z)(1-u)}\right)\log^2\left(\frac{1}{1-z}\right)\\
&+\left[-\frac{1}{3}\frac{(u+1)(uz-2u+1)}{(1-u)^2}+\frac{4}{3}\frac{1}{(1-z)(1-u)}-\frac{1}{3}\frac{1}{(1-z)^2u}\right.\\
&\qquad\left.-\frac{1}{3}\frac{u^2+4u-1}{(1-uz)(1-u)}\right]\log\left(\frac{1}{1-z}\right)\\
&+\left(\frac{1}{3}\frac{(u+1)(uz-2u+1)}{(1-u)^2}-\frac{1}{3}\frac{u+1}{(1-z)u}\right)\log\left(\frac{1}{1-uz}\right)\\
&-\frac{1}{3}\frac{2uz-3u+3}{1-u}-\frac{1}{3}\frac{8u-3}{(1-uz)(1-u)}+\frac{1}{3}\frac{2u^2+4u-1}{(1-z)(1-u)u}+\frac{1}{3}\frac{1}{(1-z)^2u}\,.
\end{aligned}
\tag{1.201}
$$

Das Ablesen der Koeffizienten ergibt schließlich

Hilfssatz 1.7.3. *Der Erwartungswert $B^{[I]}_{n,j}$ für die Anzahl der Schlüsselvergleiche, die beim Einfügen eines Elements mit Schlüssel $j+\frac{1}{2}$ mit der Prozedur* `Insert` *in ein Objekt der Größe n aus der Familie $\mathcal{B}$ gemacht werden, ist gegeben durch*

$$\boxed{\begin{aligned}B^{[I]}_{n,j} &= -\frac{2}{3}\sum_{k=j+1}^{n}\frac{H_{k-j}}{k}+\frac{1}{3}H^2_{n-j}+\left(-\frac{1}{3}-\frac{2}{3j}\right)H_n\\ &\quad+\left(\frac{4}{3}+\frac{2}{3j}-\frac{1}{3}\frac{j(j-3)}{n}+\frac{1}{3}\frac{j(j-1)}{n-1}\right)H_{n-j}\\ &\quad+\left(\frac{1}{3}+\frac{2}{3j}\right)H_j-\frac{1}{3}H^{(2)}_{n-j}+\frac{1}{3}\frac{j^2+j-3}{n}-\frac{1}{3}\frac{j^2+j-1}{n-1}\\ &\quad+\frac{1}{3}\frac{1}{n-j+1}+\frac{1}{3}\frac{1}{n-j}-\frac{1}{3}\frac{1}{j+1}-\frac{1}{3j}+2 \quad \text{für } 1\le j\le n-1\,,\end{aligned}}$$

$$\boxed{\begin{aligned}B^{[I]}_{n,n} &= 3-\frac{1}{n} \quad \text{für } n\ge 1\,,\\ B^{[I]}_{0,0} &= 1\,,\end{aligned}}$$

$$\boxed{B^{[I]}_{n,0} = H_n+1 \quad \text{für } n\ge 1\,.} \tag{1.202}$$

1.7.2 Asymptotische Betrachtungen und weitere Folgerungen

Mit den Beziehungen (1.92), (1.105), (1.110) und (1.112) erhalten wir aus Gleichung (1.198) die asymptotischen Entwicklungen der Anzahl der Schlüsselvergleiche beim Einfügen des Knotens $j+\frac{1}{2}$.

Korollar 1.7.4. *Der Erwartungswert $A^{[I]}_{n,j}$ für die Anzahl der Schlüsselvergleiche, die beim Einfügen eines Elements mit Schlüssel $j+\frac{1}{2}$ mit der Prozedur* `Insert` *in einen p-tree der Größe n, das heißt also in ein Objekt der Größe n aus der Familie $\mathcal{A}$, gemacht werden, ist asymptotisch für $j=\rho n$ und $0<\rho<1$ gegeben durch*

$$\boxed{A^{[I]}_{n,j} = \frac{1}{3}\log^2 n+\left(-\frac{2}{3}\rho^2+\frac{2}{3}\rho+\frac{2}{3}\gamma+\frac{7}{3}+\frac{2}{3}\log\rho+\frac{2}{3}\log(1-\rho)\right)+\mathcal{O}(1)\,.} \tag{1.203}$$

Korollar 1.7.5. *Der Erwartungswert $A^{[I]}_{n,j}$ für die Anzahl der Schlüsselvergleiche, die beim Einfügen eines Elements mit Schlüssel $j+\frac{1}{2}$ mit der Prozedur* `Insert` *in einen p-tree der Größe n, das heißt also in ein Objekt der Größe n aus der Familie $\mathcal{A}$, gemacht werden, ist asymptotisch für fixes j gegeben durch*

$$\boxed{A^{[I]}_{n,j} = 2\log n+\mathcal{O}(1)\,.} \tag{1.204}$$

Korollar 1.7.6. *Der Erwartungswert $A^{[I]}_{n,j}$ für die Anzahl der Schlüsselvergleiche, die beim Einfügen eines Elements mit Schlüssel $j+\frac{1}{2}$ mit der Prozedur* `Insert` *in einen p-tree der*

Größe n, das heißt also in ein Objekt der Größe n aus der Familie $\mathcal{A}$, gemacht werden, ist asymptotisch für fixes $l = n - j$ gegeben durch

$$\boxed{A_{n,j}^{[I]} = \frac{1}{3}H_l^2 + \frac{7}{3}H_l - \frac{1}{3} - \frac{1}{3}H_l^{(2)} + \frac{1}{3l} + \frac{1}{3(l+1)} + \mathcal{O}(1/n)\,.} \tag{1.205}$$

Weiters ergibt sich für das arithmetische Mittel $A_n^{[I]} = \frac{1}{n+1}\sum_{j=0}^{n} A_{n,j}^{[I]}$ der Anzahl an Schlüsselvergleichen beim Einfügen eines zufällig gewählten Knoten unter Verwendung der Beziehungen (1.85) und (1.117) durch Aufsummieren von (1.198)

Korollar 1.7.7. *Der Erwartungswert $A_n^{[I]}$ für die Anzahl der Schlüsselvergleiche, die beim Einfügen eines zufällig gewählten Elements mit der Prozedur* `Insert` *in einen p-tree der Größe n, das heißt also in ein Objekt der Größe n aus der Familie $\mathcal{A}$, gemacht werden, ist gegeben durch*

$$\boxed{\begin{aligned} &A_n^{[I]} = \frac{1}{3}H_{n+1}^2 + \frac{10}{9}H_{n+1} - \frac{1}{3}H_{n+1}^{(2)} - \frac{28}{27} \quad \text{für } n \geq 2\,,\\ &A_1^{[I]} = 1\,. \end{aligned}} \tag{1.206}$$

Dieses Resultat wurde bereits in [17], allerdings auf gänzlich andere Art und Weise, erhalten.

1.8 Die Anzahl der Knoten mit 0, 1 bzw. 2 Kindern

1.8.1 Die Verteilungen der Anzahl der Blätter

Zerlegt man Objekte aus den Familien $\mathcal{A}$, $\mathcal{B}$ und $\mathcal{C}$ der Größe n aufgrund des ersten Elements k einer Zufallspermutation mit n Elementen, so ergeben sich für die Wahrscheinlichkeiten $A_{n,m}$, $B_{n,m}$ bzw. $C_{n,m}$, daß ein Objekt der entsprechenden Familie genau m Blätter besitzt, die Rekursionen

$$A_{n,m} = \frac{1}{n} \cdot \sum_{k=1}^{n}\sum_{i=0}^{m} C_{n-k,i} \cdot B_{k-1,m-i} \quad \text{für } n \geq 1\,, \tag{1.207a}$$

$$B_{n,m} = \frac{1}{n} \cdot \sum_{k=1}^{n}\sum_{i=0}^{m} A_{n-k,i} \cdot B_{k-1,m-i} \quad \text{für } n \geq 1\,, \tag{1.207b}$$

$$C_{n,m} = \frac{1}{n} \cdot \sum_{k=1}^{n}\sum_{i=0}^{m} C_{n-k,i} \cdot A_{k-1,m-i} \quad \text{für } n \geq 1\,, \tag{1.207c}$$

$$A_{0,m} = \delta_{0,m}\,, \quad B_{0,m} = \delta_{1,m} \text{ und } C_{0,m} = \delta_{0,m}\,.$$

Definiert man nun mittels

$$A(z,v) = \sum_{n\geq 0}\sum_{m\geq 0} A_{n,m} z^n v^m\,, \quad B(z,v) = \sum_{n\geq 0}\sum_{m\geq 0} B_{n,m} z^n v^m\,, \quad C(z,v) = \sum_{n\geq 0}\sum_{m\geq 0} C_{n,m} z^n v^m$$

erzeugende Funktionen in zwei Variablen, so kommt man durch Multiplizieren obiger Gleichungen (1.207) mit $nz^{n-1}v^m$ und anschließendes Aufsummieren über alle $n \geq 1$ und $m \geq 0$ zu folgendem System von Differentialgleichungen

$$\frac{\partial}{\partial z}A(z,v) = B(z,v)\,C(z,v)\,, \tag{1.208a}$$

$$\frac{\partial}{\partial z}B(z,v) = A(z,v)\,B(z,v)\,, \tag{1.208b}$$

$$\frac{\partial}{\partial z}C(z,v) = A(z,v)\,C(z,v)\,, \tag{1.208c}$$

$$A(0,v) = 1\,, \quad B(0,v) = v \text{ und } C(0,v) = 1\,.$$

Dieses Differentialgleichungssystem versuchen wir nun umzuformen zu einer einzigen Differentialgleichung für $A(z,v)$.

Dazu leiten wir zunächst Gleichung (1.208a) nach z ab und bekommen so

$$\frac{\partial^2}{\partial z^2}A(z,v) = \frac{\partial}{\partial z}B(z,v)\,C(z,v) + B(z,v)\,\frac{\partial}{\partial z}C(z,v)\,. \tag{1.209}$$

Ersetzen wir nun $\frac{\partial}{\partial z}B(z,v)$ bzw. $\frac{\partial}{\partial z}C(z,v)$ durch die entsprechenden rechten Seiten obiger Gleichungen (1.208b) und (1.208c), so ergibt sich weiter

$$\frac{\partial^2}{\partial z^2}A(z,v) = 2A(z,v)\,B(z,v)\,C(z,v)\,. \tag{1.210}$$

Wenn man nun den Faktor $B(z,v)\,C(z,v)$ wieder durch $\frac{\partial}{\partial z}A(z,v)$ substituiert, erhält man unter Angabe der Anfangswerte folgende nichtlineare Differentialgleichung zweiter Ordnung für $A(z,v)$:

$$\begin{aligned} &\frac{\partial^2}{\partial z^2}A(z,v) = 2\,A(z,v)\,\frac{\partial}{\partial z}A(z,v)\,, \\ &A(0,v) = 1\,, \quad \frac{\partial}{\partial z}A(z,v)\Big|_{z=0} = v\,. \end{aligned} \tag{1.211}$$

Da diese Differentialgleichung (1.211) von der Form $A'' = f(A, A')$ ist, also die Variable z nicht explizit vorkommt, definieren wir uns zur Lösung eine Funktion $q(A,v)$, wobei q gleich $\frac{\partial}{\partial z}A(z,v)$ als Funktion von A und v sein soll, mit anderen Worten, es gilt $\frac{\partial}{\partial z}A(z,v) = q(A,v)$ (siehe dazu beispielsweise [48]).

Nun leiten wir uns eine neue Differentialgleichung für q her. Es gilt nämlich

$$\frac{dq}{dA} = \frac{dq}{dz}\,\frac{dz}{dA} = \frac{dq}{dz}\,\frac{1}{\frac{dA}{dz}} = \frac{\partial^2}{\partial z^2}A\,\frac{1}{\frac{\partial}{\partial z}A} = 2\,A\,. \tag{1.212}$$

Die letzte Gleichung obiger Gleichungskette folgt durch Einsetzen von (1.211).

Diese Differentialgleichung

$$\frac{\partial}{\partial A}q(A,v) = 2\,A \tag{1.213}$$

besitzt nun die Lösung

$$q(A,v) = A(z,v)^2 + c_1(v) \tag{1.214}$$

mit einer nur von v abhängigen Funktion $c_1(v)$.

Da nun aber $q(A,v) = \frac{\partial}{\partial z}A(z,v)$ ist, bekommt man daraus zunächst folgende Differentialgleichung erster Ordnung für $A(z,v)$:

$$\begin{gathered} \frac{\partial}{\partial z}A(z,v) = A(z,v)^2 + c_1(v)\,, \\ A(0,v) = 1\,, \quad \frac{\partial}{\partial z}A(z,v)\Big|_{z=0} = v\,. \end{gathered} \tag{1.215}$$

Setzt man in obiger Gleichung (1.215) für $z = 0$ ein, kann man $c_1(v)$ bestimmen und erhält so

$$\frac{\partial}{\partial z}A(z,v) = A(z,v)^2 - (1-v)\,, \quad A(0,v) = 1\,. \tag{1.216}$$

Diese Differentialgleichung ist nun durch Trennung der Variablen A und z zu lösen. Es ergibt sich zunächst

$$\frac{dA}{A^2 - (1-v)} = dz\,, \tag{1.217}$$

welches man umformt zu

$$-\frac{1}{1-z}\frac{1}{2}\left(\frac{1}{1-\frac{A}{\sqrt{1-v}}} + \frac{1}{1+\frac{A}{\sqrt{1-v}}}\right) dA = dz \tag{1.218}$$

und schließlich zu

$$\frac{1}{2\sqrt{1-v}}\, d\log\left(\frac{1+\frac{A}{\sqrt{1-v}}}{1-\frac{A}{\sqrt{1-v}}}\right) = -dz\,. \tag{1.219}$$

Integration ergibt nun vorerst

$$\frac{1}{2\sqrt{1-v}}\log\left(\frac{1+\frac{A(z,v)}{\sqrt{1-v}}}{1-\frac{A(z,v)}{\sqrt{1-v}}}\right) = -z + c_2(v) \tag{1.220}$$

mit einer von v abhängigen Funktion $c_2(v)$.

Daraus ergibt sich nach einigen Umformungen

$$A(z,v) = \sqrt{1-v}\left(\frac{e^{2\sqrt{1-v}(-z+c_2(v))} - 1}{e^{2\sqrt{1-v}(-z+c_2(v))} + 1}\right)\,, \quad A(0,v) = 1\,. \tag{1.221}$$

Einsetzen für $z = 0$ in obige Gleichung (1.221) bestimmt $c_2(v)$ als

$$c_2(v) = \frac{1}{2\sqrt{1-v}}\log\left(\frac{1+\frac{1}{\sqrt{1-v}}}{1-\frac{1}{\sqrt{1-v}}}\right)\,. \tag{1.222}$$

Somit ergibt sich für $A(z,v)$

$$A(z,v)=\sqrt{1-v}\left(\frac{e^{-2\sqrt{1-v}\,z}\left(\frac{1+\frac{1}{\sqrt{1-v}}}{1-\frac{1}{\sqrt{1-v}}}\right)-1}{e^{-2\sqrt{1-v}\,z}\left(\frac{1+\frac{1}{\sqrt{1-v}}}{1-\frac{1}{\sqrt{1-v}}}\right)+1}\right), \tag{1.223}$$

woraus man nach etlichen Umformungen zu folgendem Satz kommt:

Satz 1.8.1. *Die bivariate erzeugende Funktion $A(z,v)$ der Wahrscheinlichkeiten $A_{n,m}$, daß ein durch Zufallspermutationen erzeugter p-tree der Größe n genau m Blätter hat, ist gegeben durch*

$$\boxed{A(z,v)=\sqrt{1-v}-\frac{2\,v\sqrt{1-v}}{v-\left(1+\sqrt{1-v}\right)^2 e^{-2\sqrt{1-v}\,z}}\,.} \tag{1.224}$$

Zur Ermittlung von $B(z,v)$ geht man von (1.208) aus und bekommt so die Differentialgleichung

$$\frac{\partial}{\partial z}\log B(z,v)=A(z,v)\,,\quad B(0,v)=v\,, \tag{1.225}$$

die man durch Trennung der Variablen löst.

Man findet zunächst

$$B(z,v)=c(v)e^{\int_0^z A(t,v)dt}\,, \tag{1.226}$$

woraus man durch Einsetzen des soeben gefundenen Wertes für $A(z,v)$ und Anpassen der von v abhängigen Funktion $c(v)$ an die Anfangswerte schon folgendes erhält:

Hilfssatz 1.8.2. *Die bivariate erzeugende Funktion $B(z,v)$ der Wahrscheinlichkeiten $B_{n,m}$, daß ein durch Zufallspermutationen erzeugtes Objekt aus der Familie $\mathcal{B}$ der Größe n genau m Blätter hat, ist gegeben durch*

$$\boxed{B(z,v)=\frac{v\left(v-\left(1+\sqrt{1-v}\right)^2\right)e^{-\sqrt{1-v}\,z}}{v-\left(1+\sqrt{1-v}\right)^2 e^{-2\sqrt{1-v}\,z}}\,.} \tag{1.227}$$

Eine analoge Vorgangsweise wählt man zur Ermittlung von $C(z,v)$. Die von (1.208) herrührende Differentialgleichung

$$\frac{\partial}{\partial z}\log C(z,v)=A(z,v)\,,\quad C(0,v)=1\,, \tag{1.228}$$

löst man wieder durch Trennung der Variablen.

Setzt man in der daraus resultierenden Gleichung

$$C(z,v)=k(v)e^{\int_0^z A(t,v)dt}\,, \tag{1.229}$$

für $A(z,v)$ ein und paßt $k(v)$ an die Anfangswerte an, folgt

Hilfssatz 1.8.3. *Die bivariate erzeugende Funktion $C(z,v)$ der Wahrscheinlichkeiten $C_{n,m}$, daß ein durch Zufallspermutationen erzeugtes Objekt aus der Familie $\mathcal{C}$ der Größe n genau m Blätter hat, ist gegeben durch*

$$\boxed{C(z,v) = \frac{\left(v - \left(1+\sqrt{1-v}\,\right)^2\right) e^{-\sqrt{1-v}\,z}}{v - \left(1+\sqrt{1-v}\,\right)^2 e^{-2\sqrt{1-v}\,z}}\,.} \tag{1.230}$$

1.8.2 Die Erwartungswerte der Anzahl der Blätter

Hier bezeichnen wir mit A_n, B_n und C_n die Erwartungswerte der Anzahl von Blättern in Objekten der Größe n aus den Familien $\mathcal{A}$, $\mathcal{B}$ und $\mathcal{C}$.

Die diesen Erwartungswerten entsprechenden erzeugenden Funktionen

$$A(z) = \sum_{n\geq 0} A_n\, z^n\,, \quad B(z) = \sum_{n\geq 0} B_n\, z^n \text{ und } C(z) = \sum_{n\geq 0} C_n\, z^n$$

sind durch die Beziehungen

$$A(z) = \frac{\partial}{\partial v} A(z,v)\Big|_{v=1}\,, \quad B(z) = \frac{\partial}{\partial v} B(z,v)\Big|_{v=1}\,, \quad C(z) = \frac{\partial}{\partial v} C(z,v)\Big|_{v=1}$$

gegeben.

Leitet man also (1.224), (1.227) und (1.230) partiell nach v ab und führt den Grenzübergang für $v=1$ durch, erhält man

$$A(z) = \frac{1}{3}z - \frac{1}{3} + \frac{1}{3}\frac{1}{(1-z)^2}\,, \tag{1.231a}$$

$$B(z) = -\frac{1}{6}z + \frac{1}{6} + \frac{1}{3}\frac{1}{(1-z)^2} + \frac{1}{2}\frac{1}{1-z}\,, \tag{1.231b}$$

$$C(z) = -\frac{1}{6}z + \frac{1}{6} + \frac{1}{3}\frac{1}{(1-z)^2} - \frac{1}{2}\frac{1}{1-z}\,. \tag{1.231c}$$

Ablesen der Koeffizienten von z in (1.231) führt schon zu

Satz 1.8.4. *Die Erwartungswerte A_n, B_n und C_n der Anzahl von Blättern in Objekten der Größe n aus den Familien $\mathcal{A}$, $\mathcal{B}$ und $\mathcal{C}$ sind gegeben durch*

$$\boxed{A_n = \frac{1}{3}(n+1) \quad \text{für } n \geq 2\,, \quad A_0 = 0\,, \quad A_1 = 1\,.} \tag{1.232a}$$

$$\boxed{B_n = \frac{1}{6}(2n+5) \quad \text{für } n \geq 2\,, \quad B_0 = 1\,, \quad B_1 = 1\,.} \tag{1.232b}$$

$$\boxed{C_n = \frac{1}{6}(2n-1) \quad \text{für } n \geq 2\,, \quad C_0 = 0\,, \quad C_1 = 0\,.} \tag{1.232c}$$

1.8.3 Die zweiten Faktoriellen Momente der Anzahl der Blätter

Diese Hilfsgrößen sind dienlich für die nachfolgende Ermittlung der Varianzen. Wir bezeichnen hier mit $\breve{A}_n$, $\breve{B}_n$ und $\breve{C}_n$ die zweiten Faktoriellen Momente der Anzahl von Blättern in Objekten der Größe n aus den Familien $\mathcal{A}$, $\mathcal{B}$ und $\mathcal{C}$.

Die diesen Größen entsprechenden gewöhnlichen erzeugenden Funktionen

$$\breve{A}(z) = \sum_{n\geq 0} \breve{A}_n \, z^n \,, \quad \breve{B}(z) = \sum_{n\geq 0} \breve{B}_n \, z^n \,, \quad \breve{C}(z) = \sum_{n\geq 0} \breve{C}_n \, z^n$$

sind gegeben durch die Beziehungen

$$\breve{A}(z) = \frac{\partial^2}{\partial v^2} A(z,v)\bigg|_{v=1} \,, \quad \breve{B}(z) = \frac{\partial^2}{\partial v^2} B(z,v)\bigg|_{v=1} \,, \quad \breve{C}(z) = \frac{\partial^2}{\partial v^2} C(z,v)\bigg|_{v=1} \,.$$

Durch zweimalige partielle Differentiation von (1.224), (1.227) und (1.230) nach v und anschließendem Grenzübergang für $v = 1$ ergibt sich also

$$\breve{A}(z) = \frac{2}{45} z^3 - \frac{2}{15} z^2 + \frac{2}{15} z + \frac{8}{45} + \frac{2}{9} \frac{1}{(1-z)^3} - \frac{2}{5} \frac{1}{(1-z)^2} \,, \tag{1.233a}$$

$$\breve{B}(z) = -\frac{7}{180} z^3 + \frac{7}{60} z^2 - \frac{17}{60} z + \frac{17}{180} + \frac{2}{9} \frac{1}{(1-z)^3} - \frac{1}{15} \frac{1}{(1-z)^2} - \frac{1}{4} \frac{1}{1-z} \,, \tag{1.233b}$$

$$\breve{C}(z) = -\frac{7}{180} z^3 + \frac{7}{60} z^2 + \frac{1}{20} z - \frac{43}{180} + \frac{2}{9} \frac{1}{(1-z)^3} - \frac{11}{15} \frac{1}{(1-z)^2} + \frac{3}{4} \frac{1}{1-z} \,. \tag{1.233c}$$

Liest man die Koeffizienten von z in obigen Ausdrücken (1.233) ab, folgt

Hilfssatz 1.8.5. *Die zweiten Faktoriellen Momente $\breve{A}_n$, $\breve{B}_n$ und $\breve{C}_n$ der Anzahl von Blättern in Objekten der Größe n aus den Familien $\mathcal{A}$, $\mathcal{B}$ und $\mathcal{C}$ sind gegeben durch*

$$\boxed{\begin{aligned} &\breve{A}_n = \frac{1}{45}(n+1)(5n-8) \quad \text{für } n \geq 4 \,, \\ &\breve{A}_0 = 0 \,, \quad \breve{A}_1 = 1 \,, \quad \breve{A}_2 = 0 \,, \quad \breve{A}_3 = \frac{2}{3} \,. \end{aligned}} \tag{1.234a}$$

$$\boxed{\begin{aligned} &\breve{B}_n = \frac{1}{9} n^2 + \frac{4}{15} n - \frac{17}{180} \quad \text{für } n \geq 4 \,, \\ &\breve{B}_0 = 0 \,, \quad \breve{B}_1 = 0 \,, \quad \breve{B}_2 = 1 \,, \quad \breve{B}_3 = \frac{5}{3} \,. \end{aligned}} \tag{1.234b}$$

$$\boxed{\begin{aligned} &\breve{C}_n = \frac{1}{9} n^2 - \frac{2}{5} n + \frac{43}{180} \quad \text{für } n \geq 4 \,, \\ &\breve{C}_0 = 0 \,, \quad \breve{C}_1 = 0 \,, \quad \breve{C}_2 = 0 \,, \quad \breve{C}_3 = 0 \,. \end{aligned}} \tag{1.234c}$$

1.8.4 Die Varianzen der Anzahl der Blätter

Wir bezeichnen hier mit $\hat{A}_n$, $\hat{B}_n$ und $\hat{C}_n$ die Varianzen der Anzahl von Blättern in Objekten der Größe n aus den Familien $\mathcal{A}$, $\mathcal{B}$ und $\mathcal{C}$.

Unter Zuhilfenahme der in (1.232) bzw. (1.234) eben errechneten Größen gelten dann die Beziehungen

$$\hat{A}_n = \breve{A}_n + A_n - A_n^2\,, \quad \hat{B}_n = \breve{B}_n + B_n - B_n^2\,, \quad \hat{C}_n = \breve{C}_n + C_n - C_n^2\,.$$

Dies führt unmittelbar zu

Satz 1.8.6. *Die Varianzen $\hat{A}_n$, $\hat{B}_n$ und $\hat{C}_n$ der Anzahl von Blättern in Objekten der Größe n aus den Familien $\mathcal{A}$, $\mathcal{B}$ und $\mathcal{C}$ sind gegeben durch*

$$\boxed{\begin{aligned}&\hat{A}_n = \frac{2}{45}(n+1) \quad \text{für } n \geq 4\,,\\ &\hat{A}_0 = 0\,, \quad \hat{A}_1 = 0\,, \quad \hat{A}_2 = 0\,, \quad \hat{A}_3 = \frac{2}{9}\,.\end{aligned}} \tag{1.235a}$$

$$\boxed{\begin{aligned}&\hat{B}_n = \frac{2}{45}(n+1) \quad \text{für } n \geq 4\,,\\ &\hat{B}_0 = 0\,, \quad \hat{B}_1 = 0\,, \quad \hat{B}_2 = \frac{1}{4}\,, \quad \hat{B}_3 = \frac{5}{36}\,.\end{aligned}} \tag{1.235b}$$

$$\boxed{\begin{aligned}&\hat{C}_n = \frac{2}{45}(n+1) \quad \text{für } n \geq 4\,,\\ &\hat{C}_0 = 0\,, \quad \hat{C}_1 = 0\,, \quad \hat{C}_2 = \frac{1}{4}\,, \quad \hat{C}_3 = \frac{5}{36}\,.\end{aligned}} \tag{1.235c}$$

1.8.5 Asymptotische Betrachtungen zu Erwartungswert, Varianz und Verteilung der Anzahl der Knoten mit 0, 1 bzw. 2 Kindern

Es gibt in diesem Zusammenhang eine nicht offensichtliche Verbindung zwischen den p-trees und den in Kapitel 2 näher betrachteten Binary Search Trees. Es gilt nämlich,

Satz 1.8.7. *Die Verteilung der Anzahl der Knoten mit 0, 1 bzw. 2 Kindern in p-trees ist ident mit der Verteilung der Anzahl der Knoten 0, 1, bzw. 2 Kindern in den Binary Search Trees (Unter einem Kind eines Knotens verstehen wir einen direkten Nachfolger dieses Knotens).*

Beweis. Zunächst ist zu bemerken, daß es zum Beweis des Satzes genügt, wenn man zeigen kann, daß die Anzahl der Blätter in p-trees und in Binary Search Trees ident verteilt sind. Bezeichnen wir nämlich in einem beliebigen Binärbaum der Größe n die Anzahl der Knoten mit 0, 1 bzw. 2 Kindern mit n_0, n_1 bzw. n_2, so gelten die Beziehungen

$$\begin{aligned} n_0 + n_1 + n_2 &= n \quad \text{(Zählen der Knoten im Baum)},\\ n_1 + 2n_2 &= n - 1 \quad \text{(Zählen der Kanten im Baum)}. \end{aligned}$$

Damit sind also n_1 bzw. n_2 aus n_0 zu erhalten durch

$$n_1 = n + 1 - 2n_0 \quad \text{und} \quad n_2 = n_0 - 1\,.$$

Somit sind also die Verteilungen der Anzahl der Knoten mit 1 bzw. 2 Kindern aus der Verteilung der Anzahl der Blätter zu erhalten oder anders ausgedrückt, bei identer Verteilung der Anzahl der Blätter zweier Binärbaumklassen müssen zwangsläufig auch die Anzahl der Knoten mit 1 bzw. 2 Kindern dieser Baumklassen ident verteilt sein.

Nun betrachten wir die Verteilung der Anzahl der Blätter in Binary Search Trees. Bezeichnen wir dazu mit $F_{n,m}$ die Anzahl der Binary Search Trees der Größe n, die genau m Blätter haben. Wir bekommen unmittelbar aus der Art der Konstruktion dieser Bäume die rekursive Beziehung

$$\begin{aligned}
&F_{n,m} = \frac{1}{n}\sum_{k=1}^{n}\sum_{l=1}^{m} F_{k-1,l}\cdot F_{n-k,m-l} \quad \text{für } n \geq 2\,,\\
&F_{0,0} = 1 \;\text{ und }\; F_{1,1} = 1\,.
\end{aligned} \tag{1.236}$$

Definieren wir durch

$$F(z,v) = \sum_{n\geq 0}\sum_{m\geq 0} F_{n,m} z^n v^m$$

die bivariate erzeugende Funktion $F(z,v)$ der $F_{n,m}$, so übersetzt sich obige Rekursion (1.236) durch Multiplizieren mit $nz^{n-1}v^m$ und Summieren über alle $n \geq 2$ und $m \geq 0$ in folgende Differentialgleichung

$$\begin{aligned}
&\frac{\partial}{\partial z}F(z,v) = F^2(z,v) + v - 1\,,\\
&F(0,v) = 1\,.
\end{aligned} \tag{1.237}$$

Da nun die Differentialgleichungen (1.216) und (1.237) für die erzeugenden Funktionen der Verteilung der Anzahl der Blätter in p-trees bzw. Binary Search Trees übereinstimmen, ist die Gleichheit der Verteilungen der Blattanzahlen in p-trees und in Binary Search Trees und mit dem eingangs Gesagten auch dieser Satz gezeigt. □

Dies erlaubt uns nun, die von Devroye in [5] gezeigten Aussagen über die Verteilung der Knoten mit 0, 1 bzw. 2 Kindern in Binary Search Trees für die p-trees zu übernehmen. Bezeichnen wir die Zufallsvariablen für die Anzahl der Knoten mit 0,1 bzw. 2 Kindern in p-trees der Größe n mit $I_n^{(0)}$, $I_n^{(1)}$ bzw. $I_n^{(2)}$, so bekommen wir

Satz 1.8.8. *Für die Erwartungswerte $\mathbb{E}\left(I_n^{(i)}\right)$ und die Varianzen $\mathbb{V}\left(I_n^{(i)}\right)$ mit $0 \leq i \leq 2$ gelten für $n \to \infty$ die Beziehungen*

$$\begin{aligned}
&\mathbb{E}\left(I_n^{(0)}\right) \sim \frac{n}{3}\,,\; \mathbb{E}\left(I_n^{(1)}\right) \sim \frac{n}{3}\,,\; \mathbb{E}\left(I_n^{(2)}\right) \sim \frac{n}{3}\,,\\
&\mathbb{V}\left(I_n^{(0)}\right) \sim \frac{2}{45}n\,,\; \mathbb{V}\left(I_n^{(1)}\right) \sim \frac{8}{45}n\,,\; \mathbb{V}\left(I_n^{(2)}\right) \sim \frac{2}{45}n\,.
\end{aligned}$$

Weiters ist die Folge der Zufallsvariablen

$$\frac{I_n^{(0)} - \frac{n}{3}}{\sqrt{\frac{2}{45}n}}\,,\; \frac{I_n^{(1)} - \frac{n}{3}}{\sqrt{\frac{8}{45}n}} \;\; bzw. \;\; \frac{I_n^{(2)} - \frac{n}{3}}{\sqrt{\frac{2}{45}n}}$$

für $n \to \infty$ schwach konvergent gegen die standardisierte Normalverteilung $\mathcal{N}(0,1)$ mit Mittelwert 0 und Varianz 1.

1.9 Analyse einer Prozedur zum Aufteilen der p-trees

1.9.1 Der Algorithmus `Cut`

Nach den schon früher besprochenen Algorithmen `Insert` und `Delete` zum Einfügen bzw. Löschen eines Elements aus dem p-tree, wollen wir jetzt einen Algorithmus zum Aufteilen eines p-trees an einer beliebigen Stelle in zwei p-trees zuwenden. Diese Prozedur wurde bereits in [32] vorgestellt. Dabei wird der Baum T bei vorgegebenem Schlüssel s so in zwei p-trees aufgeteilt, daß der eine, wir nennen ihn T_1, alle Knoten von T mit Schlüssel kleiner oder gleich s besitzt, während im restlichen Baum $T - T_1$ alle Knoten von T mit Schlüssel größer als s verbleiben. Die entstehenden p-trees können auch leer sein.

Im Detail ist die Vorgangsweise beim Aufteilen des p-trees T bei vorgegebenem Schlüssel s durch den rekursiven Algorithmus `Cut` wie folgt beschrieben. (Ist x ein Knoten vom Baum T, dann soll $T(x)$ den Unterbaum mit Wurzel x bezeichnen.)

1. (a) Wenn T leer ist, oder die Wurzel von T einen niedrigeren Schlüssel als s hat, so setze man den Baum T_1 gleich T, während T danach der leere Baum ist.
   (b) Sonst setze man T_1 als leeren Baum und führe Schritt 2 aus.
2. Gehe entlang des linken Pfades von T, bis man auf den ersten Knoten y mit Schlüssel kleiner oder gleich s trifft.
   (a) Falls man keinen solchen Knoten findet, terminiert der Algorithmus.
   (b) Ansonsten sei z der Vorgänger von y und es gilt, daß s zwischen den Schlüsseln von y und z liegt.
      i. Ist der rechte Unterbaum von z leer, dann setze $T_1 := T(y)$ und $T := T - T(y)$.
      ii. Ansonsten hänge den rechten Unterbaum von z zwischen dem linken Nachfolger von z und z selbst, während der rechte Unterbaum von z leer gelassen wird. Dann suche man das linke Blatt von $T(z)$ und verbinde den Knoten y mit diesem. Schließlich wiederholt man Schritt 2 beginnend mit dem Knoten z.

1.9.2 Erwartungswerte der Anzahl der Schlüsselvergleiche in `Cut`

In [32] wurde angenommen, daß die Anzahl der Schlüsselvergleiche beim Aufteilen eines p-trees an der Stelle s im `Cut`-Algorithmus gleich sind wie die Anzahl der Schlüsselvergleiche beim Einfügen dieses Elements s in den p-tree im Algorithmus `Insert`. Das ist allerdings nicht ganz korrekt. Man sieht dies ein, wenn man bedenkt, daß im Algorithmus `Cut` in manchen Fällen ein Teil des Baumes an einen anderen Teil angehängt wird, in dem noch weiter nach der Stelle des Aufteilens gesucht wird. So kommt es in speziellen Fällen dazu, daß bei einem Element zweimal der Schlüssel verglichen wird. Dies ist in der Prozedur `Insert` nicht der Fall.

Nun gehen wir also in unserer Analyse von durch Zufallspermutationen erzeugten Objekten aus den Familien $\mathcal{A}$, $\mathcal{B}$ und $\mathcal{C}$ aus. Wir bezeichnen dann hier mit $A_{n,j}^{[C]}$, $B_{n,j}^{[C]}$ und $C_{n,j}^{[C]}$ die

Erwartungswerte der Anzahl an Schlüsselvergleichen, die beim Aufteilen eines entsprechenden Objekts der Größe n durch den Schlüssel $j + \frac{1}{2}$ mit $j \in \mathbb{Z}$ und $0 \leq j \leq n$ gemacht werden.

Für die Erstellung einer Rekursionsformel der gesuchten Größen teilen wir die Objektfamilien wieder aufgrund des ersten Elements k der Zufallspermutation auf. Wird nun an der Stelle $k+\frac{1}{2}$ aufgeteilt, ist also j gleich k, dann wird in weiterer Folge k noch einmal verglichen, wenn der Knoten mit Schlüssel k einen Vorgänger mit nichtleerem rechten Unterbaum hat. Dies spiegelt sich in den Rekursionen durch die Funktionen $r_1(n,j)$, $r_2(n,j)$ und $r_3(n,j)$ wider.

Man erhält so die Beziehungen

$$A_{n,j}^{[C]} = \frac{1}{n}\left(\sum_{k=1}^{j} C_{n-k,j-k}^{[C]} + \sum_{k=j+1}^{n}\left(B_{k-1,j}^{[C]} + H_{n-k}\right) + r_1(n,j)\right)$$
$$\text{für } n \geq 1 \text{ und } n \geq j \geq 0\,, \tag{1.238a}$$

$$B_{n,j}^{[C]} = \frac{1}{n}\left(\sum_{k=1}^{j}\left(A_{n-k,j-k}^{[C]} + 2\right) + \sum_{k=j+1}^{n}\left(B_{k-1,j}^{[C]} + 1\right) + r_2(n,j)\right)$$
$$\text{für } n \geq 1 \text{ und } n \geq j \geq 0\,, \tag{1.238b}$$

$$C_{n,j}^{[C]} = \frac{1}{n}\left(\sum_{k=1}^{j} C_{n-k,j-k}^{[C]} + \sum_{k=j+1}^{n}\left(A_{k-1,j}^{[C]} + H_{n-k} + 2\right) + r_3(n,j)\right)$$
$$\text{für } n \geq 1 \text{ und } n \geq j \geq 0\,, \tag{1.238c}$$

$$A_{0,0}^{[C]} = 0\,, \quad B_{0,0}^{[C]} = 1\,, \quad C_{0,0}^{[C]} = 1\,,$$

mit

$$r_1(n,j) = \begin{cases} 1 - \frac{1}{n-j} & \text{für } 1 \leq j \leq n-1\,, \\ 0 & \text{sonst}\,, \end{cases}$$
$$r_2(n,j) = \begin{cases} 1 & \text{für } 1 \leq j \leq n-1\,, \\ 0 & \text{sonst}\,, \end{cases}$$
$$r_3(n,j) = \begin{cases} 1 - \frac{1}{n-j} & \text{für } 1 \leq j \leq n-1\,, \\ 0 & \text{sonst}\,. \end{cases}$$

Das Lösen dieser Rekursion geschieht wie gehabt. Zunächst definiert man sich also durch

$$A^{[C]}(z,u) = \sum_{j\geq 0}\sum_{n\geq j} A_{n,j}^{[C]} z^n u^j\,, \quad B^{[C]}(z,u) = \sum_{j\geq 0}\sum_{n\geq j} B_{n,j}^{[C]} z^n u^j\,,$$
$$C^{[C]}(z,u) = \sum_{j\geq 1}\sum_{n\geq j} C_{n,j}^{[C]} z^n u^j$$

die den gesuchten Erwartungswerten entsprechenden bivariaten erzeugenden Funktionen in z und u.

Nun multipliziert man (1.238) mit $nz^{n-1}u^j$ und summiert die entstehenden Gleichungen über alle $n \geq j \geq 0$ auf. Das daraus resultierende System von Differentialgleichungen lautet

$$\frac{\partial}{\partial z} A^{[C]}(z,u) = \frac{u}{1-uz} C^{[C]}(z,u) + \frac{1}{(1-z)^2(1-uz)} \log\left(\frac{1}{1-z}\right)$$
$$\frac{uz}{(1-z)(1-uz)} - \frac{u}{1-uz} \log\left(\frac{1}{1-z}\right) + \frac{1}{1-z} B^{[C]}(z,u) , \tag{1.239a}$$

$$\frac{\partial}{\partial z} B^{[C]}(z,u) = \frac{u}{1-uz} A^{[C]}(z,u) + \frac{2u}{(1-z)(1-uz)^2} + \frac{1}{(1-z)^2(1-uz)}$$
$$+ \frac{uz}{(1-z)(1-uz)} + \frac{1}{1-z} B^{[C]}(z,u) , \tag{1.239b}$$

$$\frac{\partial}{\partial z} C^{[C]}(z,u) = \frac{u}{1-uz} C^{[C]}(z,u) + \frac{1}{(1-z)^2(1-uz)} \log\left(\frac{1}{1-z}\right)$$
$$+ \frac{2}{(1-z)^2(1-uz)} + \frac{uz}{(1-z)(1-uz)} - \frac{u}{1-uz} \log\left(\frac{1}{1-z}\right)$$
$$+ \frac{1}{1-z} A^{[C]}(z,u) . \tag{1.239c}$$

Kürzen wir die Störfunktionen ab mit

$$\begin{aligned}
f_1(z,u) &= \frac{1}{(1-z)^2(1-uz)} \log\left(\frac{1}{1-z}\right) + \frac{uz}{(1-z)(1-uz)} - \frac{u}{1-uz} \log\left(\frac{1}{1-z}\right) , \\
f_2(z,u) &= \frac{2u}{(1-z)(1-uz)^2} + \frac{1}{(1-z)^2(1-uz)} + \frac{uz}{(1-z)(1-uz)} , \\
f_3(z,u) &= \frac{1}{(1-z)^2(1-uz)} \log\left(\frac{1}{1-z}\right) + \frac{2}{(1-z)^2(1-uz)} + \frac{uz}{(1-z)(1-uz)} \\
&\quad - \frac{u}{1-uz} \log\left(\frac{1}{1-z}\right) ,
\end{aligned}$$

dann lautet das zu lösende Differentialgleichungssystem

$$\frac{\partial}{\partial z} A^{[C]}(z,u) = \frac{u}{1-uz} C^{[C]}(z,u) + \frac{1}{1-z} B^{[C]}(z,u) + f_1(z,u) , \tag{1.240a}$$

$$\frac{\partial}{\partial z} B^{[C]}(z,u) = \frac{u}{1-uz} A^{[C]}(z,u) + \frac{1}{1-z} B^{[C]}(z,u) + f_2(z,u) , \tag{1.240b}$$

$$\frac{\partial}{\partial z} C^{[C]}(z,u) = \frac{u}{1-uz} C^{[C]}(z,u) + \frac{1}{1-z} A^{[C]}(z,u) + f_3(z,u) . \tag{1.240c}$$

Dieses System von Differentialgleichungen wandelt man um in folgende lineare Differentialgleichung dritter Ordnung für $A^{[C]}(z,u)$:

$$\begin{aligned}
&\frac{2uz-u-1}{(1-z)(1-u)} A^{[C]}(z,u) - 2\frac{1-uz}{1-u} \frac{\partial}{\partial z} A^{[C]}(z,u) \\
&- \frac{3}{2} \frac{(1-uz)(2uz-u-1)}{u(1-u)} \frac{\partial^2}{\partial z^2} A^{[C]}(z,u) - \frac{1}{2} \frac{(1-uz)^2(1-z)}{u(1-u)} \frac{\partial^3}{\partial z^3} A^{[C]}(z,u) \\
&= f_3(z,u) + \frac{1}{1-uz} \frac{\partial}{\partial z} \left(\frac{(1-uz)^3}{2u(u-1)} f_2(z,u) \right. \\
&\quad \left. + \frac{(1-uz)^3}{2u(u-1)(1-z)} \frac{\partial}{\partial z} \left((1-z)^2 f_1(z,u) \right) + \frac{(1-uz)^2(1-z)}{2(u-1)} f_3(z,u) \right) .
\end{aligned} \tag{1.241}$$

Setzt man für $f_1(z,u)$, $f_2(z,u)$ und $f_3(z,u)$ wieder ein und berücksichtigt die Anfangswerte $A^{[C]}_{0,0}=0$, $A^{[C]}_{1,0}=1$, $A^{[C]}_{1,1}=1$, $A^{[C]}_{2,0}=2$, $A^{[C]}_{2,1}=2$ und $A^{[C]}_{2,2}=1$, so bekommt man aus (1.241) als zu lösende Differentialgleichung

$$\begin{aligned}
&\frac{2uz-u-1}{(1-z)(1-u)}A^{[C]}(z,u) - 2\frac{1-uz}{1-u}\frac{\partial}{\partial z}A^{[C]}(z,u)\\
&\quad-\frac{3}{2}\frac{(1-uz)(2uz-u-1)}{u(1-u)}\frac{\partial^2}{\partial z^2}A^{[C]}(z,u)-\frac{1}{2}\frac{(1-uz)^2(1-z)}{u(1-u)}\frac{\partial^3}{\partial z^3}A^{[C]}(z,u)\\
&\quad=\frac{2u^2z^3-4u^2z^2-uz^2+2u^2z+3uz-2}{(1-z)^3(1-u)u}\,,\\
&A^{[C]}(0,u)=0\,,\quad \frac{\partial}{\partial z}A^{[C]}(z,u)\bigg|_{z=0}=1+u\,,\quad \frac{\partial^2}{\partial z^2}A^{[C]}(z,u)\bigg|_{z=0}=2\left(2+2u+u^2\right)\,.
\end{aligned}\tag{1.242}$$

Mit den Transformationen

$$\begin{aligned}
A^{[C]}(z,u)&=\frac{E^{[C]}(z,u)}{(1-z)(1-uz)}\,,\\
F^{[C]}(z,u)&=\frac{\partial}{\partial z}E^{[C]}(z,u)\,,\\
F^{[C]}(z,u)&=(1-z)(1-uz)G^{[C]}(z,u)\,,\\
H^{[C]}(z,u)&=\frac{\partial}{\partial z}G^{[C]}(z,u)\,,
\end{aligned}$$

ergibt sich daraus die lineare Differentialgleichung erster Ordnung

$$\frac{\partial}{\partial z}H^{[C]}(z,u)-2\frac{1+u-2uz}{(1-z)(1-uz)}H^{[C]}(z,u)=-2\frac{2u^2z^3-4u^2z^2-uz^2+2u^2z+3uz-2}{(1-uz)^2(1-z)^4}\,.\tag{1.243}$$

Durch Variation der Konstanten findet man die Partikulärlösung

$$H_p^{[C]}(z,u)=2\frac{u}{(1-z)^2(1-uz)^2}\log\left(\frac{1}{1-z}\right)+2\frac{u^2z^3-u^2z^2-uz^2+uz-2u+u}{(1-z)^3(1-uz)^2}\,.\tag{1.244}$$

Durch Rücksubstitution ergibt sich daraus eine Partikulärlösung von (1.242):

$$A_p^{[C]}(z,u)=\frac{1}{(1-z)(1-uz)}\int_0^z\left[(1-x)(1-ux)\int_0^x H_p^{[C]}(t,u)\,dt\right]dx\,.\tag{1.245}$$

Für ein leichteres Ablesen der Koeffizienten vereinfacht man durch partielle Integration diesen Ausdruck weiter zu

$$\begin{aligned}
A_p^{[C]}(z,u)=&\left(\frac{2}{3}\frac{(2uz-u-1)u^2}{(1-u)^3}+\frac{2}{3}\frac{u}{(1-z)(1-u)}\right)\int_0^z\frac{1}{1-ut}\log\left(\frac{1}{1-t}\right)dt\\
&+\left(-\frac{1}{3}\frac{(2uz-u-1)u}{(1-u)^3}-\frac{1}{3}\frac{u}{(1-uz)(1-u)}\right)\log^2\left(\frac{1}{1-z}\right)
\end{aligned}$$

$$+\left[\frac{1}{3}\frac{12u^2z-2u^2-2uz-9u+1}{(1-u)^3}-\frac{1}{3}\frac{1}{(1-uz)(1-u)}\right.$$
$$\left.+2\frac{1}{(1-z)(1-u)}\right]\log\left(\frac{1}{1-z}\right)$$
$$+\left(-\frac{1}{3}\frac{(6u-1)(2uz-u-1)}{(1-u)^3}-\frac{1}{3}\frac{6u-1}{(1-z)(1-u)u}\right)\log\left(\frac{1}{1-uz}\right)$$
$$+\frac{1}{3}\frac{3u^2+10u+1}{(1-u)^2}+\frac{1}{3}\frac{(u+4)u}{(1-uz)(1-u)}+\frac{2}{3}\frac{u-3}{(1-z)(1-u)} \tag{1.246}$$

und schließlich zu

$$A_p^{[C]}(z,u)=\left(\frac{2}{3}uz-\frac{1}{3}u-\frac{1}{3}+\frac{1}{3}\frac{(u-1)^2}{(1-z)u}\right)\int_0^z\frac{1}{(1-t)^2(1-ut)^2}\log\left(\frac{1}{1-t}\right)dt$$
$$+\left(-\frac{1}{3}\frac{u}{(1-uz)(1-u)}+\frac{1}{3}\frac{1}{(1-z)(1-u)}\right)\log^2\left(\frac{1}{1-z}\right)$$
$$+\left[\frac{1}{3}\frac{(7u-1)(2uz-u-1)}{(1-u)^3}+\frac{7}{3}\frac{1}{(1-z)(1-u)}-\frac{1}{3}\frac{1}{(1-z)^2u}\right.$$
$$\left.-\frac{1}{3}\frac{1}{(1-uz)(1-u)}\right]\log\left(\frac{1}{1-z}\right)$$
$$+\left(-\frac{1}{3}\frac{(7u-1)(2uz-u-1)}{(1-u)^3}-\frac{1}{3}\frac{7u-1}{(1-z)(1-u)u}\right)\log\left(\frac{1}{1-uz}\right)$$
$$+\frac{1}{3}\frac{3u^2+8u+1}{(1-u)^2}+\frac{1}{3}\frac{2u-7}{(1-z)(1-u)}+\frac{1}{3}\frac{1}{(1-z)^2u}+\frac{1}{3}\frac{(u+4)u}{(1-uz)(1-u)}. \tag{1.247}$$

Paßt man die drei Funktionen $c_1(u)$, $c_2(u)$ und $c_3(u)$ in der Lösung

$$A_h^{[I]}(z,u)=\frac{c_1(u)}{(1-z)(1-uz)}+c_2(u)\frac{z(6-3uz-3z+2uz^2)}{(1-z)(1-uz)}$$
$$+\frac{c_3(u)}{(1-z)(1-uz)}\int_0^z\left[(1-x)(1-ux)\int_0^x\frac{1}{(1-t)^2(1-ut)^2}dt\right]dx\,, \tag{1.248}$$

der (1.242) entsprechenden homogenen Differentialgleichung an die Anfangswerte an, so ermittelt man für diese

$$c_1(u)=\frac{1}{18}\frac{u^4-14u^3-108u^2+14u-1}{(1-u)^2u}\,,\quad c_2(u)=\frac{1}{3}\frac{u}{(1-u)}\,,$$
$$c_3(u)=-\frac{1}{18}\frac{u^2+6u-1}{(1-u)^3u}\,.$$

Zusammengefaßt ergibt sich damit als Lösung der doppelten erzeugenden Funktion $A^{[C]}(z,u)$ für die Erwartungswerte $A_{n,j}^{[C]}$ der Anzahl der Key-Comparisons beim Aufteilen eines p-trees der Größe n mit der Prozedur `Cut` durch einen Knoten mit Schlüssel $j+\frac{1}{2}$

$$A^{[C]}(z,u)=\left(\frac{2}{3}uz-\frac{1}{3}u-\frac{1}{3}+\frac{1}{3}\frac{(u-1)^2}{(1-z)u}\right)\int_0^z\frac{1}{(1-t)^2(1-ut)^2}\log\left(\frac{1}{1-t}\right)dt$$

$$
\begin{aligned}
&+\left(-\frac{1}{3}\frac{u}{(1-uz)(1-u)}+\frac{1}{3}\frac{1}{(1-z)(1-u)}\right)\log^2\left(\frac{1}{1-z}\right)\\
&+\left(\frac{1}{3}\frac{(2uz-u-1)u}{(1-u)^2}+\frac{7}{3}\frac{1}{(1-z)(1-u)}-\frac{1}{3}\frac{1}{(1-z)^2u}\right.\\
&\qquad\left.-\frac{1}{3}\frac{(u+6)u}{(1-uz)(1-u)}\right)\log\left(\frac{1}{1-z}\right)\\
&+\left(-\frac{1}{3}\frac{(2uz-u-1)u}{(1-u)^2}-\frac{1}{3}\frac{1}{(1-z)}\right)\log\left(\frac{1}{1-uz}\right)\\
&+\frac{2}{3}\frac{uz-u+1}{1-u}+\frac{2}{3}\frac{1}{1-uz}-\frac{1}{3}\frac{4u+1}{(1-z)u}+\frac{1}{3}\frac{1}{(1-z)^2u}\,. \qquad (1.249)
\end{aligned}
$$

Alle für das Ablesen der Koeffizienten $[z^n u^j]$ in $A^{[C]}(z,u)$ notwendigen Beziehungen (1.19) und (1.62) wurden schon in den vorangegangenen Abschnitten hergeleitet, und es ergibt sich mit diesen aus (1.249)

Satz 1.9.1. *Der Erwartungswert $A^{[C]}_{n,j}$ für die Anzahl der Schlüsselvergleiche, die beim Aufteilen eines p-trees der Größe n mit dem Algorithmus* `Cut` *durch ein Element mit Schlüssel $j+\frac{1}{2}$ gemacht werden, ist gegeben durch*

$$
\boxed{
\begin{aligned}
A^{[C]}_{n,j} = &-\frac{2}{3}\sum_{k=j+1}^{n}\frac{H_{k-j}}{k}+\frac{1}{3}H^2_{n-j}+\left(-\frac{1}{3}-\frac{2}{3}\frac{1}{j}\right)H_n\\
&+\left(\frac{7}{3}+\frac{2}{3}\frac{1}{j}+\frac{2}{3}\frac{j^2}{n}-\frac{2}{3}\frac{j(j-1)}{n-1}\right)H_{n-j}\\
&+\left(\frac{1}{3}+\frac{2}{3}\frac{1}{j}\right)H_j-\frac{1}{3}H^{(2)}_{n-j}-\frac{1}{3}\frac{2j^2+3j-1}{n}+\frac{2}{3}\frac{(j-1)(j+1)}{n-1}\\
&+\frac{1}{3}\frac{1}{n-j+1}-\frac{1}{3}\frac{1}{j}+\frac{1}{3}\quad \text{für } 1\le j\le n-1\,,
\end{aligned}}
$$

$$
\boxed{A^{[C]}_{n,n}=1\quad\text{für } n\ge 1\,,\quad A^{[C]}_{0,0}=0\,,}
$$

$$
\boxed{A^{[C]}_{n,0}=2H_n-1\quad\text{für } n\ge 1\,.} \qquad (1.250)
$$

Als Lösung der doppelten erzeugenden Funktion $C^{[C]}(z,u)$ erhält man

$$
\begin{aligned}
C^{[C]}(z,u) = &\left(-\frac{1}{3}uz-\frac{1}{3}u+\frac{2}{3}+\frac{1}{3}\frac{(u-1)^2}{(1-z)u}\right)\int_0^z\frac{1}{(1-t)^2(1-ut)^2}\log\left(\frac{1}{1-t}\right)dt\\
&+\left(-\frac{1}{3}\frac{u}{(1-uz)(1-u)}+\frac{1}{3}\frac{1}{(1-z)(1-u)}\right)\log^2\left(\frac{1}{1-z}\right)\\
&+\left(-\frac{1}{3}\frac{u^2z-2u^2+4u-3}{(1-u)^2}-\frac{1}{3}\frac{1}{(1-z)^2u}+\frac{7}{3}\frac{1}{(1-z)(1-u)}\right.\\
&\qquad\left.-\frac{1}{3}\frac{(u+6)u}{(1-uz)(1-u)}\right)\log\left(\frac{1}{1-z}\right)\\
&+\left(\frac{1}{3}\frac{(uz+u-2)u}{(1-u)^2}-\frac{1}{3}\frac{1}{1-z}\right)\log\left(\frac{1}{1-uz}\right)\\
&-\frac{1}{3}\frac{uz+5u-5}{1-u}+\frac{2}{3}\frac{1}{1-uz}-\frac{1}{3}\frac{4u+1}{(1-z)u}+\frac{1}{3}\frac{1}{(1-z)^2u}\,. \qquad (1.251)
\end{aligned}
$$

Das Ablesen der Koeffizienten führt unmittelbar zu

Hilfssatz 1.9.2. *Der Erwartungswert $C^{[C]}_{n,j}$ für die Anzahl der Schlüsselvergleiche, die beim Auteilen eines Objekts der Größe n aus der Familie $\mathcal{C}$ mit dem Algorithmus* `Cut` *durch ein Element mit Schlüssel $j+\frac{1}{2}$ gemacht werden, ist gegeben durch*

$$\boxed{\begin{aligned} C^{[C]}_{n,j} &= -\frac{2}{3}\sum_{k=j+1}^{n}\frac{H_{k-j}}{k} + \frac{1}{3}H^2_{n-j} + \left(-\frac{1}{3}-\frac{2}{3}\frac{1}{j}\right)H_n \\ &\quad + \left(\frac{10}{3}+\frac{2}{3}\frac{1}{j}-\frac{1}{3}\frac{j(j+3)}{n}+\frac{1}{3}\frac{j(j-1)}{n-1}\right)H_{n-j} \\ &\quad + \left(\frac{1}{3}+\frac{2}{3}\frac{1}{j}\right)H_j - \frac{1}{3}H^{(2)}_{n-j} + \frac{1}{3}\frac{j^2+3j+1}{n} - \frac{1}{3}\frac{(j-1)(j+1)}{n-1} \\ &\quad + \frac{1}{3}\frac{1}{n-j+1} - \frac{1}{3}\frac{1}{j} - \frac{2}{3} \qquad \text{für } 1\le j\le n-1\,, \end{aligned}}$$

$$\boxed{C^{[C]}_{n,n} = 1 \quad \text{für } n\ge 0\,,}$$

$$\boxed{C^{[C]}_{n,0} = 3H_n - 2 + \frac{1}{n} \quad \text{für } n\ge 1\,.} \tag{1.252}$$

Für $B^{[C]}(z,u)$ findet man

$$\begin{aligned} B^{[C]}(z,u) &= \left(-\frac{1}{3}uz+\frac{2}{3}u-\frac{1}{3}+\frac{1}{3}\frac{(u-1)^2}{(1-z)u}\right)\int_0^z \frac{1}{(1-t)^2(1-ut)^2}\log\left(\frac{1}{1-t}\right)dt \\ &\quad + \left(-\frac{1}{3}\frac{u}{(1-uz)(1-u)}+\frac{1}{3}\frac{1}{(1-z)(1-u)}\right)\log^2\left(\frac{1}{1-z}\right) \\ &\quad + \left(-\frac{1}{3}\frac{(uz-2u+1)u}{(1-u)^2}+\frac{4}{3}\frac{1}{(1-z)(1-u)}-\frac{1}{3}\frac{1}{(1-z)^2u}\right. \\ &\qquad \left. -\frac{1}{3}\frac{(u+3)u}{(1-uz)(1-u)}\right)\log\left(\frac{1}{1-z}\right) \\ &\quad + \left(\frac{1}{3}\frac{(uz-2u+1)u}{(1-u)^2}-\frac{1}{3}\frac{1}{1-z}\right)\log\left(\frac{1}{1-uz}\right) \\ &\quad - \frac{1}{3}\frac{uz-u+1}{1-u} - \frac{2}{3}\frac{4u-1}{(1-uz)(1-u)} + \frac{1}{3}\frac{(1+u)(4u-1)}{(1-z)(1-u)u} + \frac{1}{3}\frac{1}{(1-z)^2u}\,. \end{aligned} \tag{1.253}$$

Das Ablesen der Koeffizienten ergibt schließlich

Hilfssatz 1.9.3. *Der Erwartungswert $B^{[C]}_{n,j}$ für die Anzahl der Schlüsselvergleiche, die beim Auteilen eines Objekts der Größe n aus der Familie $\mathcal{B}$ mit dem Algorithmus* `Cut` *durch ein Element mit Schlüssel $j+\frac{1}{2}$ gemacht werden, ist gegeben durch*

$$\boxed{\begin{aligned} B^{[C]}_{n,j} &= -\frac{2}{3}\sum_{k=j+1}^{n}\frac{H_{k-j}}{k} + \frac{1}{3}H^2_{n-j} + \left(-\frac{1}{3}-\frac{2}{3}\frac{1}{j}\right)H_n \\ &\quad + \left(\frac{4}{3}+\frac{2}{3}\frac{1}{j}-\frac{1}{3}\frac{j(j-3)}{n}+\frac{1}{3}\frac{j(j-1)}{n-1}\right)H_{n-j} \\ &\quad + \left(\frac{1}{3}+\frac{2}{3}\frac{1}{j}\right)H_j - \frac{1}{3}H^{(2)}_{n-j} + \frac{1}{3}\frac{j^2-2}{n} - \frac{1}{3}\frac{(j-1)(j+1)}{n-1} \\ &\quad + \frac{1}{3}\frac{1}{n-j+1} - \frac{1}{3}\frac{1}{j} + \frac{7}{3} \qquad \text{für } 1\le j\le n-1\,, \end{aligned}}$$

$$\boxed{B_{n,n}^{[C]} = 3 - \frac{1}{n} \quad \text{für } n \geq 1\,, \quad B_{0,0}^{[C]} = 1\,,}$$

$$\boxed{B_{n,0}^{[C]} = H_n + 1 \quad \text{für } n \geq 1\,.} \tag{1.254}$$

1.9.3 Asymptotische Betrachtungen und weitere Folgerungen

Mit den Beziehungen (1.92), (1.105), (1.110) und (1.112) erhalten wir aus Gleichung (1.250) die asymptotischen Entwicklungen der Anzahl der Schlüsselvergleiche beim Aufteilen des p-trees mit dem Algorithmus `Cut` durch einen Knoten mit Schlüssel $j + \frac{1}{2}$.

Korollar 1.9.4. *Der Erwartungswert $A_{n,j}^{[C]}$ für die Anzahl der Schlüsselvergleiche, die beim Aufteilen eines p-trees der Größe n mit dem Algorithmus* `Cut` *durch ein Element mit Schlüssel $j + \frac{1}{2}$ gemacht werden, ist asymptotisch für $j = \rho n$ und $0 < \rho < 1$ gegeben durch*

$$\boxed{A_{n,j}^{[C]} = \frac{1}{3}\log^2 n + \left(-\frac{2}{3}\rho^2 + \frac{2}{3}\rho + \frac{2}{3}\gamma + \frac{7}{3} + \frac{2}{3}\log\rho + \frac{2}{3}\log(1-\rho)\right) + \mathcal{O}(1)\,.} \tag{1.255}$$

Korollar 1.9.5. *Der Erwartungswert $A_{n,j}^{[C]}$ für die Anzahl der Schlüsselvergleiche, die beim Aufteilen eines p-trees der Größe n mit dem Algorithmus* `Cut` *durch ein Element mit Schlüssel $j + \frac{1}{2}$ gemacht werden, ist asymptotisch für fixes j gegeben durch*

$$\boxed{A_{n,j}^{[C]} = 2\log n + \mathcal{O}(1)\,.} \tag{1.256}$$

Korollar 1.9.6. *Der Erwartungswert $A_{n,j}^{[C]}$ für die Anzahl der Schlüsselvergleiche, die beim Aufteilen eines p-trees der Größe n mit dem Algorithmus* `Cut` *durch ein Element mit Schlüssel $j + \frac{1}{2}$ gemacht werden, ist asymptotisch für fixes $l = n - j$ gegeben durch*

$$\boxed{A_{n,j}^{[C]} = \frac{1}{3}H_l^2 + \frac{7}{3}H_l - \frac{1}{3}H_l^{(2)} + \frac{1}{3(l+1)} + \mathcal{O}(1/n)\,.} \tag{1.257}$$

Weiters ergibt sich für das arithmetische Mittel $A_n^{[C]} = \frac{1}{n+1}\sum_{j=0}^{n} A_{n,j}^{[C]}$ der Anzahl an Schlüsselvergleichen beim Aufteilen eines p-trees durch einen zufällig gewählten Knoten unter Verwendung der Beziehungen (1.85) und (1.117) durch Aufsummieren von (1.250)

Korollar 1.9.7. *Der Erwartungswert $A_n^{[C]}$ für die Anzahl der Schlüsselvergleiche, die beim Aufteilen eines p-trees der Größe n mit dem Algorithmus* `Cut` *durch ein zufällig gewähltes Element gemacht werden, ist gegeben durch*

$$\boxed{\begin{aligned} &A_n^{[C]} = \frac{1}{3}H_{n+1}^2 + \frac{10}{9}H_{n+1} - \frac{1}{3}H_{n+1}^{(2)} - \frac{19n+46}{27(n+1)} \quad \text{für } n \geq 2\,,\\ &A_1^{[C]} = 1\,. \end{aligned}} \tag{1.258}$$

	j fest	$j \sim \rho n$	$l := n - j$ fest
Erwartungswert $E_{n,j}$ der Höhe eines Knotens	$2 \log n$	$\frac{1}{3} \log^2 n$	$\frac{1}{3}H_{l+1}^2+\frac{5}{3}H_{l+1}-\frac{1}{3}H_{l+1}^{(2)}-1+\frac{1}{3}\frac{1}{l+1}$
Erwartungswert $E_{n,j}$ der linken Kanten	$2 \log n$	$\frac{1}{3} \log^2 n$	$\frac{1}{3}H_l^2+H_l-\frac{5}{3}-\frac{1}{3}H_l^{(2)}+\frac{2}{3}\frac{H_l}{l+1}+\frac{5}{3}\frac{1}{l+1}$
Erwartungswert $E_{n,j}$ der rechten Kanten	$\frac{2}{3}H_j - \frac{1}{3} - \frac{1}{3j}$	$\frac{2}{3} \log n$	$\frac{2}{3}H_l - \frac{1}{3} + \frac{1}{3}\frac{1}{l+1}$
Erwartungswert $E_{n,j}$ der Anzahl der Nachfolger	$\frac{1}{3}H_j^2-\frac{1}{6}H_j+\frac{2}{3}H_j^{(2)}+\frac{1}{6}$	$\frac{1}{3} \log^2 n$	$\frac{1}{l+1}n$
Erwartungswert $E_{n,j}$ der Vergleiche beim Einfügen	$2 \log n$	$\frac{1}{3} \log^2 n$	$\frac{1}{3}H_l^2+\frac{7}{3}H_l-\frac{1}{3}-\frac{1}{3}H_l^{(2)}+\frac{1}{3l}+\frac{1}{3(l+1)}$
Erwartungswert $E_{n,j}$ der Vergleiche beim Aufteilen	$2 \log n$	$\frac{1}{3} \log^2 n$	$\frac{1}{3}H_l^2+\frac{7}{3}H_l-\frac{1}{3}H_l^{(2)}+\frac{1}{3(l+1)}$

Länge des linken Pfades	Grenzverteilung	normalverteilt
	Erwartungswert	$2 \log n$
	Varianz	$2 \log n$
Anzahl der Blätter	Grenzverteilung	normalverteilt
	Erwartungswert	$\frac{1}{3}n$
	Varianz	$\frac{2}{45}n$
Anzahl der Knoten mit 1 Kind	Grenzverteilung	normalverteilt
	Erwartungswert	$\frac{1}{3}n$
	Varianz	$\frac{2}{45}n$
Anzahl der Knoten mit 2 Kindern	Grenzverteilung	normalverteilt
	Erwartungswert	$\frac{1}{3}n$
	Varianz	$\frac{2}{45}n$

Tabelle 1.1: Asymptotische Resultate zu allen expliziten Formeln dieses Kapitels

Kapitel 2

Über locally balanced binary search trees

2.1 Allgemeines

In einem erst kürzlich erschienen Artikel von Martínez und Prodinger [30] wurden zahlreiche exakte und asymptotische Ergebnisse für die Nachkommen beziehungsweise die Höhe von Knoten in binary search trees (BSTs) und locally balanced binary search trees (LBSTs) veröffentlicht.

Wie hinreichend bekannt ist, versteht man unter Binären Suchbäumen Binärbäume, bei denen jedem Knoten p genau ein Schlüssel k zugeordnet ist, und bei der die Prozedur `Insert` zum Einfügen eines neuen Knotens p mit Schlüssel k in einen Binären Suchbaum T rekursiv so definiert ist:

1. Ist T der leere Baum, dann wird dieser Knoten zur Wurzel eines aus einem Knoten bestehenden Baumes.
2. Ansonsten wird der Schlüssel k von p mit dem Schlüssel l der Wurzel w von T verglichen.
   (a) Ist $k \leq l$, dann füge p mittels `Insert` in den linken Unterbaum von w ein.
   (b) Ist $k > l$, dann füge p mittels `Insert` in den rechten Unterbaum von w ein.

Wird ein Binärer Suchbaum der Größe n aus einer zufälligen Permutation der Zahlen von $1, 2, \ldots, n$ erzeugt, das heißt jede mögliche Permutation der Zahlen tritt mit der gleichen Wahrscheinlichkeit $\frac{1}{n!}$ auf, dann spricht man von random binary search trees.

Die Konstruktion der random LBSTs, wie sie in [30] gegeben ist, soll hier ebenfalls kurz wiedergegeben werden.

1. Für $n \leq 2$ sind die random LBSTs genau die random binary search trees.
2. Ein Binärer Suchbaum T der Größe $n \geq 3$ mit linken und rechten Unterbäumen T_1 und T_2 ist genau dann ein random LBST, wenn T_1 und T_2 voneinander unabhängige

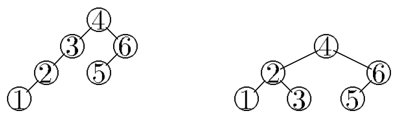


Abbildung 2.1: Links der binary search tree und rechts der locally balanced binary search tree, welche beide aus der Permutation 4 6 3 2 1 5 erzeugt wurden.

random LBSTs sind, und die Wahrscheinlichkeit $\pi_{n,k}$, daß im Baum der Größe n der Knoten mit Schlüssel k die Wurzel bildet, ist gegeben durch

$$\pi_{n,k} = \frac{(k-1)(n-k)}{\binom{n}{3}} \ , \quad \text{für alle } 1 \le k \le n \ .$$

Anders betrachtet unterscheiden sich die random BSTs und die random LBSTs nur durch die Wahrscheinlichkeiten $\pi_{n,k}$, daß k die Wurzel in einem Baum der Größe n bildet. Bei random BSTs betragen sie unabhängig von k immer $\frac{1}{n}$, während diese bei random LBSTs die Wahrscheinlichkeiten angeben, daß k der Median von 3 beliebigen Elementen aus $\{1, 2, \ldots, n\}$ ist.

Man kann die Erzeugung der random LBSTs auch beschreiben als aus einer Zufallspermutation generierte BSTs, die lokal ausbalanziert werden in der Art, daß beim Einfügen eines neuen Elementes in den Baum ein eventuell entstehender Unterbaum der Größe 3 so umgewandelt wird, daß der Median der drei Elemente als Wurzel der beiden anderen Elemente fungiert [36]. Vergleiche dazu mit Abbildung 2.1.

2.2 Die Erwartungswerte der Höhe eines Knotens

Es folgt hier kurz die Herleitung eines der Ergebnisse von [30], da dieses in der Folge benötigt wird.

Bezeichnen wir mit $F_{n,j,m}$ die Wahrscheinlichkeit, daß der Knoten mit Schlüssel j in einem random LBST der Größe n genau die Höhe m hat. Unter der Höhe eines Knotens p versteht man hier wiederum die Anzahl der Knoten, die auf dem kürzesten Weg von der Wurzel des Baumes zu p liegen.

Aus der rekursiven Definition der LBSTs erhält man nun durch Betrachten des Schlüssels k der Wurzel und Unterscheiden der möglichen Fälle $k = j$, $k < j$ und $k > j$ als Rekursion für die $F_{n,j,m}$:

$$\begin{aligned} F_{n,j,m} &= \pi_{n,j}\delta_{m,1} + \sum_{k=1}^{j-1} \pi_{n,k} F_{n-k,j-k,m-1} + \sum_{k=j+1}^{n} \pi_{n,k} F_{k-1,j,m-1} \ , \\ &\text{für } n \ge 3 \ , \ n \ge j \ge 1 \text{ und } m \ge 1 \ , \\ F_{1,1,1} &= 1 \ , \ F_{2,1,1} = \frac{1}{2} \ , \ F_{2,1,2} = \frac{1}{2} \ , \ F_{2,2,1} = \frac{1}{2} \ , \ F_{2,2,2} = \frac{1}{2} \ . \end{aligned} \tag{2.1}$$

Definieren wir die erzeugende Funktion in drei Variablen $F(z,u,v)$ durch

$$F(z,u,v) = \sum_{n,j,m} nF_{n,j,m} z^{n-1} u^j v^m \,, \tag{2.2}$$

dann ergibt sich aus der Rekursion (2.1) durch Multiplizieren mit $\binom{n}{3} z^{n-3} u^j v^m$ und Aufsummieren über alle $n \geq 3$, $1 \leq j \leq n$ und $m \geq 1$ zunächst

$$\begin{aligned}
\sum_{n\geq 3}\sum_{j=1}^{n}\sum_{m\geq 1}\binom{n}{3} F_{n,j,m} z^{n-3}u^j v^m &= \sum_{n\geq 3}\sum_{j=1}^{n}\sum_{m\geq 1}(j-1)(n-j)\delta_{m,1} z^{n-3}u^j v^m \\
&+ \sum_{n\geq 3}\sum_{j=1}^{n}\sum_{m\geq 1}\sum_{k=1}^{j-1}(k-1)(n-k)F_{n-k,j-k,m-1} z^{n-3}u^j v^m \\
&+ \sum_{n\geq 3}\sum_{j=1}^{n}\sum_{m\geq 1}\sum_{k=j+1}^{n}(k-1)(n-k)F_{k-1,j,m-1} z^{n-3}u^j v^m \,.
\end{aligned} \tag{2.3}$$

Die einzelnen Summanden werden vereinfacht zu

$$\begin{aligned}
&\sum_{n\geq 3}\sum_{j=1}^{n}\sum_{m\geq 1}\binom{n}{3} F_{n,j,m} z^{n-3}u^j v^m = \frac{1}{6}\sum_{m\geq 1}\sum_{j\geq 1}\sum_{n\geq j} nF_{n,j,m}(n-1)(n-2) z^{n-3}u^j v^m \\
&\quad = \frac{1}{6}\frac{\partial^2}{\partial z^2}F(z,u,v)\,, \\[2ex]
&\sum_{n\geq 3}\sum_{j=1}^{n}\sum_{m\geq 1}(j-1)(n-j)\delta_{m,1} z^{n-3}u^j v^m = v\sum_{n\geq 3}\sum_{j=1}^{n}(j-1)(n-j)z^{n-3}u^j \\
&\quad = v\sum_{j\geq 1}(j-1)z^{j-2}u^j\sum_{n\geq j}(n-j)z^{n-j-1} = \frac{u^2 v}{(1-z)^2(1-uz)^2}\,, \\[2ex]
&\sum_{n\geq 3}\sum_{j=1}^{n}\sum_{m\geq 1}\sum_{k=1}^{j-1}(k-1)(n-k)F_{n-k,j-k,m-1} z^{n-3}u^j v^m = \\
&\quad = v\sum_{k\geq 1}(k-1)z^{k-2}u^k\sum_{m\geq 1}\sum_{j-k\geq 1}\sum_{n-k\geq j-k}(n-k)F_{n-k,j-k,m-1} z^{n-k-1}u^{j-k}v^{m-1} \\
&\quad = \frac{u^2 v}{(1-uz)^2}F(z,u,v)\,, \\[2ex]
&\sum_{n\geq 3}\sum_{j=1}^{n}\sum_{m\geq 1}\sum_{k=j+1}^{n}(k-1)(n-k)F_{k-1,j,m-1} z^{n-3}u^j v^m = \\
&\quad = v\sum_{m\geq 1}\sum_{j\geq 1}\sum_{k\geq j+1}(k-1)F_{k-1,j,m-1} z^{k-2}u^j v^{m-1}\sum_{n\geq k}(n-k)z^{n-k-1} = \\
&\quad = \frac{v}{(1-z)^2}F(z,u,v)\,.
\end{aligned} \tag{2.4}$$

Damit erhält man aus (2.3) folgende Differentialgleichung für $F(z,u,v)$:

$$\frac{1}{6}\frac{\partial^2}{\partial z^2}F(z,u,v) = \frac{v}{(1-z)^2}F(z,u,v) + \frac{u^2 v}{(1-uz)^2}F(z,u,v) + \frac{u^2 v}{(1-z)^2(1-uz)^2}\,,$$
$$F(0,u,v) = uv\,, \quad \frac{\partial}{\partial z}F(z,u,v)\bigg|_{z=0} = uv(1+v)(1+u)\,. \tag{2.5}$$

Nun schreiben wir abkürzend $E_{n,j}$ für die Erwartungswerte der Höhe des Knotens mit Schlüssel j in einem random LBST der Größe n und definieren weiters die Funktion $G_1(z,u)$ durch

$$G_1(z,u) = \sum_{n,j} nE_{n,j} z^{n-1} u^j\,.$$

Da definitionsgemäß $E_{n,j} = \sum_m mF_{n,j,m}$ gilt, folgt zunächst

$$G_1(z,u) = \frac{\partial}{\partial v}F(z,u,v)\bigg|_{v=1}\,.$$

Somit bekommt man aus (2.5) durch Ableiten nach v und anschließendes Auswerten an der Stelle $v=1$ folgende lineare Differentialgleichung zweiter Ordnung für $G_1(z,u)$:

$$\frac{1}{6}\frac{\partial^2}{\partial z^2}G_1(z,u) - \left(\frac{1}{(1-z)^2} + \frac{u^2}{(1-uz)^2}\right)G_1(z,u) = \frac{u}{1-u}\left(\frac{1}{(1-z)^4} - \frac{u^3}{(1-uz)^4}\right)\,,$$
$$G_1(0,u) = u \quad \text{und} \quad \frac{\partial}{\partial z}G_1(z,u)\bigg|_{z=0} = 3u(1+u)\,. \tag{2.6}$$

Als Lösung der entsprechenden homogenen Differentialgleichung

$$\frac{1}{6}\frac{\partial^2}{\partial z^2}G_1^{[h]}(z,u) - \left(\frac{1}{(1-z)^2} + \frac{u^2}{(1-uz)^2}\right)G_1^{[h]}(z,u) = 0 \tag{2.7}$$

ergibt sich

$$G_1^{[h]}(z,u) = c_1(u)\frac{2uz-u-1}{(1-z)^2(1-uz)^2} \tag{2.8}$$
$$+\, c_2(u)\frac{(5z^3u^3 + 5z^2u^3 - 20z^2u^2 + 3zu^3 - 16zu^2 + 28zu + u^3 - 6u^2 + 14u - 14)(1-z)^2}{(1-uz)^2}\,,$$

mit zwei nur von u abhängigen Funktionen $c_1(u)$ und $c_2(u)$.

Mittels Variation der Konstanten ermittelt man daraus eine Partikulärlösung für (2.6), woraus sich schließlich durch Anpassen der inhomogenen Lösung an die Anfangsbedingungen die Lösung der gesuchten Differentialgleichung ergibt:

$$G_1(z,u) = \frac{D_1(z,u)}{(1-z)^2(1-u)^7(1-uz)^2} + \frac{D_2(z,u)}{(1-z)^2(1-u)^7(1-uz)^2}\log\left(\frac{1}{1-z}\right)$$
$$+ \frac{D_3(z,u)}{(1-z)^2(1-u)^7(1-uz)^2}\log\left(\frac{1}{1-uz}\right)\,, \tag{2.9}$$

wobei die Polynome $D_i(z,u)$ für $1 \leq i \leq 3$ folgendermaßen gegeben sind:

$$\begin{aligned}
D_1(z,u) = &-\frac{u}{350}(-1960uz^3 - 19000u^3z^3 + 19050u^3z^2 - 84uz^5 + 370z + 840uz^4 \\
&+ 4844u^3z^5 + 400u^3z^4 + 2520u^9z^4 + 1960u^9z^3 + 140u^9z^2 - 8900u^8z^3 - 1650u^8z^2 \\
&+ 336u^2z^6 - 644u^2z^5 - 2520u^2z^4 + 8900u^2z^3 - 7650u^2z^2 - 28uz^6 + 2450u \\
&- 1008u^3z^6 + 7650u^7z^2 + 1650uz^2 + 20u^2z^7 - 350 - 7350u^2 + 7350u^5 - 2450u^6 \\
&- 140z^2 - 70z^4 + 140z^3 + 14z^5 - 2340uz + 5900u^2z - 6860u^3z + 6860u^5z - 5900u^6z \\
&+ 2340u^7z - 370u^8z - 20790u^4z^2 + 5840u^4z^3 + 22630u^4z^4 - 15546u^4z^5 + 20790u^5z^2 \\
&- 17040u^5z^4 - 1528u^5z^5 - 19050u^6z^2 - 5840u^6z^3 + 17040u^6z^4 - 400u^8z^4 \\
&+ 15546u^8z^5 - 4844u^9z^5 + 12250u^3 - 12250u^4 + 350u^7 + 19000u^7z^3 \\
&- 22630u^7z^4 + 1)\,,
\end{aligned}$$

$$\begin{aligned}
D_2(z,u) = &-\frac{6u}{35}(-7 + 42u - 98u^2 - 98u^5 + 82u^6 + 14uz - 98u^2z + 294u^3z - 630u^4z \\
&+ 434u^5z - 238u^6z + 74u^7z - 10u^8z + 280u^4z^2 - 280u^4z^3 + 140u^4z^4 - 28u^4z^5 \\
&+ 280u^5z^2 - 560u^5z^3 + 560u^5z^4 - 280u^5z^5 - 280u^6z^3 + 560u^6z^4 - 504u^6z^5 \\
&- 28u^8z^5 + 98u^3 + 28u^4 - 32u^7 + 5u^8 + 140u^7z^4 - 280u^7z^5 + 224u^7z^6 \\
&+ 224u^6z^6 + 56u^8z^6 - 80u^7z^7 - 40u^8z^7 + 10u^8z^8 + 56u^5z^6 - 40u^6z^7 + 10u^7z^8)\,,
\end{aligned}$$

$$\begin{aligned}
D_3(z,u) = &-\frac{1}{35}(30 - 192u + 492u^2 + 588u^5 - 588u^6 - 60uz + 444u^2z - 1428u^3z \\
&+ 2604u^4z - 3780u^5z + 1764u^6z - 588u^7z + 84u^8z + 1680u^5z^2 - 1680u^5z^3 + 840u^5z^4 \\
&- 168u^5z^5 + 1680u^6z^2 - 3360u^6z^3 + 3360u^6z^4 - 1680u^6z^5 + 840u^8z^4 - 1680u^8z^5 \\
&- 168u^9z^5 - 588u^3 + 168u^4 + 252u^7 - 42u^8 - 1680u^7z^3 + 3360u^7z^4 - 3024u^7z^5 \\
&+ 1344u^7z^6 + 336u^6z^6 + 1344u^8z^6 + 336u^9z^6 - 240u^7z^7 - 240u^9z^7 - 480u^8z^7 \\
&+ 60u^8z^8 + 60u^9z^8)\,.
\end{aligned}$$

Liest man nun von (2.9) die Koeffizienten ab, wobei die Beziehungen aus Abschnitt 1.4 hilfreich sind, führt das mittels

$$E_{n,j} = \frac{1}{n}[z^{n-1}u^j]G_1(z,u)$$

direkt zu

Satz 2.2.1. *Die Erwartungswerte $E_{n,j}$ der Höhe des Knotens mit Schlüssel j in einem random locally balanced binary search tree der Größe n sind gegeben durch*

$$\boxed{\begin{aligned}
E_{n,j} = &\\
&\frac{24}{35}H_n + \frac{18}{35}H_j + \frac{18}{35}H_{n+1-j} + \frac{12}{35}\frac{1}{j} + \frac{12}{35}\frac{1}{n+1-j} - \frac{279}{175} - \frac{6}{7}\frac{1}{n} + \frac{18}{35}\frac{j}{n} \\
&- \frac{12}{35}\frac{(j-1)^2}{n^{\underline{2}}} - \frac{4}{35}\frac{(2j-3)(j-1)^{\underline{2}}}{n^{\underline{3}}} - \frac{6}{35}\frac{(j-2)(j-1)^{\underline{3}}}{n^{\underline{4}}} \\
&+ \frac{6}{35}\frac{(2j-5)(j-1)^{\underline{4}}}{n^{\underline{5}}} - \frac{4}{35}\frac{(j-3)(j-1)^{\underline{5}}}{n^{\underline{6}}} \\
&\text{für n} \geq 8 \text{ und } 5 \leq \text{j} \leq \text{n} - 4\,.
\end{aligned}}$$

$$\boxed{\begin{array}{l} E_{n,1} = \frac{6}{5}H_n - \frac{6}{25} \quad \text{für n} \geq 5 \; , \\ E_{1,1} = 1 \; , \; E_{2,1} = \frac{3}{2} \; , \; E_{3,1} = 2 \; , \; E_{4,1} = \frac{9}{4} \; . \end{array}}$$
$$\boxed{\begin{array}{l} E_{n,2} = \frac{6}{5}H_n - \frac{21}{5} \quad \text{für n} \geq 6 \; , \\ E_{2,2} = \frac{3}{2} \; , \; E_{3,2} = 1 \; , \; E_{4,2} = \frac{7}{4} \; , \; E_{5,2} = \frac{19}{10} \; . \end{array}}$$
$$\boxed{\begin{array}{l} E_{n,3} = \frac{6}{5}H_n - \frac{87}{175} - \frac{6}{5}\frac{1}{n^{\underline{2}}} \quad \text{für n} \geq 7 \; , \\ E_{3,3} = 2 \; , \; E_{4,3} = \frac{7}{4} \; , \; E_{5,3} = \frac{11}{5} \; , \; E_{6,3} = \frac{12}{5} \; . \end{array}}$$
$$\boxed{\begin{array}{l} E_{n,4} = \frac{6}{5}H_n - \frac{11}{25} - \frac{18}{5}\frac{1}{n^{\underline{2}}} - \frac{12}{5}\frac{1}{n^{\underline{3}}} \quad \text{für n} \geq 8 \; , \\ E_{4,4} = \frac{9}{4} \; , \; E_{5,4} = \frac{19}{10} \; , \; E_{6,4} = \frac{12}{5} \; , \; E_{7,4} = \frac{18}{7} \; . \end{array}} \tag{2.10}$$

Dabei bezeichnet $n^{\underline{k}}$ die absteigenden Faktoriellen, also

$$n^{\underline{k}} = n(n-1)\cdots(n-k+1) \; .$$

Aus diesem Resultat lassen sich leicht einige Folgerungen gewinnen. So ergibt sich für das asymptotische Verhalten von $E_{n,j}$ für ein festes Verhältnis $\rho = j/n$

Korollar 2.2.2. *Als asymptotische Entwicklung für den Erwartungswert der Höhe des Knotens j in einem random locally balanced binary search tree der Größe n für $j = \rho n$ und $0 < \rho < 1$ erhält man*

$$\boxed{E_{n,j} = \frac{12}{7}\log n + \mathcal{O}(1) \; .} \tag{2.11}$$

Weiters erhält man für das asymptotische Verhalten von $E_{n,j}$ für fixes j, also am 'linken Rand' des Baumes

Korollar 2.2.3. *Als asymptotische Entwicklung für den Erwartungswert der Höhe des Knotens j in einem random locally balanced binary search tree der Größe n für fixes j ergibt sich*

$$\boxed{E_{n,j} = \frac{6}{5}\log n + \mathcal{O}(1) \; .} \tag{2.12}$$

Aus Symmetriegründen gilt für fixes $l = n+1-j$, das heißt am 'rechten Rand' des Baumes

Korollar 2.2.4. *Als asymptotische Entwicklung für den Erwartungswert der Höhe des Knotens $l = n+1-j$ in einem random locally balanced binary search tree der Größe n für fixes l ergibt sich*

$$\boxed{E_{n,l} = \frac{6}{5}\log n + \mathcal{O}(1) \; .} \tag{2.13}$$

Bildet man das arithmetische Mittel über die $E_{n,j}$ bei festem n, sei also E_n definiert durch

$$E_n = \frac{1}{n}\sum_{j=1}^{n} E_{n,j}\,,$$

dann erhält man damit den Erwartungswert der Höhe eines zufälligen Knotens in einem Baum der Größe n. Es gilt dann

Korollar 2.2.5. *Die erwartete Höhe E_n eines zufälligen Knotens in einem locally balanced binary search tree der Größe n oder anders ausgedrückt der Erwartungswert der Suchkosten für eine erfolgreiche Suche eines zufälligen Elements in einem LBST der Größe n ist gegeben durch*

$$\boxed{\begin{array}{l} E_n = \frac{12}{7}\left(1+\frac{1}{n}\right)H_n - \frac{1}{49}\left(124-\frac{9}{n}\right) \quad \text{für n} \geq 6\,, \\ E_1 = 1\,,\ E_2 = \frac{3}{2}\,,\ E_3 = \frac{5}{3}\,,\ E_4 = 2\,,\ E_5 = \frac{11}{5}\,. \end{array}} \tag{2.14}$$

2.3 Die zweiten Faktoriellen Momente der Höhe eines Knotens

Definieren wir

$$G_2(z,u) = \frac{\partial^2}{\partial v^2}F(z,u,v)\bigg|_{v=1}\,,$$

so sind die zweiten Faktoriellen Momente $M_{n,j}^{(2)}$ der Höhe des Knotens j in einem LBST der Größe n gegeben durch die Beziehung

$$G_2(z,u) = \sum_{n,j} n M_{n,j}^{(2)} z^{n-1} u^j\,.$$

Durch zweimaliges Ableiten von (2.5) nach v und anschließendes Auswerten an der Stelle $v=1$ ergibt sich folgende Differentialgleichung für $G_2(z,u)$:

$$\begin{aligned}
&\frac{1}{6}\frac{\partial^2}{\partial z^2}G_2(z,u) - \left(\frac{1}{(1-z)^2} - \frac{u^2}{(1-uz)^2}\right)G_2(z,u) \\
&\quad = \frac{r_1(z,u)}{(1-u)^7(1-z)^4(1-uz)^4}\log\left(\frac{1}{1-z}\right) + \frac{r_2(z,u)}{(1-u)^7(1-z)^4(1-uz)^4}\log\left(\frac{1}{1-uz}\right) \\
&\quad\quad + \frac{r_3(z,u)}{(1-u)^6(1-z)^4(1-uz)^4}\,,
\end{aligned} \tag{2.15}$$

wobei die Polynome $r_1(z,u)$, $r_2(z,u)$ und $r_3(z,u)$ gegeben sind durch

$$\begin{aligned}
r_1(z,u) = &-\frac{12}{35}u\,(\,2\,u^2\,z^2 + 1 - 2\,u\,z - 2\,u^2\,z + u^2\,)(-7 + 14\,u\,z + 42\,u - 98\,u^2 - 98\,u^2\,z \\
&+ 294\,u^3\,z + 28\,u^4 + 98\,u^3 - 98\,u^5 + 82\,u^6 - 32\,u^7 + 140\,u^4\,z^4 - 28\,u^4\,z^5
\end{aligned}$$

$$
\begin{aligned}
&- 10\, u^8\, z - 238\, u^6\, z - 40\, u^8\, z^7 + 10\, u^8\, z^8 + 224\, u^7\, z^6 - 80\, u^7\, z^7 - 280\, u^5\, z^5 \\
&+ 56\, u^5\, z^6 + 140\, u^7\, z^4 - 280\, u^6\, z^3 - 28\, u^8\, z^5 + 434\, u^5\, z - 280\, u^4\, z^3 \\
&- 504\, u^6\, z^5 - 40\, u^6\, z^7 + 10\, u^7\, z^8 - 630\, u^4\, z + 280\, u^4\, z^2 + 280\, u^5\, z^2 \\
&+ 560\, u^6\, z^4 + 224\, u^6\, z^6 + 56\, u^8\, z^6 - 280\, u^7\, z^5 - 560\, u^5\, z^3 + 74\, u^7\, z \\
&+ 560\, u^5\, z^4 + 5\, u^8)\,,
\end{aligned}
$$

$$
\begin{aligned}
r_2(z,u) = {}& \frac{12}{35}(\, 2\, u^2\, z^2 + 1 - 2\, u\, z - 2\, u^2\, z + u^2\,)(5 - 10\, u\, z - 32\, u + 82\, u^2 + 74\, u^2\, z \\
&- 238\, u^3\, z + 28\, u^4 - 98\, u^3 + 98\, u^5 - 98\, u^6 + 42\, u^7 + 14\, u^8\, z + 294\, u^6\, z \\
&+ 280\, u^6\, z^2 - 80\, u^8\, z^7 + 10\, u^8\, z^8 + 224\, u^7\, z^6 - 40\, u^7\, z^7 - 28\, u^5\, z^5 - 280\, u^7\, z^3 \\
&+ 560\, u^7\, z^4 - 560\, u^6\, z^3 - 280\, u^8\, z^5 - 630\, u^5\, z - 280\, u^6\, z^5 + 434\, u^4\, z \\
&+ 280\, u^5\, z^2 + 140\, u^8\, z^4 + 560\, u^6\, z^4 + 56\, u^6\, z^6 + 224\, u^8\, z^6 - 504\, u^7\, z^5 \\
&- 280\, u^5\, z^3 - 98\, u^7\, z + 140\, u^5\, z^4 - 7\, u^8 - 40\, u^9\, z^7 + 10\, u^9\, z^8 - 28\, u^9\, z^5 \\
&+ 56\, u^9\, z^6)\,,
\end{aligned}
$$

$$
\begin{aligned}
r_3(z,u) = {}& \frac{1}{175} u(-370\, z + 1270\, u\, z - 2100\, u + 960\, u^3\, z^7 + 101\, u^3\, z^8 + 5600\, u^2 \\
&+ 3780\, u^2\, z^2 + 140\, z^2 - 800\, u^2\, z - 770\, u\, z^2 - 5220\, u^2\, z^3 + 1540\, u\, z^3 - 1260\, u^2\, z^4 \\
&+ 8420\, u^3\, z^3 - 14700\, u^3\, z^2 + 56\, u\, z^6 - 70\, u\, z^5 - 280\, u^2\, z^6 - 104\, u^2\, z^7 - 140\, z^3 \\
&- 1400\, u^3\, z + 3780\, u^8\, z^2 - 14700\, u^7\, z^2 - 2380\, u^3\, z^5 + 8080\, u^3\, z^4 - 5220\, u^9\, z^3 \\
&+ 8420\, u^8\, z^3 + 10500\, u^4 - 9100\, u^3 - 1260\, u^{10}\, z^4 + 1820\, u^{11}\, z^5 - 2380\, u^{11}\, z^6 \\
&+ 10304\, u^{10}\, z^6 + 8080\, u^9\, z^4 - 490\, u\, z^4 - 9100\, u^5 + 5600\, u^6 - 2380\, u^{10}\, z^5 \\
&- 2100\, u^7 + 350 - 2600\, z^9\, u^{11} + 660\, z^{10}\, u^{11} - 16930\, u^4\, z^4 + 3295\, z^8\, u^{11} \\
&- 416\, u^{11}\, z^7 - 7800\, u^{10}\, z^7 - 2620\, z^8\, u^{10} - 416\, u^4\, z^7 + 10304\, u^4\, z^6 \\
&- 8230\, u^4\, z^5 - 14\, z^5 + 70\, z^4 - 50\, u^4\, z^9 - 1166\, u^4\, z^8 + 1270\, u^8\, z - 1400\, u^6\, z \\
&+ 26520\, u^6\, z^2 + 62864\, u^8\, z^7 - 23852\, u^8\, z^8 - 81480\, u^7\, z^6 + 62864\, u^7\, z^7 \\
&+ 3295\, u^5\, z^8 + 650\, u^5\, z^9 + 72514\, u^5\, z^5 + 10\, u^5\, z^{10} - 7800\, u^5\, z^7 \\
&- 21784\, u^5\, z^6 + 4680\, u^7\, z^3 - 64710\, u^7\, z^4 + 33920\, u^6\, z^3 + 72514\, u^8\, z^5 \\
&- 2620\, u^6\, z^8 + 2500\, u^5\, z + 4680\, u^4\, z^3 + 87560\, u^6\, z^5 + 45296\, u^6\, z^7 \\
&- 30884\, u^7\, z^8 + 2500\, u^4\, z + 26520\, u^4\, z^2 - 51540\, u^5\, z^2 - 16930\, u^8\, z^4 \\
&- 51120\, u^6\, z^4 - 2600\, u^6\, z^9 - 140\, u^6\, z^{10} - 96376\, u^6\, z^6 - 96376\, u^8\, z^6 \\
&+ 87560\, u^7\, z^5 + 33920\, u^5\, z^3 - 800\, u^7\, z - 64710\, u^5\, z^4 + 70\, u^{12}\, z^4 + 56\, u^{13}\, z^6 \\
&- 140\, u^{11}\, z^3 + 350\, u^8 + 101\, u^{13}\, z^8 - 104\, u^{13}\, z^7 + 45296\, u^9\, z^7 - 30884\, u^9\, z^8 \\
&+ 960\, u^{12}\, z^7 - 1166\, u^{12}\, z^8 - 8230\, u^9\, z^5 + 20\, u^9\, z^{10} - 21784\, u^9\, z^6 \\
&+ 7400\, u^9\, z^9 - 490\, u^{11}\, z^4 + 650\, u^{12}\, z^9 - 140\, u^{12}\, z^{10} - 50\, u^{13}\, z^9 + 10\, u^{13}\, z^{10} \\
&+ 1820\, u^2\, z^5 - 2380\, u^3\, z^6 + 140\, u^{10}\, z^2 + 1540\, u^{10}\, z^3 - 1740\, z^{10}\, u^{10} \\
&- 770\, u^9\, z^2 + 5400\, u^7\, z^9 + 660\, u^7\, z^{10} + 7400\, u^8\, z^9 - 1740\, u^8\, z^{10} \\
&+ 5400\, u^{10}\, z^9 - 14\, u^{13}\, z^5 - 70\, u^{12}\, z^5 - 280\, u^{12}\, z^6 - 370\, u^9\, z)\,.
\end{aligned}
$$

Diese Differentialgleichung hat als homogene Differentialgleichung ebenfalls (2.7) und besitzt daher (2.8) als Lösung derselben. Aus dieser gewinnt man wiederum durch Variation der Konstanten eine Partikulärlösung und schließlich durch Anpassen der inhomogenen Lösung an die Anfangswerte die Lösung obiger Differentialgleichung (2.15):

$$
\begin{aligned}
G_2(z,u) = \\
&+ \frac{D_1(z,u)}{(1-u)^7} \int_0^z \frac{1}{1-ut} \log\left(\frac{1}{1-t}\right) dt \\
&+ \left(\frac{72}{175}\,\frac{u^2\,(-2+5\,u)}{(1-z\,u)^2} - \frac{72}{1225}\,\frac{u-20}{(1-z)^2}\right) \int_0^z \frac{1}{(1-t)(1-ut)} \log\left(\frac{1}{1-t}\right) dt \\
&+ \frac{D_2(z,u)}{(1-u)^7} \int_0^z \frac{1}{1-t} \log\left(\frac{1}{1-ut}\right) dt \\
&+ \left(\frac{72}{1225}\,\frac{u^2\,(20\,u-1)}{(1-z\,u)^2} - \frac{72}{175}\,\frac{2\,u-5}{(1-z)^2}\right) \int_0^z \frac{1}{(1-t)(1-ut)} \log\left(\frac{1}{1-ut}\right) dt \\
&+ \left(\frac{D_3(z,u)}{u^2(1-u)^7} - \frac{72}{1225}\,\frac{(20\,u-1)\,u^2}{(1-uz)(1-u)^2} - \frac{6}{875}\,\frac{48\,u^4+240\,u^3-80\,u^2+15\,u-1}{(1-uz)^2(1-u)u^2}\right. \\
&\qquad \left. + \frac{72}{175}\,\frac{(2\,u+5)\,u}{(1-z)(1-u)^2} + \frac{D_4(z,u)}{(1-z)^2(1-u)u^4}\right) \log\left(\frac{1}{1-uz}\right) \\
&+ \left(\frac{D_5(z,u)}{(1-u)^6} - \frac{72}{245}\,\frac{(2\,u-5)\,u^2}{(1-uz)^2(1-u)} - \frac{72}{35}\,\frac{1}{(1-z)^2(1-u)}\right) \log^2\left(\frac{1}{1-uz}\right) \\
&+ \left(\frac{D_6(z,u)}{(1-u)^6} + \frac{72}{35}\,\frac{u^3}{(1-uz)^2(1-u)} - \frac{72}{245}\,\frac{-2+5\,u}{(1-z)^2(1-u)}\right) \log^2\left(\frac{1}{1-z}\right) \\
&+ \left(\frac{D_7(z,u)}{(1-u)^7} + \frac{72}{1225}\,\frac{(u-20)\,u}{(1-z)(1-u)^2} - \frac{6}{875}\,\frac{(u^4-15\,u^3+80\,u^2-240\,u-48)\,u}{(1-z)^2(1-u)}\right. \\
&\qquad \left. + \frac{72}{175}\,\frac{(5\,u+2)\,u^2}{(1-uz)(1-u)^2} + \frac{D_8(z,u)}{(1-uz)^2(1-u)u^2}\right) \log\left(\frac{1}{1-z}\right) \\
&+ \frac{D_9(z,u)}{(1-u)^6} + \frac{D_{10}(z,u)}{(1-z)(1-u)^2u^3} + \frac{D_{11}(z,u)}{(1-z)^2(1-u)u^3} + \frac{D_{12}(z,u)}{(1-uz)(1-u)^2u^2} \\
&+ \frac{D_{13}(z,u)}{(1-uz)^2(1-u)u^2}\,,
\end{aligned}
\tag{2.16}
$$

wobei die Polynome $D_i(z,u)$ für $1 \le i \le 13$ gegeben sind durch:

$$
\begin{aligned}
D_1(z,u) = {} & -\frac{2448}{245}\,z^4\,u^8 - \frac{432}{245}\,z^4\,u^7 + \frac{4896}{245}\,z^3\,u^8 + \frac{1152}{49}\,z^3\,u^7 - \frac{7344}{1225}\,z^2\,u^8 \\
& + \frac{864}{245}\,z^3\,u^6 - \frac{60048}{1225}\,z^2\,u^7 - \frac{4896}{1225}\,z\,u^8 - \frac{17712}{1225}\,z^2\,u^6 + \frac{28512}{1225}\,z\,u^7 - \frac{2448}{1225}\,u^8 \\
& - \frac{1296}{1225}\,z^2\,u^5 + \frac{6912}{245}\,z\,u^6 + \frac{14256}{1225}\,u^7 + \frac{288}{1225}\,z\,u^5 - \frac{34128}{1225}\,u^6 - \frac{864}{1225}\,z\,u^4 \\
& + \frac{8208}{1225}\,u^5 + \frac{144}{1225}\,u^4 - \frac{432}{1225}\,u^3\,,
\end{aligned}
$$

$$
\begin{aligned}
D_2(z,u) = {} & \frac{432}{245}\,z^4\,u^7 + \frac{2448}{245}\,z^4\,u^6 - \frac{864}{245}\,z^3\,u^7 - \frac{1152}{49}\,z^3\,u^6 + \frac{1296}{1225}\,z^2\,u^7 \\
& - \frac{4896}{245}\,z^3\,u^5 + \frac{17712}{1225}\,z^2\,u^6 + \frac{864}{1225}\,z\,u^7 + \frac{60048}{1225}\,z^2\,u^5 - \frac{288}{1225}\,z\,u^6 + \frac{432}{1225}\,u^7
\end{aligned}
$$

$$+\frac{7344}{1225}z^2u^4-\frac{6912}{245}z\,u^5-\frac{144}{1225}u^6-\frac{28512}{1225}z\,u^4-\frac{8208}{1225}u^5+\frac{4896}{1225}z\,u^3$$
$$+\frac{34128}{1225}u^4-\frac{14256}{1225}u^3+\frac{2448}{1225}u^2\,,$$

$$\begin{aligned}D_3(z,u)=&-\frac{6}{42875}(-21+9065\,u^{11}+141047\,u^9-46599\,u^{10}-10760\,z^4u^8+25200\,z^4u^7\\&-21360\,z^3u^8-79280\,z^3u^7+128376\,z^2u^8-33600\,z^3u^6-16944\,z^2u^7\\&-7600\,z\,u^8+183384\,z^2u^6+75168\,z\,u^7-24255\,z^2u^5-179520\,z\,u^6\\&-38954\,z\,u^5+26418\,z\,u^4-8400\,z^4u^6+13650\,z^3u^5+2457\,z^2u^4\\&-6888\,z\,u^3+50840\,z^4u^9-3675\,u^{12}z^4-105\,z^4u^4+1575\,z^4u^5+245\,u^{13}z^4\\&-58800\,z^4u^{10}+19600\,z^4u^{11}+210\,z^3u^3+7520\,z^3u^9-490\,u^{13}z^3\\&-31850\,z^3u^{11}+6860\,u^{12}z^3-2940\,z^3u^4+78400\,z^3u^{10}-63\,z^2u^2\\&+56595\,z^2u^{10}-1029\,u^{12}z^2+147\,u^{13}z^2-5733\,z^2u^{11}-229056\,z^2u^9\\&+441\,z^2u^3-2058\,u^{12}z+16072\,z\,u^{11}-61642\,z\,u^{10}+115186\,z\,u^9\\&+882\,z\,u^2-240575\,u^8+191752\,u^7-145368\,u^6+165215\,u^5-74343\,u^4\\&+19971\,u^3-3885\,u^2+98\,u^{13}z+441\,u-1029\,u^{12}+49\,u^{13}-42\,z\,u)\,,\end{aligned}$$

$$\begin{aligned}D_4(z,u)=&\frac{6}{42875}(49\,u^9-735\,u^8+3920\,u^7-11760\,u^6+5548\,u^5-4420\,u^4+16800\,u^3\\&-5600\,u^2+1050\,u-70)\,,\end{aligned}$$

$$\begin{aligned}D_5(z,u)=&\frac{72}{175}(5\,z^4u^4-10\,z^3u^3-10\,z^3u^4+3\,z^2u^4+3\,z^2u^2+24\,z^2u^3+2\,z\,u\\&-12\,z\,u^3+2\,z\,u^4-12\,z\,u^2-6\,u^3+15\,u^2-6\,u+u^4+1)u^2\,,\end{aligned}$$

$$\begin{aligned}D_6(z,u)=&\frac{72}{175}(5\,z^4u^4-10\,z^3u^3-10\,z^3u^4+3\,z^2u^4+3\,z^2u^2+24\,z^2u^3+2\,z\,u\\&-12\,z\,u^3+2\,z\,u^4-12\,z\,u^2-6\,u^3+15\,u^2-6\,u+u^4+1)u^2\,,\end{aligned}$$

$$\begin{aligned}D_7(z,u)=&-\frac{18}{6125}u^{11}-\frac{666}{1225}u^9+\frac{54}{875}u^{10}+\frac{864}{245}z^4u^8-\frac{12912}{8575}z^4u^7-\frac{1152}{245}z^3u^8\\&-\frac{95136}{8575}z^3u^7-\frac{594}{175}z^2u^8-\frac{25632}{8575}z^3u^6+\frac{1100304}{42875}z^2u^7+\frac{22644}{6125}z\,u^8\\&-\frac{101664}{42875}z^2u^6-\frac{233724}{42875}z\,u^7+\frac{770256}{42875}z^2u^5-\frac{215424}{8575}z\,u^6+\frac{451008}{42875}z\,u^5\\&-\frac{1824}{1715}z\,u^4+\frac{61008}{8575}z^4u^6+\frac{9024}{8575}z^3u^5-\frac{1374336}{42875}z^2u^4+\frac{691116}{42875}z\,u^3+\frac{96}{35}z^4u^4\\&-\frac{288}{245}z^4u^9-\frac{288}{35}z^4u^5+\frac{54}{245}z^4u^{10}-\frac{18}{1225}z^4u^{11}-\frac{156}{35}z^3u^3+\frac{468}{245}z^3u^9\\&+\frac{36}{1225}z^3u^{11}+\frac{384}{35}z^3u^4-\frac{72}{175}z^3u^{10}-\frac{702}{875}z^2u^2+\frac{54}{875}z^2u^{10}-\frac{54}{6125}z^2u^{11}\\&+\frac{2106}{6125}z^2u^9+\frac{198}{25}z^2u^3-\frac{36}{6125}z\,u^{11}+\frac{108}{875}z\,u^{10}-\frac{5904}{6125}z\,u^9-\frac{7548}{875}z\,u^2\\&+\frac{17118}{6125}u^8-\frac{446058}{42875}u^7+\frac{198258}{8575}u^6-\frac{872208}{42875}u^5+\frac{1150512}{42875}u^4-\frac{57738}{1715}u^3\end{aligned}$$

$$
\begin{aligned}
&+ \frac{846282}{42875}\,u^2 - \frac{18}{35}\,z^4\,u^3 + \frac{6}{175}\,z^4\,u^2 + \frac{24}{25}\,z^3\,u^2 - \frac{12}{175}\,z^3\,u - \frac{18}{125}\,z^2\,u + \frac{12}{875}\,\frac{z}{u} \\
&- \frac{18}{125}\,\frac{1}{u} + \frac{6}{875}\,\frac{1}{u^2} + \frac{222}{175} - \frac{36}{125}\,z - \frac{5706}{875}\,u + \frac{18}{875}\,z^2 + \frac{1968}{875}\,z\,u\,,
\end{aligned}
$$

$$
\begin{aligned}
D_8(z,u) = {} & \frac{6}{42875}(70\,u^9 - 1050\,u^8 + 5600\,u^7 - 16800\,u^6 + 4420\,u^5 - 5548\,u^4 \\
&+ 11760\,u^3 - 3920\,u^2 + 735\,u - 49)\,,
\end{aligned}
$$

$$
\begin{aligned}
D_9(z,u) = {} & \frac{272}{1715}\,u^9 - \frac{2407}{214375}\,u^{10} - \frac{46092}{42875}\,z^4\,u^8 + \frac{192578}{42875}\,z^4\,u^7 \\
&+ \frac{35988}{42875}\,z^3\,u^8 - \frac{192172}{42875}\,z^3\,u^7 + \frac{58542}{42875}\,z^2\,u^8 - \frac{172044}{6125}\,z^3\,u^6 - \frac{172836}{42875}\,z^2\,u^7 \\
&- \frac{64476}{214375}\,z\,u^8 + \frac{2920044}{214375}\,z^2\,u^6 + \frac{28548}{42875}\,z\,u^7 + \frac{9099852}{214375}\,z^2\,u^5 + \frac{649876}{214375}\,z\,u^6 \\
&- \frac{3037236}{214375}\,z\,u^5 - \frac{3037236}{214375}\,z\,u^4 + \frac{601456}{42875}\,z^4\,u^6 - \frac{172044}{6125}\,z^3\,u^5 \\
&+ \frac{2920044}{214375}\,z^2\,u^4 + \frac{649876}{214375}\,z\,u^3 + \frac{1614}{6125}\,z^4\,u^9 - \frac{46092}{42875}\,z^4\,u^4 + \frac{192578}{42875}\,z^4\,u^5 \\
&- \frac{1672}{42875}\,z^4\,u^{10} + \frac{35988}{42875}\,z^3\,u^3 - \frac{12952}{42875}\,z^3\,u^9 - \frac{192172}{42875}\,z^3\,u^4 + \frac{2924}{42875}\,z^3\,u^{10} \\
&+ \frac{58542}{42875}\,z^2\,u^2 - \frac{2286}{214375}\,z^2\,u^{10} - \frac{43614}{214375}\,z^2\,u^9 - \frac{172836}{42875}\,z^2\,u^3 - \frac{2784}{214375}\,z\,u^{10} \\
&+ \frac{5536}{42875}\,z\,u^9 + \frac{28548}{42875}\,z\,u^2 - \frac{44158}{42875}\,u^8 + \frac{931502}{214375}\,u^7 - \frac{1325557}{214375}\,u^6 \\
&- \frac{311998}{214375}\,u^5 + \frac{14324}{1715}\,u^4 - \frac{311998}{214375}\,u^3 - \frac{1325557}{214375}\,u^2 + \frac{1614}{6125}\,z^4\,u^3 \\
&- \frac{1672}{42875}\,z^4\,u^2 - \frac{12952}{42875}\,z^3\,u^2 + \frac{2924}{42875}\,z^3\,u - \frac{43614}{214375}\,z^2\,u - \frac{2784}{214375}\,\frac{z}{u} \\
&+ \frac{272}{1715}\,\frac{1}{u} - \frac{2407}{214375}\,\frac{1}{u^2} - \frac{44158}{42875} + \frac{5536}{42875}\,z + \frac{931502}{214375}\,u - \frac{2286}{214375}\,z^2 \\
&- \frac{64476}{214375}\,z\,u\,,
\end{aligned}
$$

$$
\begin{aligned}
D_{10}(z,u) = {} & -\frac{6}{42875}(49\,u^9 - 735\,u^8 + 3920\,u^7 - 11760\,u^6 - 752\,u^5 \\
&+ 18680\,u^4 + 16800\,u^3 - 5600\,u^2 + 1050\,u - 70)\,,
\end{aligned}
$$

$$
\begin{aligned}
D_{11}(z,u) = {} & \frac{1}{214375}(2827\,u^8 - 28195\,u^7 + 119340\,u^6 - 362920\,u^5 - 779633\,u^4 \\
&- 429975\,u^3 + 152950\,u^2 - 30450\,u + 2100)\,,
\end{aligned}
$$

$$
\begin{aligned}
D_{12}(z,u) = {} & \frac{6}{42875}(70\,u^9 - 1050\,u^8 + 5600\,u^7 - 16800\,u^6 - 18680\,u^5 + 752\,u^4 \\
&+ 11760\,u^3 - 3920\,u^2 + 735\,u - 49)\,,
\end{aligned}
$$

$$
\begin{aligned}
D_{13}(z,u) = {} & -\frac{1}{214375}(2100\,u^8 - 30450\,u^7 + 152950\,u^6 - 429975\,u^5 - 779633\,u^4 \\
&- 362920\,u^3 + 119340\,u^2 - 28195\,u + 2827)\,.
\end{aligned}
$$

Das Ablesen der Koeffizienten dieses Ausdrucks, wofür im wesentlichen die Beziehungen von Abschnitt 1.4 genügen, liefert aufgrund von

$$M_{n,j}^{(2)} = \frac{1}{n}[z^{n-1}u^j]G_2(z,u)$$

die gesuchten zweiten Faktoriellen Momente und man erhält

Hilfssatz 2.3.1. *Die zweiten Faktoriellen Momente $M_{n,j}^{(2)}$ der Höhe des Knotens mit Schlüssel j in einem random locally balanced binary search tree der Größe n sind gegeben durch*

$$
\begin{aligned}
M_{n,j}^{(2)} = {} & \\
& \frac{576}{245}H_nH_{n-j} + \frac{576}{245}H_nH_j - \frac{288}{245}H_n^2 + \frac{324}{1225}H_{n-j}^2 - \frac{1368}{1225}H_{n-j}H_j + \frac{324}{1225}H_j^2 + \\
& + \left(\frac{72}{42875}\frac{227j+525}{j} + \frac{792}{245}\frac{1}{n+1-j} - \right. \\
& \quad - \frac{2}{1225}\frac{2j^6-6j^5-55j^4+120j^3+953j^2-1014j+540}{n} + \\
& \quad + \frac{4}{245}\frac{(j-1)^2(j^4-4j^3-11j^2+30j+72)}{n-1} - \\
& \quad - \frac{4}{245}\frac{(j-1)(j-2)(2j^4-12j^3+13j^2+15j-36)}{n-2} + \\
& \quad + \frac{8}{245}\frac{(j-1)^2(j-2)^2(j-3)^2}{n-3} - \\
& \quad - \frac{2}{245}\frac{(j-1)(j-2)(j-3)(j-4)(2j^2-10j+15)}{n-4} + \\
& \quad \left. + \frac{4}{1225}\frac{(j-1)(j-2)(j-3)^2(j-4)(j-5)}{n-5}\right) H_n + \\
& + \left(-\frac{36}{42875}\frac{5697j+70}{j} - \frac{288}{245}\frac{1}{n+1-j} + \right. \\
& \quad + \frac{1}{6125}\frac{20j^6-102j^5-445j^4+2040j^3+8165j^2+402j+360}{n} - \\
& \quad - \frac{2}{1225}\frac{(j-1)(10j^5-71j^4+14j^3+557j^2-42j-1224)}{n-1} + \\
& \quad + \frac{2}{1225}\frac{(j-1)(j-2)(20j^4-162j^3+319j^2+129j-612)}{n-2} - \\
& \quad - \frac{4}{1225}\frac{(j-1)(j-2)(j-3)(10j^3-81j^2+194j-102)}{n-3} + \\
& \quad + \frac{1}{1225}\frac{(j-1)(j-2)(j-3)(j-4)(20j^2-142j+255)}{n-4} - \\
& \quad \left. - \frac{2}{6125}\frac{(j-1)(j-2)(j-3)(j-4)(j-5)(10j-51)}{n-5}\right) H_{n-j} + \\
& + \left(-\frac{36}{42875}\frac{2267j+2030}{j} - \frac{288}{245}\frac{1}{n+1-j} + \right. \\
& \quad + \frac{1}{6125}\frac{20j^6-18j^5-655j^4+360j^3+10895j^2-20682j+10440}{n} -
\end{aligned}
$$

$$
\begin{aligned}
&- \frac{2}{1225}\frac{(j-1)(10j^5-29j^4-154j^3+263j^2+882j-216)}{n-1}+\\
&+ \frac{2}{1225}\frac{(j-1)(j-2)(20j^4-78j^3-59j^2+171j-108)}{n-2}-\\
&- \frac{4}{1225}\frac{(j-1)(j-2)(j-3)(10j^3-39j^2+26j-18)}{n-3}+\\
&+ \frac{1}{1225}\frac{(j-1)(j-2)(j-3)(j-4)(20j^2-58j+45)}{n-4}-\\
&- \frac{2}{6125}\frac{(j-1)(j-2)(j-3)(j-4)(j-5)(10j-9)}{n-5}\Bigg) H_j +\\
+ &\frac{288}{245}H_n^{(2)} - \frac{3204}{1225}H_{n-j}^{(2)} - \frac{3204}{1225}H_j^{(2)} +\\
+ &\frac{12}{1225}\frac{1}{j+4} - \frac{24}{175}\frac{1}{j+3} + \frac{792}{1225}\frac{1}{j+2} - \frac{2088}{1225}\frac{1}{j+1} -\\
- &\frac{53148}{42875}\frac{1}{j} - \frac{1044}{875}\frac{1}{j-1} + \frac{396}{875}\frac{1}{j-2} - \frac{12}{125}\frac{1}{j-3} + \frac{6}{875}\frac{1}{j-4} + \frac{144}{175}\frac{1}{j^2} +\\
+ &\frac{6}{875}\frac{1}{n-3-j} - \frac{12}{125}\frac{1}{n-2-j} + \frac{396}{875}\frac{1}{n-1-j} - \frac{1044}{875}\frac{1}{n-j} -\\
- &\frac{24}{8575}\frac{1123j-420}{(n+1-j)j} - \frac{2088}{1225}\frac{1}{n+2-j} + \frac{792}{1225}\frac{1}{n+3-j} - \frac{24}{175}\frac{1}{n+4-j} +\\
+ &\frac{12}{1225}\frac{1}{n+5-j} - \frac{792}{245}\frac{1}{(n+1-j)^2} + \frac{2297696}{214375} -\\
- &\frac{1}{643125}(5522j^7-8166j^6-167395j^5+142845j^4+3249683j^3-1044579j^2-\\
&\qquad -2445930j+1096200)\frac{1}{nj} +\\
+ &\frac{2}{128625}(2761j^7-12366j^6-25787j^5+116883j^4+176746j^3-361269j^2+\\
&\qquad +120672j+22680)\frac{1}{(n-1)j} -\\
- &\frac{2}{128625}(5522j^7-41298j^6+82913j^5+11427j^4-226090j^3+256698j^2-\\
&\qquad -49482j-22680)\frac{1}{(n-2)j} +\\
+ &\frac{4}{128625}(j-1)(j-3)(2761j^5-17888j^4+38198j^3-28798j^2-\\
&\qquad -5718j+3780)\frac{1}{(n-3)j} -\\
- &\frac{1}{128625}(5522j^7-74430j^6+407885j^5-1158375j^4+1755053j^3-1233465j^2+\\
&\qquad +169920j+113400)\frac{1}{(n-4)j} +\\
+ &\frac{2}{643125}(2761j^7-45498j^6+299185j^5-999255j^4+1757614j^3-1456887j^2+\\
&\qquad +313560j+113400)\frac{1}{(n-5)j} +
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{2}{1225}(2j^6-6j^5-55j^4+120j^3+953j^2-1014j+540)\frac{1}{n^2}-\\
&-\frac{4}{245}(j-1)^2(j^4-4j^3-11j^2+30j+72)\frac{1}{(n-1)^2}+\\
&+\frac{4}{245}(j-1)(j-2)(2j^4-12j^3+13j^2+15j-36)\frac{1}{(n-2)^2}-\\
&-\frac{8}{245}(j-1)^2(j-2)^2(j-3)^2)\frac{1}{(n-3)^2}+\\
&+\frac{2}{245}(j-1)(j-2)(j-3)(j-4)(2j^2-10j+15)\frac{1}{(n-4)^2}-\\
&-\frac{4}{1225}\frac{(j-1)(j-2)(j-3)^2(j-4)(j-5)}{(n-5)^2}
\end{aligned}
$$

für n ≥ 8 und $5 \leq$ j $\leq$ n -4 .

$$
\begin{aligned}
M_{n,1}^{(2)} &= \\
&\frac{36}{25}H_n^2-\frac{144}{125}H_n-\frac{36}{25}H_n^{(2)}+\frac{766}{625}+\frac{3}{125}\frac{1}{n}-\frac{12}{125}\frac{1}{n-1}+\frac{18}{125}\frac{1}{n-2}-\frac{12}{125}\frac{1}{n-3}+\\
&+\frac{3}{125}\frac{1}{n-4} \quad \text{für n} \geq 5\,,
\end{aligned}
$$

$M_{1,1}^{(2)}=0\,,\ M_{2,1}^{(2)}=1\,,\ M_{3,1}^{(2)}=2\,,\ M_{4,1}^{(2)}=3\,.$

$$
\begin{aligned}
M_{n,2}^{(2)} &= \\
&\frac{36}{25}H_n^2-\frac{324}{125}H_n-\frac{36}{25}H_n^{(2)}+\frac{1726}{625}-\frac{27}{125}\frac{1}{n}+\frac{108}{125}\frac{1}{n-1}-\frac{162}{125}\frac{1}{n-2}+\\
&+\frac{108}{125}\frac{1}{n-3}-\frac{27}{125}\frac{1}{n-4} \quad \text{für n} \geq 6\,,
\end{aligned}
$$

$M_{2,2}^{(2)}=1\,,\ M_{3,2}^{(2)}=0\,,\ M_{4,2}^{(2)}=2\,,\ M_{5,2}^{(2)}=\frac{11}{5}\,.$

$$
\begin{aligned}
M_{n,3}^{(2)} &= \\
&\frac{36}{25}H_n^2-\left(\frac{1548}{875}+\frac{72}{25}\frac{1}{n(n-1)}\right)H_n-\frac{36}{25}H_n^{(2)}+\frac{77094}{30625}+\frac{2169}{875}\frac{1}{n}-\frac{2892}{875}\frac{1}{n-1}+\\
&+\frac{1476}{875}\frac{1}{n-2}-\frac{1044}{875}\frac{1}{n-3}+\frac{321}{875}\frac{1}{n-4}-\frac{36}{875}\frac{1}{n-5}+\frac{6}{875}\frac{1}{n-6}-\\
&-\frac{72}{25}\frac{1}{n^2}+\frac{72}{25}\frac{1}{(n-1)^2} \quad \text{für n} \geq 7\,,
\end{aligned}
$$

$M_{3,3}^{(2)}=2\,,\ M_{4,3}^{(2)}=2\,,\ M_{5,3}^{(2)}=\frac{18}{5}\,,\ M_{6,3}^{(2)}=\frac{22}{5}\,.$

$$
\begin{aligned}
M_{n,4}^{(2)} &= \\
&\frac{36}{25}H_n^2+\left(-\frac{204}{125}+\frac{144}{25}\frac{1}{n}-\frac{72}{25}\frac{1}{n-1}-\frac{72}{25}\frac{1}{n-2}\right)H_n-\frac{36}{25}H_n^{(2)}+\frac{11077}{4375}+
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{5517}{875}\frac{1}{n}-\frac{786}{125}\frac{1}{n-1}-\frac{246}{125}\frac{1}{n-2}+\frac{378}{125}\frac{1}{n-3}-\frac{177}{125}\frac{1}{n-4}+\frac{54}{125}\frac{1}{n-5}-\\
&-\frac{12}{125}\frac{1}{n-6}+\frac{6}{875}\frac{1}{n-7}-\frac{144}{25}\frac{1}{n^2}+\frac{72}{25}\frac{1}{(n-1)^2}+\frac{72}{25}\frac{1}{(n-2)^2}\quad \text{für n} \geq 8\,,
\end{aligned}
$$

$$
M_{4,4}^{(2)}=3\,,\ M_{5,4}^{(2)}=\frac{11}{5}\,,\ M_{6,4}^{(2)}=\frac{22}{5}\,,\ M_{7,4}^{(2)}=\frac{36}{7}\,. \tag{2.17}
$$

2.4 Die Varianz der Höhe eines Knoten

Bezeichnet man die Varianz der Höhe des Knotens mit Schlüssel j in einem random LBST der Größe n mit $V_{n,j}$, dann erhält man mittels der Beziehung

$$
V_{n,j}=M_{n,j}^{(2)}+E_{n,j}-E_{n,j}^2\,,
$$

durch Kombinieren der zuvor errechneten Resultate unmittelbar

Satz 2.4.1. *Die Varianz $V_{n,j}$ der Höhe des Knotens mit Schlüssel j in einem random locally balanced binary search tree der Größe n ist gegeben durch*

$$
\begin{aligned}
V_{n,j} &= \\
&\frac{288}{175}H_nH_j+\frac{288}{175}H_nH_{n+1-j}-\frac{288}{175}H_n^2-\frac{288}{175}H_jH_{n+1-j}+\\
&+\Bigg(\frac{139488}{42875}+\frac{72}{175}\frac{1}{j}+\frac{72}{175}\frac{1}{n+1-j}-\frac{72}{175}\frac{6j-5}{n}+\frac{288}{175}\frac{(j-1)^2}{n^{\underline{2}}}+\\
&\quad+\frac{96}{175}\frac{(2j-3)(j-1)^{\underline{2}}}{n^{\underline{3}}}+\frac{144}{175}\frac{(j-2)(j-1)^{\underline{3}}}{n^{\underline{4}}}-\frac{144}{175}\frac{(2j-5)(j-1)^{\underline{4}}}{n^{\underline{5}}}+\\
&\quad+\frac{96}{175}\frac{(j-3)(j-1)^{\underline{5}}}{n^{\underline{6}}}\Bigg)H_n+\\
&+\Bigg(-\frac{112734}{42875}-\frac{72}{175}\frac{1}{j}-\frac{72}{35}\frac{1}{n+1-j}+\frac{72}{35}\frac{2j-1}{n}-\frac{144}{175}\frac{(j-1)^{\underline{2}}}{n^{\underline{2}}}-\\
&\quad-\frac{96}{175}\frac{(j-1)^{\underline{3}}}{n^{\underline{3}}}-\frac{72}{175}\frac{(j-1)^{\underline{4}}}{n^{\underline{4}}}+\frac{144}{175}\frac{(j-1)^{\underline{5}}}{n^{\underline{5}}}-\frac{48}{175}\frac{(j-1)^{\underline{6}}}{n^{\underline{6}}}\Bigg)H_{n+1-j}+\\
&+\Bigg(\frac{10746}{42875}-\frac{72}{35}\frac{1}{j}-\frac{72}{175}\frac{1}{n+1-j}-\frac{72}{175}\frac{4j-5}{n}-\frac{144}{175}\frac{j^{\underline{2}}}{n^{\underline{2}}}-\frac{96}{175}\frac{j^{\underline{3}}}{n^{\underline{3}}}-\\
&\quad-\frac{72}{175}\frac{j^{\underline{4}}}{n^{\underline{4}}}+\frac{144}{175}\frac{j^{\underline{5}}}{n^{\underline{5}}}-\frac{48}{175}\frac{j^{\underline{6}}}{n^{\underline{6}}}\Bigg)H_j+\\
&+\frac{288}{245}H_n^{(2)}-\frac{3204}{1225}H_{n+1-j}^{(2)}-\frac{3204}{1225}H_j^{(2)}+\\
&+\frac{1}{926100000}\frac{1}{j^{\underline{5}}}(49336n^2j^{17}-49336nj^{18}-1865416n^2j^{16}+1671908nj^{17}+\\
&\quad+242844j^{18}+32784304n^2j^{15}-25494056nj^{16}-9398508j^{17}-355430740n^2j^{14}+\\
&\quad+227842328nj^{15}+169528380j^{16}+2660193907n^2j^{13}-1283644187nj^{14}-\\
&\quad-1892353176j^{15}-14573920102n^2j^{12}+4328697506nj^{13}+14637396963j^{14}+\\
&\quad+60447522743n^2j^{11}-4676604687nj^{12}-83250254661j^{13}-
\end{aligned}
$$

$$
\begin{aligned}
&- 193475395850n^2j^{10} - 36170564954nj^{11} + 360438321705j^{12} + \\
&+ 482054276453n^2j^9 + 246940887347nj^{10} - 1212564043197j^{11} - \\
&- 935042746298n^2j^8 - 863949222686nj^9 + 3203516805087j^{10} + \\
&+ 1401970954277n^2j^7 + 2046716285867nj^8 - 6664776231879j^9 - \\
&- 1606665630530n^2j^6 - 3484586512526nj^7 + 10876953083775j^8 + \\
&+ 1399475657604n^2j^5 + 4312068202636nj^6 - 13811250629703j^7 - \\
&- 951315080184n^2j^4 - 3893967885528nj^5 + 13559528011686j^6 + \\
&+ 548876398176n^2j^3 + 2697126473376nj^4 - 10418235359052j^5 - \\
&- 276724874880n^2j^2 - 1671052102848nj^3 + 6626911103640j^4 + \\
&+ 88189171200n^2j + 957718391040nj^2 - 3803838011424j^3 + 2612736000n^2 - \\
&- 274471027200nj + 2026211137920j^2 - 108428544000n - 1064459750400j + \\
&+ 555206400000)\frac{1}{(n+1-j)(n-4)(n-5)} -
\end{aligned}
$$

$$
\begin{aligned}
&- \frac{1}{926100000}\frac{1}{j^2}(3836j^{14} - 31892j^{13} - 92960j^{12} + 1268876j^{11} + 1190343j^{10} - \\
&\quad - 26514831j^9 + 11837170j^8 + 245828918j^7 - 345582349j^6 - 868029827j^5 + \\
&\quad + 4697373240j^4 - 4702798044j^3 - 3343999680j^2 + 4329547200j - \\
&\quad - 1360800000)\frac{1}{n} - \\
&- \frac{1}{9261000}\frac{1}{j(j-2)}(455j^{14} - 7630j^{13} + 46900j^{12} - 96320j^{11} - 262290j^{10} + \\
&\quad + 1720740j^9 - 2168948j^8 - 5125976j^7 + 18165755j^6 - 13883206j^5 - \\
&\quad - 41023200j^4 + 119810088j^3 - 98394480j^2 + 16137792j + 10886400)\frac{1}{n-1} - \\
&- \frac{1}{2315250}\frac{1}{j(j-3)}(140j^{14} - 4200j^{13} + 54040j^{12} - 389550j^{11} + 1721685j^{10} - \\
&\quad - 4737915j^9 + 7600984j^8 - 4804158j^7 - 6353236j^6 + 20120466j^5 - \\
&\quad - 28218987j^4 + 24854661j^3 - 8887482j^2 - 6263568j + 4082400)\frac{1}{n-2} + \\
&+ \frac{1}{1157625}\frac{(j-1)(j-3)}{j(j-4)}(70j^{12} - 1050j^{11} + 3500j^{10} + 33810j^9 - 404670j^8 + \\
&\quad + 2022930j^7 - 6035383j^6 + 11464746j^5 - 13898405j^4 + 11040600j^3 - \\
&\quad - 7054452j^2 + 4642704j - 907200)\frac{1}{n-3} + \\
&+ \frac{12}{1225}\frac{1}{n+5-j} - \frac{24}{175}\frac{1}{n+4-j} + \frac{792}{1225}\frac{1}{n+3-j} - \frac{2088}{1225}\frac{1}{n+2-j} - \\
&- \frac{1044}{875}\frac{1}{n-j} + \frac{396}{875}\frac{1}{n-1-j} - \frac{12}{125}\frac{1}{n-2-j} + \frac{6}{875}\frac{1}{n-3-j} + \\
&+ \frac{864}{1225}\frac{1}{(n+1-j)^2} +
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{2}{214375}(705517j^{10}-194925j^9-21064710j^8+5812050j^7+189569541j^6-\\
&\quad-52136385j^5-550488940j^4+146201100j^3+318001392j^2+12130560j+\\
&\quad+43545600)\frac{1}{j\,(j+4)^{\underline{9}}}-\\
&-\frac{1}{4410000}(4j^{12}-24j^{11}-184j^{10}+1140j^9+5397j^8-28692j^7-95542j^6+\\
&\quad+392100j^5+1140049j^4-2969484j^3-7205724j^2+8760960j-3369600)\frac{1}{n^2}-\\
&-\frac{1}{44100}(j-1)^2(j^4-4j^3-11j^2+30j+72)\\
&\quad(j^6-6j^5-2j^4+48j^3+j^2-114j+792)\frac{1}{(n-1)^2}-\\
&-\frac{1}{44100}(j-1)(j-2)(2j^6-18j^5+53j^4-48j^3-55j^2+138j-792)\\
&\quad(2j^4-12j^3+13j^2+15j-36)\frac{1}{(n-2)^2}-\\
&-\frac{1}{11025}(j-1)^2(j-2)^2(j-3)^2\\
&\quad(j^6-12j^5+58j^4-144j^3+193j^2-132j+396)\frac{1}{(n-3)^2}-\\
&-\frac{1}{176400}(j-1)^{\underline{4}}(2j^2-10j+15)\\
&\quad(2j^6-30j^5+185j^4-600j^3+1073j^2-990j-1080)\frac{1}{(n-4)^2}-\\
&-\frac{1}{1102500}(j-3)(j-1)^{\underline{5}}\\
&\quad(j^6-18j^5+130j^4-480j^3+949j^2-942j+3960)\frac{1}{(n-5)^2}
\end{aligned}
$$

für $n \geq 8$ und $5 \leq j \leq n-4$.

$V_{n,1} =$

$$
\begin{aligned}
&\frac{78}{125}H_n-\frac{36}{25}H_n^{(2)}+\frac{116}{125}+\frac{3}{125}\frac{1}{n}-\frac{12}{125}\frac{1}{n-1}+\frac{18}{125}\frac{1}{n-2}-\frac{12}{125}\frac{1}{n-3}+\\
&+\frac{3}{125}\frac{1}{n-4}\qquad \text{für } n \geq 5\ ,
\end{aligned}
$$

$V_{1,1}=0\ ,\ V_{2,1}=\frac{1}{4}\ ,\ V_{3,1}=0\ ,\ V_{4,1}=\frac{3}{16}\ .$

$V_{n,2} =$

$$
\begin{aligned}
&\frac{78}{125}H_n-\frac{36}{25}H_n^{(2)}+\frac{152}{125}-\frac{27}{125}\frac{1}{n}+\frac{108}{125}\frac{1}{n-1}-\frac{162}{125}\frac{1}{n-2}+\frac{108}{125}\frac{1}{n-3}-\\
&-\frac{27}{125}\frac{1}{n-4}\qquad \text{für } n \geq 6\ ,
\end{aligned}
$$

$V_{2,2}=\frac{1}{4}\ ,\ V_{3,2}=0\ ,\ V_{4,2}=\frac{11}{16}\ ,\ V_{5,2}=\frac{49}{100}\ .$

$$
\begin{aligned}
V_{n,3} &= \\
&\frac{78}{125}H_n - \frac{36}{25}H_n^{(2)} + \frac{2172}{1225} + \frac{249}{125}\frac{1}{n} - \frac{2466}{875}\frac{1}{n-1} + \frac{1476}{875}\frac{1}{n-2} - \frac{1044}{875}\frac{1}{n-3} + \\
&+ \frac{321}{875}\frac{1}{n-4} - \frac{36}{875}\frac{1}{n-5} + \frac{6}{875}\frac{1}{n-6} - \frac{108}{25}\frac{1}{n^2} + \frac{36}{25}\frac{1}{(n-1)^2} \quad \text{für n} \geq 7 \ , \\
V_{3,3} &= 0 \ , \ V_{4,3} = \frac{11}{16} \ , \ V_{5,3} = \frac{24}{25} \ , \ V_{6,3} = \frac{26}{25} \ . \\
V_{n,4} &= \\
&\frac{78}{125}H_n - \frac{36}{25}H_n^{(2)} + \frac{1661}{875} + \frac{381}{175}\frac{1}{n} - \frac{12}{125}\frac{1}{n-1} - \frac{528}{125}\frac{1}{n-2} + \frac{378}{125}\frac{1}{n-3} - \\
&- \frac{177}{125}\frac{1}{n-4} + \frac{54}{125}\frac{1}{n-5} - \frac{12}{125}\frac{1}{n-6} + \frac{6}{875}\frac{1}{n-7} - \frac{288}{25}\frac{1}{n^2} + \frac{36}{25}\frac{1}{(n-1)^2} + \\
&+ \frac{36}{25}\frac{1}{(n-2)^2} \quad \text{für n} \geq 8 \ , \\
V_{4,4} &= \frac{3}{16} \ , \ V_{5,4} = \frac{49}{100} \ , \ V_{6,4} = \frac{26}{25} \ , \ V_{7,4} = \frac{54}{49} \ .
\end{aligned} \tag{2.18}
$$

Wiederum lassen sich leicht einige Folgerungen aus diesem Resultat gewinnen. Zunächst wird das asymptotische Verhalten von $V_{n,j}$ für festes $\rho = j/n$ beschrieben durch

Korollar 2.4.2. *Als asymptotische Entwicklung für die Varianz der Höhe des Knotens j in einem random locally balanced binary search tree der Größe n für $j = \rho n$ und $0 < \rho < 1$ erhält man*

$$\boxed{V_{n,j} = \frac{300}{343}\log n + \mathcal{O}\left(1\right) \ .} \tag{2.19}$$

Für das asymptotische Verhalten von $V_{n,j}$ für fixes j, also am 'linken Rand' des Baumes erhält man

Korollar 2.4.3. *Als asymptotische Entwicklung für die Varianz der Höhe des Knotens j in einem random locally balanced binary search tree der Größe n für fixes j ergibt sich*

$$\boxed{V_{n,j} = \frac{78}{125}\log n + \mathcal{O}\left(1\right) \ .} \tag{2.20}$$

Wiederum gilt aus Symmetriegründen für fixes $l = n + 1 - j$, das heißt am 'rechten Rand' des Baumes

Korollar 2.4.4. *Als asymptotische Entwicklung für die Varianz der Höhe des Knotens $l = n + 1 - j$ in einem random locally balanced binary search tree der Größe n für fixes l ergibt sich*

$$\boxed{V_{n,l} = \frac{78}{125}\log n + \mathcal{O}\left(1\right) \ .} \tag{2.21}$$

Bildet man das arithmetische Mittel über die $V_{n,j}$ bei festem n, sei also $\bar{V}_n$ definiert durch

$$\bar{V}_n = \frac{1}{n}\sum_{j=1}^{n} V_{n,j}\,,$$

so gilt

Korollar 2.4.5. *Der Mittelwert $\bar{V}_n$ der Varianzen der Höhen $V_{n,j}$ der Knoten in einem LBST der Größe n ist gegeben durch*

$$\boxed{\begin{aligned}\bar{V}_n = {} & \frac{12}{42875}\big(3125n^8 - 28914n^7 + 11046n^6 + 615720n^5 - 2144415n^4 + 2457714n^3 \\ & \quad - 1044956n^2 + 836280n - 504000\big)\frac{1}{n(n+1)^{\underline{7}}}H_n \\ & - \frac{72}{35}\frac{H_n^2}{n} - \frac{72}{1225}\frac{41n+30}{n}H_n^{(2)} \\ & + \frac{1}{1287536250}\big(4362842618n^7 - 69153941563n^6 + 424049263925n^5 \\ & \quad - 1260094812235n^4 + 1824373821137n^3 - 1066873113922n^2 \\ & \quad + 46549806120n + 74863303200\big)\frac{1}{nn^{\underline{6}}} \quad \text{für n} \geq 8\,, \\ \bar{V}_1 = 0\,,\ & \bar{V}_2 = \frac{1}{4}\,,\ \bar{V}_3 = 0\,,\ \bar{V}_4 = \frac{7}{16}\,,\ \bar{V}_5 = \frac{61}{125}\,,\ \bar{V}_6 = \frac{97}{150}\,,\ \bar{V}_7 = \frac{13247}{17150}\,.\end{aligned}} \tag{2.22}$$

Zu unterscheiden von den eben errechneten $\bar{V}_n$ ist die Varianz V_n der Höhe eines zufällig gewählten Knotens in einem random LBST. Aus den hier errechneten Resultaten kann man diese V_n mittels

$$V_n = \frac{1}{n}\sum_j M_{n,j}^{(2)} + \frac{1}{n}\sum_j E_{n,j} - \left(\frac{1}{n}\sum_j E_{n,j}\right)^2$$

ebenfalls erhalten. Zur Gewinnung des arithmetischen Mittels der $M_{n,j}^{(2)}$ gibt es allerdings einen einfacheren Weg als diese mühsame Summation. Man kann nämlich in der Differentialgleichung (2.5) $u = 1$ setzen, zweimal nach v ableiten und $v = 1$ setzen. Löst man nun diese Differentialgleichung, dann erhält man durch Ablesen der Koeffizienten ebenfalls das Gewünschte. Details dazu kann man [30] entnehmen. Jedenfalls gilt

Korollar 2.4.6. *Die Varianz V_n der Höhe eines zufälligen Knotens in einem random locally balanced binary search tree der Größe n oder äquivalent dazu die Varianz der Suchkosten für eine erfolgreiche Suche eines zufälligen Elementes in einem random LBST der Größe n ist gegeben durch*

$$\boxed{\begin{aligned}V_n = {} & \frac{1}{343}\left(300 + \frac{2100}{n} - \frac{216}{n^2}\right)H_n - \frac{144}{49}\left(1+\frac{1}{n}\right)\left(\frac{H_n^2}{n} + H_n^{(2)}\right) \\ & + \frac{1}{2401}\left(10758 + \frac{2431}{n} - \frac{81}{n^2}\right) + \frac{2304}{343nn^{\underline{6}}} \quad \text{für n} \geq 6\,, \\ V_1 = 0\,,\ & V_2 = \frac{1}{4}\,,\ V_3 = \frac{2}{9}\,,\ V_4 = \frac{1}{2}\,,\ V_5 = \frac{14}{25}\,.\end{aligned}} \tag{2.23}$$

Kapitel 3

Über die Höhe und die Anzahl der Nachfolger von Knoten in traversierten Binärbäumen

3.1 Einleitung

Bäume und insbesondere Binärbäume sind eine überaus vielseitig verwendbare und auch verwendete Datenstruktur. Als solche muß die in diesem Baum abgelegte Information natürlich bei Bedarf wieder abgerufen werden können. Das heißt, man braucht ein Verfahren mit dem man systematisch jeden Knoten besuchen kann, mit anderen Worten wie man ihn traversieren kann. In der Praxis verwendet man meist eine der folgenden rekursiven Traversierungen: Preorder-, Postorder- bzw. Inorder-Traversierung. Diese unterscheiden sich nur darin, ob die Wurzel *vor* (bei Preorder) *nach* (bei Postorder) oder *zwischen* (bei Inorder) dem linken und rechten Unterbaum besucht wird.

Im folgenden bezeichnen wir mit $\mathcal{B}$ die Familie der Binärbäume, manchmal auch erweiterte Binärbäume genannt. Sei $\bigcirc$ das Symbol für einen inneren Knoten und $\Box$ das Symbol für ein Blatt (oder auch externer Knoten genannt), dann ist die Familie $\mathcal{B}$ durch folgende formale Gleichung gegeben:

$$\mathcal{B} \quad = \quad \Box \quad + \quad \begin{array}{c} \bigcirc \\ \diagup \quad \diagdown \\ \mathcal{B} \qquad \mathcal{B} \end{array} \quad . \tag{3.1}$$

Die Gesamtheit der Objekte aus $\mathcal{B}$ mit n inneren Knoten sei fortan mit $\mathcal{B}_n$ bezeichnet. Wir schreiben weiters für die Mächtigkeit der $\mathcal{B}_n$, also die Anzahl der Binärbäume mit n inneren Knoten, abkürzend B_n. Schließlich soll $B(z) = \sum_{n \geq 0} B_n z^n$ als die gewöhnliche erzeugende Funktion der B_n definiert sein.

Unter der Größe $|B|$ eines Binärbaumes B verstehen wir in Hinkunft weiters die Anzahl der inneren Knoten von B.

Die Preorder-Traversierung eines Binärbaumes definiert man dann rekursiv wie folgt: Falls der Baum innere Knoten hat, besuche die Wurzel, anschließend traversiere den linken Unterbaum und nachfolgend traversiere den rechten Unterbaum.

Die Postorder-Traversierung eines Binärbaumes kann man rekursiv wie folgt definieren: Falls der Baum innere Knoten hat traversiere zuerst den linken Unterbaum, dann traversiere den rechten Unterbaum und schließlich besuche die Wurzel.

Die Inorder-Traversierung eines Binärbaumes kann man rekursiv wie folgt beschreiben: Falls der Baum innere Knoten hat traversiere zuerst den linken Unterbaum, dann besuche die Wurzel und schließlich traversiere den rechten Unterbaum.

Die Preorder- und die Postorder-Traversierung ist natürlich nicht nur auf Binärbäume beschränkt, sondern kann auch für viel allgemeinere Familien von Bäumen, beispielsweise für die in [31] eingeführten einfach erzeugten Familien von Bäumen, sinngemäß übernommen werden. Damit ist gemeint, man traversiert bei einem Baum mit Knotengrad der Wurzel $k \geq 1$ nacheinander von links nach rechts alle k Unterbäume.

Bei einer solchen Traversierung ist es oft von Bedeutung zu wissen, wie hoch der Knoten liegt, den man als j-tes besucht hat, bzw. wie viele Nachfolger dieser Knoten besitzt. Zur Untersuchung solcher Fragestellungen belegt man den Baum mit Marken von 1 bis n, wenn n die Knotenzahl des Baumes darstellt, und zwar so, daß diese Belegung genau dem Ablauf der Traversierung entspricht. Siehe hierzu die Abbildungen 3.1, 3.2 und 3.3 für Inorder-, Preorder- bzw. Postorder-Traversierung.

Unter der Höhe eines Knotens versteht man die Anzahl der Knoten, die auf dem direkten Weg von der Wurzel zu diesem Knoten liegen.

Unter der Anzahl der Nachkommen eines Knotens p in einem Baum B versteht man die Größe des Unterbaums von B mit p als Wurzel. Das heißt also der Knoten p wird quasi als sein eigener Nachkomme gezählt.

Mit einem ähnlichen Problem hat sich schon vor 15 Jahren Kirschenhofer [19, 20] beschäftigt. Er untersuchte damals die Höhe der Blätter in einem Binärbaum. Dabei denkt man sich die $n+1$ Blätter eines Binärbaumes mit n inneren Knoten von links nach rechts mit den Zahlen von 0 bis n durchnumeriert (siehe Abbildung 3.1). Etwas genauer formuliert, würde ein Algorithmus, der die Blätter in der angegebenen Reihenfolge traversiert, so lauten: Wenn der Binärbaum nur aus einem externen Knoten besteht, dann besuche ihn, ansonsten traversiere die Blätter des linken Unterbaums und anschließend traversiere die Blätter des rechten Unterbaums.

Die große Ähnlichkeit dieses Algorithmus mit den oben genannten erlaubt es, alle Probleme mit den gleichen Methoden zu behandeln. Es ergibt sich so beispielsweise nicht nur ein exaktes Resultat für den Erwartungswert der Höhe des j-ten Blattes in einem Binärbaum der Größe n, sondern man erhält auch ein exaktes Resultat für das entsprechende zweite Faktorielle Moment und die Varianz.

Bei allen Untersuchungen dieses Abschnitts beginnen wir mit algebraischen Gleichungen von gewöhnlichen erzeugenden Funktionen

$$F(z,u,v) = \sum_{j\geq 1\,(0)} \sum_{n\geq j} \sum_{m\geq 0} f_{n,j,m}\, z^n u^j v^m$$

in drei Variablen. Die Koeffizienten $f_{n,j,m}$ sind definiert als die Anzahl der Objekte B aus $\mathcal{B}_n$, für die der Knoten mit Marke j eine Bedingung der Art $H(B,j) = m$ erfüllt. Dabei ist $H(B,j)$ allgemein eine nichtnegative ganzzahlige Funktion auf $\mathcal{B} \times \mathbb{N}$. Im speziellen ist bei

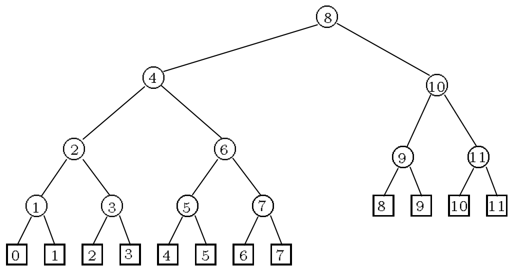


Abbildung 3.1: Ein Binärbaum der Größe 11, bei dem die Knoten entsprechend der Reihenfolge der Inorder-Traversierung und die Blätter von links nach rechts markiert wurden. Knoten 4 hat 7 Nachkommen und die Höhe 2, Blatt 6 besitzt die Höhe 5.

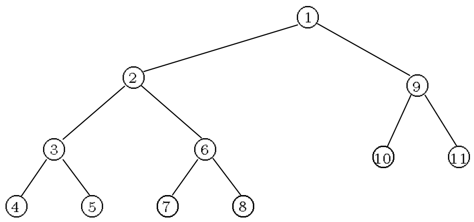


Abbildung 3.2: Ein Binärbaum der Größe 11, bei dem die Knoten entsprechend der Reihenfolge der Preorder-Traversierung markiert wurden. Knoten 3 hat 3 Nachkommen und besitzt die Höhe 3. (Die Blätter sind hier nicht extra angegeben).

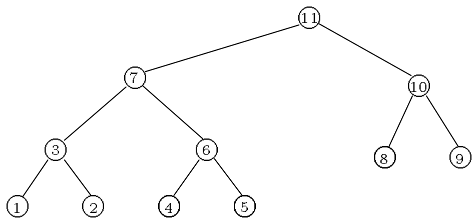


Abbildung 3.3: Ein Binärbaum der Größe 11, bei dem die Knoten entsprechend der Reihenfolge der Postorder-Traversierung markiert wurden. Knoten 5 hat 1 Nachkommen und besitzt die Höhe 4. (Die Blätter sind hier nicht extra angegeben).

uns mit der Funktion $H(B,j)$ immer die Höhe des mit j markierten Knotens oder die Anzahl der Nachfolger des j-ten Knotens im Baum B gemeint.

Diese algebraischen Gleichungen für die Funktionen $F(z,u,v)$ werden nicht aus Rekursionen der $f_{n,j,m}$ ermittelt, sondern mittels der sogenannten symbolischen Operatormethode (siehe hierzu z. B. [44]) direkt aus der formalen Gleichung (3.1) hergeleitet.

Um zu den für uns besonders interessanten Erwartungswerten $E_{n,j}$, zweiten Faktoriellen Momenten $M_{n,j}^{(2)}$ bzw. Varianzen $V_{n,j}$ der $H(\mathcal{B}_n,j) := H(B,j)\big|_{B\in\mathcal{B}_n}$, also der Funktion H

eingeschränkt auf die Menge $\mathcal{B}_n$ zu gelangen, verwendet man die Beziehungen

$$\begin{aligned}
E_{n,j} &= \frac{\sum_{m\geq 0} m\cdot |\{B\in\mathcal{B}_n : H(B,j)=m\}|}{B_n} = \frac{[z^n u^j]\ \frac{\partial}{\partial v}F(z,u,v)\big|_{v=1}}{B_n}\,,\\
M_{n,j}^{(2)} &= \frac{\sum_{m\geq 0} m(m-1)\cdot |\{B\in\mathcal{B}_n : H(B,j)=m\}|}{B_n} = \frac{[z^n u^j]\ \frac{\partial^2}{\partial v^2}F(z,u,v)\Big|_{v=1}}{B_n}\,,\\
V_{n,j} &= M_{n,j}^{(2)} + E_{n,j} - (E_{n,j})^2\ .
\end{aligned} \tag{3.2}$$

Dabei soll $|\mathcal{M}|$ die Mächtigkeit einer Menge $\mathcal{M}$ bedeuten.

Mithilfe der oben angesprochenen symbolischen Operatormethode übersetzt sich die formale Gleichung (3.1) mittels

$$B(z) = 1 + zB^2(z) \tag{3.3}$$

direkt in eine Gleichung für die erzeugende Funktion $B(z)$ der B_n. Aus der daraus resultierenden Lösung

$$B(z) = \frac{1-\sqrt{1-4z}}{2z} \tag{3.4}$$

erhält man nach dem Koeffizientenablesen ein klassisches Resultat der Kombinatorik. Die Anzahl der Binärbäume B_n der Größe n sind ja bekanntlich gegeben durch

$$B_n = \frac{1}{n+1}\binom{2n}{n}\,, \tag{3.5}$$

also durch die sogenannten Catalan-Zahlen.

In diesem Kapitel wollen wir weiters die folgenden Abkürzungen verwenden, um die Berechnungen optisch besser aufzubereiten.

$$\begin{aligned}
X &= \sqrt{1-4z}\,, & Y &= \sqrt{1-4uz}\,,\\
G_1(z,u) &= \frac{\partial}{\partial v}F(z,u,v)\bigg|_{v=1}\,, & G_2(z,u) &= \frac{\partial^2}{\partial v^2}F(z,u,v)\bigg|_{v=1}\,.
\end{aligned} \tag{3.6}$$

3.2 Die Höhe der Blätter

Sei hier nun $f_{n,j,m}$ die Anzahl der Binärbäume mit n inneren Knoten, bei denen das Blatt mit Marke j genau die Höhe m hat. Wir wollen nun aus der formalen Gleichung für die Binärbäume (3.1) direkt eine Gleichung für die erzeugende Funktion der $f_{n,j,m}$ in 3 Variablen

$$F(z,u,v) = \sum_{j\geq 0}\sum_{n\geq j}\sum_{m\geq 0} f_{n,j,m}\, z^n u^j v^m$$

gewinnen.

Dazu wollen wir zunächst zeigen, wie man die Funktion $F(z,u,v)$ auch anders deuten kann. Wir definieren uns dazu eine Funktion $\widetilde{F}$, die jedem Objekt B aus $\mathcal{B}$ ein Polynom in z, u und v durch folgende Vorschrift

$$\widetilde{F}(B) := \sum_{j=0}^{|B|} z^{|B|} u^j v^{H(B,j)}$$

zuweist. Dabei soll $H(B,j)$ die Höhe des j-ten Knotens im Baum B bezeichnen.

Wir können diese Definition mittels

$$\widetilde{F}\left((B_i)_{i\in I}\right) := \sum_{i\in I} \widetilde{F}(B_i)$$

auf eine Familie von Bäumen $(B_i)_{i\in I}$, $B_i \in \mathcal{B}$ ausdehnen, sofern gewährleistet ist, daß für jedes n nur endlich viele $B_i \in \mathcal{B}_n$ sind. Dann weist $\widetilde{F}$ jeder Familie $(B_i)_{i\in I}$ eine formale Potenzreihe in z, u und v zu.

Da ja $\mathcal{B} = \mathcal{B}_0 + \mathcal{B}_1 + \mathcal{B}_2 + \ldots$ gilt, folgt

$$\begin{aligned} \widetilde{F}(\mathcal{B}) &= \sum_{n\ge 0} \sum_{B\in\mathcal{B}_n} \sum_{j=0}^{n} z^n u^j v^{H(B,j)} \\ &= \sum_{j\ge 0} u^j \sum_{n\ge j} z^n \sum_{m\ge 0} \left|\{B \in \mathcal{B}_n : H(B,j) = m\}\right| \cdot v^m \\ &= F(z,u,v)\ , \end{aligned} \tag{3.7}$$

wenn $F(z,u,v)$ als formale Potenzreihe aufgefaßt wird.

Wenden wir nun $\widetilde{F}$ auf die Gleichung (3.1) an, so ergibt sich zunächst aufgrund der Additivität

$$\widetilde{F}(\mathcal{B}) \quad = \quad \widetilde{F}(\Box) \quad + \quad \widetilde{F}\left(\begin{smallmatrix} & \circ & \\ \diagup & & \diagdown \\ \mathcal{B} & & \mathcal{B} \end{smallmatrix}\right) . \tag{3.8}$$

Man erhält unmittelbar

$$\widetilde{F}(\mathcal{B}) = F(z,u,v) \quad \text{und} \quad \widetilde{F}(\Box) = v\ . \tag{3.9}$$

Weiters behandeln wir zunächst für beliebige $B' \in \mathcal{B}$ bzw. $B'' \in \mathcal{B}$ den Ausdruck
$\widetilde{F}\left(\begin{smallmatrix} & \circ & \\ \diagup & & \diagdown \\ B' & & B'' \end{smallmatrix}\right)$.

Indem man die Blätter im linken Unterbaum B' und im rechten Unterbaum B'' getrennt betrachtet, folgert man leicht

$$\widetilde{F}\left(\begin{smallmatrix} & \circ & \\ \diagup & & \diagdown \\ B' & & B'' \end{smallmatrix}\right) = \widetilde{F}(B') z v z^{|B''|} + z v (uz)^{|B'|} u \widetilde{F}(B'')\ .$$

Für die Blätter des linken Unterbaums kommen nämlich durch die Wurzel ein Faktor zv und durch den rechten Ast ein Faktor $z^{|B''|}$ zu $\widetilde{F}(B')$ dazu. Für die Blätter des rechten Unterbaums kommen durch die Wurzel wiederum ein Faktor zv und durch den linken Ast ein Faktor $(uz)^{|B'|}u$ zu $\widetilde{F}(B'')$ dazu, da ein Binärbaum mit n inneren Knoten immer $n+1$ Blätter besitzt.

Dann bekommt man weiter

$$
\begin{aligned}
\widetilde{F}\left(\genfrac{}{}{0pt}{}{\circ}{\mathcal{B}\quad\mathcal{B}}\right) &= \sum_{B'\in\mathcal{B},B''\in\mathcal{B}} \widetilde{F}\left(\genfrac{}{}{0pt}{}{\circ}{B'\quad B''}\right) \\
&= F(z,u,v)zv\sum_{B''\in\mathcal{B}} z^{|B''|} + \sum_{B'\in\mathcal{B}} (uz)^{|B'|} zuvF(z,u,v) \qquad (3.10)\\
&= zvF(z,u,v)B(z) + zuvB(zu)F(z,u,v)\ ,
\end{aligned}
$$

wenn $B(z)$ die durch (3.4) gegebene erzeugende Funktion der Catalan-Zahlen bezeichnet.

Zusammenfassend erhält man als Funktionalgleichung für $F(z,u,v)$:

$$F(z,u,v) = v + zvF(z,u,v)B(z) + zuvF(z,u,v)B(zu)\ , \tag{3.11}$$

woraus sich sofort als Lösung für die erzeugende Funktion $F(z,u,v)$ in 3 Variablen

$$F(z,u,v) = \frac{v}{1 - zvB(z) - zuvB(zu)} \tag{3.12}$$

ergibt.

Durch Ableiten von (3.12) nach v und Auswerten bei $v=1$ erhält man unter Verwendung der oben eingeführten Abkürzungen (3.6)

$$G_1(z,u) = \frac{4}{(\sqrt{1-4z}+\sqrt{1-4uz}\,)^2} = \frac{4}{(X+Y)^2}\ . \tag{3.13}$$

Um zu dem Erwartungswert $E_{n,j}$ der Höhe des j-ten Blattes im Binärbaum der Größe n zu gelangen bleibt aufgrund der Beziehung (3.2) nur mehr das Ablesen der Koeffizienten $[z^n u^j]$ im Ausdruck $\frac{1}{(X+Y)^2}$.

Alternativ zu der in [19] verwendeten Methode, bietet sich ein Weg an, der die neuen Möglichkeiten nutzt, die sich durch Zeilberger's Algorithmus ergeben (siehe hierzu beispielsweise [34]).

Dazu entwickeln wir zunächst $\frac{1}{(X+Y)^2}$ um $u=1$ in eine Potenzreihe und bekommen so

$$
\begin{aligned}
\frac{1}{(X+Y)^2} &= \frac{(X-Y)^2}{(X^2-Y^2)^2} = \frac{2-4z-4uz-2\sqrt{1-4z}\sqrt{1-4uz}}{16z^2(1-u)^2} \\
&= \sum_{k\geq 0} \frac{1}{4(k+2)}\binom{2(k+1)}{k+1}\frac{z^k}{(1-4z)^{k+1}}(u-1)^k\ .
\end{aligned} \tag{3.14}
$$

Als Koeffizient von $[z^n u^j]$ ergibt sich dann die Summe

$$
\begin{aligned}
[z^n u^j]\frac{1}{(X+Y)^2} &= \sum_{k=j}^{n} [z^n u^j] \frac{1}{4(k+2)}\binom{2(k+1)}{k+1}\frac{z^k}{(1-4z)^{k+1}}(u-1)^k \\
&= \sum_{k=j}^{n} \frac{1}{4(k+2)}\binom{2(k+1)}{k+1}\binom{n}{k}\binom{k}{j}(-1)^{k-j}4^{n-k}\ ,
\end{aligned} \tag{3.15}
$$

bei der man die Summationsvariable auch über ganz $\mathbb{Z}$ laufen lassen kann.
Schreibt man abkürzend

$$F(n,j,k) = \frac{1}{4(k+2)}\binom{2(k+1)}{k+1}\binom{n}{k}\binom{k}{j}(-1)^{k-j}4^{n-k},$$

dann sind die gesuchten Koeffizienten mittels

$$[z^n u^j]\frac{1}{(X+Y)^2} = f(n,j) = \sum_{k\in\mathbb{Z}} F(n,j,k)\,. \tag{3.16}$$

als Summe passender hypergeometrischer Terme dargestellt (siehe [34]).
Zeilberger's Algorithmus liefert dann

$$\begin{aligned} &-2(n+1)(2n-2j+3)F(n,j,k) + (n+3)(n-j+1)F(n+1,j,k) \\ &\qquad = G(n,j,k+1) - G(n,j,k) \end{aligned} \tag{3.17}$$

mit

$$G(n,j,k) = R(n,j,k)F(n,j,k) \quad \text{und } R(n,j,k) = -4\frac{(n+1)(k+2)(k-j)}{n+1-k}\,.$$

Beim Aufsummieren über alle $k \in \mathbb{Z}$ teleskopiert die rechte Seite obiger Gleichung und liefert 0, sodaß sich folgende lineare Rekursion ergibt:

$$\begin{aligned} (n+3)(n-j+1)f(n+1,j) &= 2(n+1)(2n-2j+3)f(n,j) \quad \text{für } n \geq j\,, \\ f(j,j) &= \frac{1}{4(j+2)}\binom{2(j+1)}{j+1}\,. \end{aligned} \tag{3.18}$$

Man bekommt so

$$f(n,j) = f(j,j)\cdot\prod_{k=j+1}^{n}\frac{2k(2k-2j+1)}{(k+2)(k-j)} = \frac{1}{4(j+2)}\binom{2(j+1)}{j+1}\prod_{k=j+1}^{n}\frac{2k(2k-2j+1)}{(k+2)(k-j)}\,, \tag{3.19}$$

woraus sich nach Vereinfachungen schließlich für $0 \leq j \leq n$

$$[z^n u^j]\frac{1}{(X+Y)^2} = f(n,j) = \frac{(2j+1)(2n-2j+1)}{2(n+1)(n+2)}\binom{2j}{j}\binom{2(n-j)}{n-j} \tag{3.20}$$

ergibt.
Dies zusammen mit (3.2) und (3.5) liefert

Satz 3.2.1. *Der Erwartungswert $E_{n,j}$ der Höhe des Blattes mit Marke j in einem Binärbaum der Größe n ist gegeben durch*

$$\boxed{E_{n,j} = \frac{2(2j+1)(2n-2j+1)}{n+2}\,\frac{\binom{2j}{j}\binom{2n-2j}{n-j}}{\binom{2n}{n}} \quad \text{für } 0 \leq j \leq n\,.} \tag{3.21}$$

Da wir die Höhe eines Blattes als die Anzahl der Knoten auf dem Weg zur Wurzel definiert haben, unterscheidet sich dieses Resultat um genau den Wert 1 von dem in [19] gefundenen. Dort betrachtet man nämlich die Anzahl der *Kanten*, die auf dem Weg von der Wurzel zum gesuchten Blatt liegen.

Für alle exakten Ergebnisse dieses Abschnitts ist es aufgrund der Komplexität der Formeln von Interesse auch asymptotische Aussagen über die untersuchten Größen zu erhalten. Von besonderer Bedeutung sind hierbei die Fälle j fix (also am 'linken Rand' des Baumes), $n-j$ fix (also am 'rechten Rand' des Baumes) und $j \sim \rho n$ für $0 < \rho < 1$ (also ein fixes Verhältnis von j/n). Zu allen Sätzen dieses Kapitels werden diese asymptotischen Entwicklungen als Korollare angegeben.

Um aus den exakten Formeln asymptotische Aussagen zu erhalten, benötigt man meist nicht mehr als Stirling's Formel für die asymptotische Entwicklung der Faktoriellen, wie man sie beispielsweise in [1] findet.

Es gilt

$$\begin{aligned} n! &= \frac{n^n\sqrt{2\pi}\sqrt{n}}{e^n}\left(1+\frac{1}{12n}+\frac{1}{288n^2}+\mathcal{O}\left(\frac{1}{n^3}\right)\right) \\ \text{und} \qquad \binom{2n}{n} &= \frac{4^n}{\sqrt{\pi}\sqrt{n}}\left(1-\frac{1}{8n}+\frac{1}{128n^2}+\mathcal{O}\left(\frac{1}{n^3}\right)\right). \end{aligned} \tag{3.22}$$

Dies führt schon zu

Korollar 3.2.2. *Der Erwartungswert $E_{n,j}$ der Höhe des Blattes mit Marke j in einem Binärbaum der Größe n ist asymptotisch für $j \sim \rho n$ und $0 < \rho < 1$ gegeben durch*

$$\boxed{E_{n,j} \sim 8\frac{\sqrt{\rho(1-\rho)}}{\sqrt{\pi}}\sqrt{n}\,.} \tag{3.23}$$

Korollar 3.2.3. *Der Erwartungswert $E_{n,j}$ der Höhe des Blattes mit Marke j in einem Binärbaum der Größe n ist asymptotisch für fixes j gegeben durch*

$$\boxed{E_{n,j} \sim (2j+1)\binom{2j}{j}4^{-j+1}\,.} \tag{3.24}$$

Korollar 3.2.4. *Der Erwartungswert $E_{n,j}$ der Höhe des Blattes mit Marke j in einem Binärbaum der Größe n ist asymptotisch für fixes $l = n-j$ gegeben durch*

$$\boxed{E_{n,j} \sim (2l+1)\binom{2l}{l}4^{-l+1}\,.} \tag{3.25}$$

Zweimaliges Ableiten von (3.12) nach v und Auswerten bei $v = 1$ führt unter Verwendung der abkürzenden Schreibweisen zu

$$G_2(z,u) = \frac{16}{(X+Y)^3} - \frac{8}{(X+Y)^2}\,. \tag{3.26}$$

Man benötigt für die zweiten Faktoriellen Momente der Höhen der Blätter also die Koeffizienten von $\frac{1}{(X+Y)^3}$. Dazu formen wir diesen Ausdruck um zu

$$\begin{aligned}\frac{1}{(X+Y)^3} &= \frac{(X-Y)^3}{(X^2-Y^2)^3} = \frac{X^3-3X^2Y+3XY^2-Y^3}{-64z^3(1-u)^3}\\ &= \frac{\sqrt{1-4z}}{(1-u)^3}\left(-\frac{1}{16z^3}+\frac{1}{16z^2}+\frac{3u}{16z^2}\right)+\frac{\sqrt{1-4uz}}{(1-u)^3}\left(\frac{1}{16z^3}-\frac{3}{16z^2}-\frac{u}{16z^2}\right).\end{aligned} \tag{3.27}$$

Für $n \geq j$ ergibt der zweite Summand beim Ablesen der Koeffizienten keinen Beitrag und es gilt

$$[z^n u^j]\frac{1}{(X+Y)^3} = [z^n u^j]\frac{\sqrt{1-4z}}{(1-u)^3}\left(-\frac{1}{16z^3}+\frac{1}{16z^2}+\frac{3u}{16z^2}\right). \tag{3.28}$$

Mittels

$$[z^n u^j]\frac{\sqrt{1-4z}}{(1-u)^3} = -\frac{2}{n}\binom{j+2}{2}\binom{2(n-1)}{n-1} \tag{3.29}$$

folgert man

$$[z^n u^j]\frac{1}{(X+Y)^3} = \frac{3}{4}\frac{(2n+1)(j+1)(n-j+1)}{(n+1)(n+2)(n+3)}\binom{2n}{n} \quad \text{für } 0 \leq j \leq n\,. \tag{3.30}$$

Mit (3.20) und (3.30) bekommt man unter Verwendung von (3.2) sofort

Hilfssatz 3.2.5. *Das zweite Faktorielle Moment $M_{n,j}^{(2)}$ der Höhe des Blattes mit Marke j in einem Binärbaum der Größe n ist für $0 \leq j \leq n$ gegeben durch*

$$\boxed{M_{n,j}^{(2)} = \frac{12(j+1)(2n+1)(n-j+1)}{(n+2)(n+3)} - \frac{4(2j+1)(2n-2j+1)}{n+2}\frac{\binom{2j}{j}\binom{2n-2j}{n-j}}{\binom{2n}{n}}.} \tag{3.31}$$

Unter Berücksichtigung von (3.2) folgt sofort weiter

Satz 3.2.6. *Die Varianz $V_{n,j}$ der Höhe des Blattes mit Marke j in einem Binärbaum der Größe n ist gegeben durch*

$$\boxed{\begin{aligned}V_{n,j} = {}&\frac{12(j+1)(2n+1)(n-j+1)}{(n+2)(n+3)} - \frac{2(2j+1)(2n-2j+1)}{n+2}\frac{\binom{2j}{j}\binom{2n-2j}{n-j}}{\binom{2n}{n}}\\ &-\frac{4(2j+1)^2(2n-2j+1)^2}{(n+2)^2}\frac{\binom{2j}{j}^2\binom{2n-2j}{n-j}^2}{\binom{2n}{n}^2} \quad \text{für } 0 \leq j \leq n\,.\end{aligned}} \tag{3.32}$$

Die asymptotischen Aussagen über diese Größe sind nachfolgend angeführt.

Korollar 3.2.7. *Die Varianz $V_{n,j}$ der Höhe des Blattes mit Marke j in einem Binärbaum der Größe n ist asymptotisch für $j \sim \rho n$ und $0 < \rho < 1$ gegeben durch*

$$\boxed{V_{n,j} \sim 8\frac{\rho(1-\rho)(3\pi-8)}{\pi} n\,.} \tag{3.33}$$

Korollar 3.2.8. *Die Varianz $V_{n,j}$ der Höhe des Blattes mit Marke j in einem Binärbaum der Größe n ist asymptotisch für fixes j gegeben durch*

$$\boxed{V_{n,j} \sim 24(j+1) - 16(2j+1)^2\binom{2j}{j}^2 4^{-2j} - (2j+1)\binom{2j}{j}4^{-j+1}\,.} \tag{3.34}$$

Korollar 3.2.9. *Die Varianz $V_{n,j}$ der Höhe des Blattes mit Marke j in einem Binärbaum der Größe n ist asymptotisch für fixes $l = n - j$ gegeben durch*

$$\boxed{V_{n,j} \sim 24(l+1) - 16(2l+1)^2\binom{2l}{l}^2 4^{-2l} - (2l+1)\binom{2l}{l}4^{-l+1}\,.} \tag{3.35}$$

3.3 Die Anzahl der Nachkommen bei Inorder-Traversierung

Wir beschäftigen uns hier mit der Anzahl der Nachkommen von entsprechend der Reihenfolge der Inorder-Traversierung von 1 bis n markierten Knoten in Binärbäumen. Bezeichnen wir also nun mit $f_{n,j,m}$ die Anzahl der Binärbäume mit n inneren Knoten, bei denen der entsprechend der Inorder-Traversierung markierte Knoten j genau m Nachkommen hat. $F(z,u,v)$ bezeichnet dann die erzeugende Funktion der $f_{n,j,m}$ in 3 Variablen, ist also definiert durch

$$F(z,u,v) = \sum_{j\geq 1}\sum_{n\geq j}\sum_{m\geq 0} f_{n,j,m}\, z^n u^j v^m\,.$$

Weiters definieren wir wiederum eine Funktion $\widetilde{F}(B)$ die für jeden Baum $B \in \mathcal{B}$ definiert ist durch

$$\widetilde{F}(B) = \sum_{j=1}^{|B|} z^{|B|} u^j v^{H(B,j)}\,, \tag{3.36}$$

wobei $H(B,j)$ hier die Anzahl der Nachkommen des Knotens mit Marke j im Binärbaum B bezeichnet.

Diese Definition soll für eine Familie von Bäumen $(B_i)_{i\in I}$, $B_i \in \mathcal{B}$ erweitert werden durch

$$\widetilde{F}\left((B_i)_{i\in I}\right) := \sum_{i\in I}\widetilde{F}(B_i)\,.$$

Dann gilt wiederum

$$\widetilde{F}(\mathcal{B}) = F(z,u,v)\,, \tag{3.37}$$

wenn $F(z,u,v)$ als formale Potenzreihe aufgefaßt wird.

Wenden wir nun $\widetilde{F}$ auf die Gleichung (3.1) an, so ergibt sich aufgrund der Additivität

$$\widetilde{F}(\mathcal{B}) \quad = \quad \widetilde{F}(\Box) \quad + \quad \widetilde{F}\left(\begin{array}{ccc} & \bigcirc & \\ B' & & B'' \end{array}\right) \tag{3.38}$$

Sogleich erhält man

$$\widetilde{F}(\Box) = 0\,. \tag{3.39}$$

Weiters behandeln wir $\widetilde{F}\left(\begin{array}{ccc} & \bigcirc & \\ B' & & B'' \end{array}\right)$ für B' und B'' aus $\mathcal{B}$.

Betrachtet man getrennt die Wurzel $\bigcirc$, die Knoten im linken Unterbaum B' und die Knoten im rechten Unterbaum B'', so kommt man zu

$$\widetilde{F}\left(\begin{array}{ccc} & \bigcirc & \\ B' & & B'' \end{array}\right) = \widetilde{F}\,(\bigcirc)\,(zuv)^{|B'|}(zv)^{|B''|} + \widetilde{F}(B')zz^{|B''|} + (zu)^{|B'|}zuF(B'')\,. \tag{3.40}$$

Da ja

$$\widetilde{F}(\bigcirc) = zuv \tag{3.41}$$

gilt, ergibt sich weiter

$$\begin{aligned}\widetilde{F}\left(\begin{array}{ccc} & \bigcirc & \\ \mathcal{B} & & \mathcal{B} \end{array}\right) &= \sum_{B'\in\mathcal{B},B''\in\mathcal{B}} \widetilde{F}\left(\begin{array}{ccc} & \bigcirc & \\ B' & & B'' \end{array}\right)\\ &= zuv\sum_{B'\in\mathcal{B}}(zuv)^{|B'|}\sum_{B''\in\mathcal{B}}(zv)^{|B''|} + F(z,u,v)z\sum_{B''\in\mathcal{B}}z^{|B''|}\\ &\quad + \sum_{B'\in\mathcal{B}}(zu)^{w(B')}zuF(z,u,v)\\ &= zuvB(zuv)B(zv) + F(z,u,v)zB(z) + B(zu)zuF(z,u,v)\,.\end{aligned} \tag{3.42}$$

Damit erhält man also als Funktionalgleichung für $F(z,u,v)$

$$F(z,u,v) = zuvB(zuv)B(zv) + F(z,u,v)zB(z) + B(zu)zuF(z,u,v)\,, \tag{3.43}$$

woraus man sofort als Lösung für die erzeugende Funktion der Anzahl der Nachkommen eines entsprechend der Inorder-Traversierung markierten Knotens eines Binärbaumes

$$F(z,u,v) = \frac{zuvB(zv)B(zuv)}{1 - zB(z) - zuB(zu)} \tag{3.44}$$

folgert.

Durch Differenzieren erhält man aus (3.44) unter Verwendung der Abkürzungen (3.6)

$$G_1(z,u) = \frac{1}{4z}\left(\frac{1}{XY} - \frac{1}{X} - \frac{1}{Y} + 1\right)\,. \tag{3.45}$$

Liest man die Koeffizienten von $[z^n u^j]$ dieses Ausdrucks ab, so ergibt sich für $1 \leq j \leq n$

$$[z^n u^j] G_1(z,u) = [z^n u^j] \frac{1}{4z} \left(\frac{1}{XY} - \frac{1}{X} - \frac{1}{Y} + 1 \right) = [z^n u^j] \frac{1}{4z} \frac{1}{XY} . \tag{3.46}$$

Da ja für $1 \leq j \leq n$

$$[z^n u^j] \frac{1}{XY} = [(uz)^j] \frac{1}{\sqrt{1-4uz}} [z^{n-j}] \frac{1}{\sqrt{1-4z}} = \binom{2j}{j} \binom{2(n-j)}{n-j} \tag{3.47}$$

gilt, folgt damit aus (3.46)

$$[z^n u^j] G_1(z,u) = \frac{1}{4} \binom{2j}{j} \binom{2(n+1-j)}{n+1-j} \quad \text{für } 1 \leq j \leq n . \tag{3.48}$$

Mit der Beziehung (3.2) folgt nun aus (3.48) und (3.5)

Satz 3.3.1. *Der Erwartungswert $E_{n,j}$ der Anzahl der Nachkommen des entsprechend der Inorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist gegeben durch*

$$\boxed{E_{n,j} = \frac{n+1}{4} \frac{\binom{2j}{j}\binom{2(n+1-j)}{n+1-j}}{\binom{2n}{n}} \quad \text{für } 1 \leq j \leq n .} \tag{3.49}$$

Dies führt zu den asymptotischen Aussagen

Korollar 3.3.2. *Der Erwartungswert $E_{n,j}$ der Anzahl der Nachkommen des entsprechend der Inorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für $j \sim \rho n$ und $0 < \rho < 1$ gegeben durch*

$$\boxed{E_{n,j} \sim \frac{1}{\sqrt{\pi}\sqrt{\rho(1-\rho)}} \sqrt{n} .} \tag{3.50}$$

Korollar 3.3.3. *Der Erwartungswert $E_{n,j}$ der Anzahl der Nachkommen des entsprechend der Inorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für fixes j gegeben durch*

$$\boxed{E_{n,j} \sim \binom{2j}{j} 4^{-j} n .} \tag{3.51}$$

Korollar 3.3.4. *Der Erwartungswert $E_{n,j}$ der Anzahl der Nachkommen des entsprechend der Inorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für fixes $l = n - j$ gegeben durch*

$$\boxed{E_{n,j} \sim \binom{2(l+1)}{l+1} 4^{-(l+1)} n .} \tag{3.52}$$

Weiters erhält man durch zweimaliges Differenzieren von (3.44) und Auswerten bei $v = 1$

$$\begin{aligned} G_2(z,u) &= \frac{(1-X)(-4X^2Y^2+3X^2Y^3+X^2-XY^3+2XY^2-XY+Y^2-Y^3)}{8zX^3Y^3} \\ &= \frac{1}{8z}\left(-3+\frac{4}{X}-\frac{1}{X^3}\right)+\frac{1}{8zY}\left(4-\frac{6}{X}+\frac{1}{X^2}+\frac{1}{X^3}\right)+\frac{1}{8zY^2}\left(\frac{1}{X}-\frac{1}{X^2}\right) \\ &\quad +\frac{1}{8zY^3}\left(-1+\frac{1}{X}\right). \end{aligned} \tag{3.53}$$

Wegen der Beziehung

$$[z^n u^j]\frac{1}{X^pY^q} = [(uz)^j]\frac{1}{Y^q}\,[z^{n-j}]\frac{1}{X^p} = [z^j]\frac{1}{X^q}\,[z^{n-j}]\frac{1}{X^p} \quad \text{für } p,q\in\mathbb{N} \tag{3.54}$$

benötigt man zur weiteren Vereinfachung von (3.53) nur noch

$$\begin{aligned} [z^n]\frac{1}{X} &= \binom{2n}{n}, \\ [z^n]\frac{1}{X^2} &= [z^n]\frac{1}{1-4z} = 4^n \quad \text{und} \\ [z^n]\frac{1}{X^3} &= [z^n]\frac{1}{(1-4z)^{\frac{3}{2}}} = (2n+1)\binom{2n}{n}. \end{aligned} \tag{3.55}$$

Mit diesen Beziehungen ergibt sich schließlich für $1 \le j \le n$

$$[z^nu^j]G_2(z,u) = \frac{n-1}{4}\binom{2j}{j}\binom{2n+2-2j}{n+1-j} - \frac{1}{2}4^n + \frac{1}{8}4^j\binom{2n+2-2j}{n+1-j} + \frac{1}{8}4^{n+1-j}\binom{2j}{j}. \tag{3.56}$$

Mit (3.2) führt dies sofort zu

Hilfssatz 3.3.5. *Das zweite Faktorielle Moment $M_{n,j}^{(2)}$ der Anzahl der Nachkommen des entsprechend der Inorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist gegeben durch*

$$\boxed{\begin{aligned} M_{n,j}^{(2)} &= \frac{(n-1)(n+1)}{4}\frac{\binom{2j}{j}\binom{2n+2-2j}{n+1-j}}{\binom{2n}{n}} \\ &\quad + \frac{n+1}{\binom{2n}{n}}\left(-\tfrac{1}{2}4^n + \tfrac{1}{8}4^j\tbinom{2n+2-2j}{n+1-j} + \tfrac{1}{8}4^{n+1-j}\tbinom{2j}{j}\right) \quad \text{für } 1 \le j \le n. \end{aligned}} \tag{3.57}$$

Als weiteres Resultat erhält man mit (3.2)

Satz 3.3.6. *Die Varianz $V_{n,j}$ der Anzahl der Nachkommen des entsprechend der Inorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist gegeben durch*

$$\boxed{\begin{aligned} V_{n,j} &= \frac{n(n+1)}{4}\frac{\binom{2j}{j}\binom{2n+2-2j}{n+1-j}}{\binom{2n}{n}} - \frac{(n+1)^2}{16}\frac{\binom{2j}{j}^2\binom{2n+2-2j}{n+1-j}^2}{\binom{2n}{n}^2} \\ &\quad + \frac{n+1}{\binom{2n}{n}}\left(-\tfrac{1}{2}4^n + \tfrac{1}{8}4^j\tbinom{2n+2-2j}{n+1-j} + \tfrac{1}{8}4^{n+1-j}\tbinom{2j}{j}\right) \quad \text{für } 1 \le j \le n. \end{aligned}} \tag{3.58}$$

Man bekommt daraus die asymptotischen Beziehungen

Korollar 3.3.7. *Die Varianz $V_{n,j}$ der Anzahl der Nachkommen des entsprechend der Inorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für $j \sim \rho n$ und $0 < \rho < 1$ gegeben durch*

$$\boxed{V_{n,j} \sim \left(\frac{1}{\sqrt{\pi}\sqrt{\rho(1-\rho)}} - \frac{\sqrt{\pi}}{2} \right) n^{3/2} .} \tag{3.59}$$

Korollar 3.3.8. *Die Varianz $V_{n,j}$ der Anzahl der Nachkommen des entsprechend der Inorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für fixes j gegeben durch*

$$\boxed{V_{n,j} \sim \left(\binom{2j}{j} 4^{-j} - \binom{2j}{j}^2 4^{-2j} \right) n^2 .} \tag{3.60}$$

Korollar 3.3.9. *Die Varianz $V_{n,j}$ der Anzahl der Nachkommen des entsprechend der Inorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für fixes $l = n - j$ gegeben durch*

$$\boxed{V_{n,j} \sim \left(\binom{2(l+1)}{l+1} 4^{-(l+1)} - \binom{2(l+1)}{l+1}^2 4^{-2(l+1)} \right) n^2 .} \tag{3.61}$$

3.4 Die Höhe der Knoten bei Inorder-Traversierung

Sei hier $f_{n,j,m}$ die Anzahl der Binärbäume mit n inneren Knoten, bei denen der entsprechend der Reihenfolge der Inorder-Traversierung markierte Knoten j genau die Höhe m hat.

Die gewöhnliche erzeugende Funktion der $f_{n,j,m}$ in 3 Variablen ist dann definiert durch

$$F(z, u, v) = \sum_{j \geq 1} \sum_{n \geq j} \sum_{m \geq 0} f_{n,j,m} \, z^n u^j v^m \ .$$

Weiters definieren wir wiederum eine Funktion $\widetilde{F}(B)$ die für jeden Baum $B \in \mathcal{B}$ definiert ist durch

$$\widetilde{F}(B) = \sum_{j=1}^{|B|} z^{|B|} u^j v^{H(B,j)} , \tag{3.62}$$

wobei $H(B, j)$ die Höhe des Knotens mit Marke j im Binärbaum B bezeichnet.

Diese Definition soll für eine Familie von Bäumen $(B_i)_{i \in I}$, $B_i \in \mathcal{B}$ wiederum durch

$$\widetilde{F}\left((B_i)_{i \in I}\right) := \sum_{i \in I} \widetilde{F}(B_i)$$

erweitert werden.

Es gilt dann

$$\widetilde{F}(\mathcal{B}) = F(z,u,v)\,, \tag{3.63}$$

wenn $F(z,u,v)$ als formale Potenzreihe aufgefaßt wird.

Wenden wir nun $\widetilde{F}$ auf die Gleichung (3.1) an, so ergibt sich

$$\widetilde{F}(\mathcal{B}) \quad = \quad \widetilde{F}(\Box) \quad + \quad \widetilde{F}\left(\begin{array}{ccc} & \bigcirc & \\ \diagup & & \diagdown \\ \mathcal{B} & & \mathcal{B} \end{array}\right) \tag{3.64}$$

Unmittelbar aus der Definition von $\widetilde{F}$ erhält man

$$\widetilde{F}(\Box) = 0\,. \tag{3.65}$$

Zunächst untersuchen wir $\widetilde{F}\left(\begin{array}{ccc} & \bigcirc & \\ \diagup & & \diagdown \\ B' & & B'' \end{array}\right)$ für spezielle B' und B'' aus $\mathcal{B}$.

Betrachtet man getrennt die Wurzel $\bigcirc$, die Knoten im linken Unterbaum B' und die Knoten im rechten Unterbaum B'', so kommt man zu

$$\widetilde{F}\left(\begin{array}{ccc} & \bigcirc & \\ \diagup & & \diagdown \\ B' & & B'' \end{array}\right) = \widetilde{F}\,(\bigcirc)\,(zu)^{|B'|}z^{|B''|} + \widetilde{F}(B')zvz^{|B''|} + (zu)^{|B'|}zuvF(B'')\,. \tag{3.66}$$

Wegen

$$\widetilde{F}(\bigcirc) = zuv \tag{3.67}$$

ergibt sich weiter

$$\begin{aligned}
\widetilde{F}\left(\begin{array}{ccc} & \bigcirc & \\ \diagup & & \diagdown \\ \mathcal{B} & & \mathcal{B} \end{array}\right) &= \sum_{B'\in\mathcal{B},B''\in\mathcal{B}} \widetilde{F}\left(\begin{array}{ccc} & \bigcirc & \\ \diagup & & \diagdown \\ B' & & B'' \end{array}\right) \\
&= zuv \sum_{B'\in\mathcal{B}} (zu)^{|B'|} \sum_{B''\in\mathcal{B}} z^{|B''|} + F(z,u,v)zv \sum_{B''\in\mathcal{B}} z^{|B''|} \\
&\quad + \sum_{B'\in\mathcal{B}} (zu)^{|B'|} zuvF(z,u,v) \\
&= zuvB(zu)B(z) + F(z,u,v)zvB(z) + B(zu)zuvF(z,u,v)\,.
\end{aligned} \tag{3.68}$$

Damit lautet also eine Funktionalgleichung für $F(z,u,v)$

$$F(z,u,v) = zuvB(zu)B(z) + F(z,u,v)zvB(z) + B(zu)zuvF(z,u,v)\,, \tag{3.69}$$

woraus man sofort als Lösung für die erzeugende Funktion der Höhe der entsprechend der Inorder-Traversierung markierten Knoten in einem Binärbaum

$$F(z,u,v) = \frac{zuvB(z)B(zu)}{1 - zvB(z) - zuvB(zu)} \tag{3.70}$$

erhält.

Aus (3.70) erhält man wegen (3.2) durch Differenzieren

$$G_1(z,u) = \frac{(1-X)(1-Y)}{z(X+Y)^2} = \frac{1}{z(X+Y)^2} - \frac{1}{z(X+Y)} + \frac{XY}{z(X+Y)^2}\,. \tag{3.71}$$

Für das Ablesen der Koeffizienten von $G_1(z,u)$ benötigt man also noch die Koeffizienten von

$$\begin{aligned}[z^n u^j]\frac{1}{X+Y} &= [z^n u^j]\frac{X-Y}{X^2-Y^2} = [z^n u^j]\frac{\sqrt{1-4z}-\sqrt{1-4uz}}{-4z(1-u)}\\ &= [z^n u^j]\left(-\frac{\sqrt{1-4z}}{4z(1-u)}\right) = \frac{1}{2(n+1)}\binom{2n}{n} \quad \text{für } 0 \le j \le n\,.\end{aligned} \tag{3.72}$$

Wegen

$$\begin{aligned}1 = \frac{(X+Y)^2}{(X+Y)^2} &= \frac{X^2}{(X+Y)^2} + 2\frac{XY}{(X+Y)^2} + \frac{Y^2}{(X+Y)^2}\\ &= 2\frac{XY}{(X+Y)^2} + 2\frac{1}{(X+Y)^2} - 4\frac{z}{(X+Y)^2} - 4\frac{uz}{(X+Y)^2}\end{aligned} \tag{3.73}$$

ergeben sich mittels (3.20) auch noch die ebenfalls benötigten Koeffizienten von

$$\begin{aligned}[z^n u^j]\frac{XY}{(X+Y)^2} &= 2[z^{n-1}u^j]\frac{1}{(X+Y)^2} + 2[z^{n-1}u^{j-1}]\frac{1}{(X+Y)^2} - [z^n u^j]\frac{1}{(X+Y)^2}\\ &= -\frac{(n-1)(n-8j)+8j(j-1)}{2n(n+1)(n+2)}\binom{2j}{j}\binom{2n-2j}{n-j} \quad \text{für } 1 \le j \le n\,.\end{aligned} \tag{3.74}$$

Kombiniert man die Gleichungen (3.20), (3.72) und (3.74) erhält man aus (3.71) für $1 \le j \le n$

$$[z^n u^j]G_1(z,u) = \frac{4jn+n-4j^2+4j+1}{2(n+1)(n+2)}\binom{2j}{j}\binom{2n+2-2j}{n+1-j} - \frac{2n+1}{(n+1)(n+2)}\binom{2n}{n}\,. \tag{3.75}$$

Dies führt mit (3.2) zu

Satz 3.4.1. *Der Erwartungswert $E_{n,j}$ der Höhe des entsprechend der Inorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist gegeben durch*

$$\boxed{E_{n,j} = \frac{4jn+n-4j^2+4j+1}{2(n+2)}\frac{\binom{2j}{j}\binom{2n+2-2j}{n+1-j}}{\binom{2n}{n}} - \frac{2n+1}{n+2} \quad \text{für } 1 \le j \le n\,.} \tag{3.76}$$

Daraus ergeben sich die asymptotischen Beziehungen

Korollar 3.4.2. *Der Erwartungswert $E_{n,j}$ der Höhe des entsprechend der Inorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für $j \sim \rho n$ und $0 < \rho < 1$ gegeben durch*

$$\boxed{E_{n,j} \sim 8\frac{\sqrt{\rho(1-\rho)}}{\sqrt{\pi}}\sqrt{n}\,.} \tag{3.77}$$

Korollar 3.4.3. *Der Erwartungswert $E_{n,j}$ der Höhe des entsprechend der Inorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für fixes j gegeben durch*

$$\boxed{E_{n,j} \sim 2(4j+1)\binom{2j}{j}4^{-j} - 2\,.} \tag{3.78}$$

Korollar 3.4.4. *Der Erwartungswert $E_{n,j}$ der Höhe des entsprechend der Inorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für fixes $l = n - j$ gegeben durch*

$$\boxed{E_{n,j} \sim 2(4(l+1)+1)\binom{2(l+1)}{l+1}4^{-(l+1)} - 2\,.} \tag{3.79}$$

Für $G_2(z,u)$ bekommen wir aus (3.70) durch zweimaliges Differenzieren und Auswerten bei $v = 1$

$$\begin{aligned} G_2(z,u) &= -2\frac{(1-X)(1-Y)(X+Y-2)}{z(X+Y)^3} \\ &= \frac{4XY}{z(X+Y)^3} + \frac{4}{z(X+Y)^3} - \frac{6}{z(X+Y)^2} - \frac{2XY}{z(X+Y)^2} + \frac{2}{z(X+Y)}\,. \end{aligned} \tag{3.80}$$

Für das Ablesen der Koeffizienten dieses Ausdrucks benötigt man noch $[z^n u^j]\frac{XY}{(X+Y)^3}$. Um zu diesem zu gelangen kann man folgendermaßen vorgehen.

Aus der Gleichung

$$\frac{1}{X+Y} = \frac{(X+Y)^2}{(X+Y)^3} = 2\frac{XY}{(X+Y)^3} + \frac{X^2}{(X+Y)^3} + \frac{Y^2}{(X+Y)^3} \tag{3.81}$$

erhält man

$$\frac{XY}{(X+Y)^3} = \frac{1}{2(X+Y)} - \frac{1}{(X+Y)^3} + \frac{2z}{(X+Y)^3} + \frac{2uz}{(X+Y)^3}\,. \tag{3.82}$$

Das ergibt mit (3.72) und (3.30) weiter

$$\begin{aligned} [z^n u^j]\frac{XY}{(X+Y)^3} &= \frac{1}{2}[z^n u^j]\frac{1}{X+Y} - [z^n u^j]\frac{1}{(X+Y)^3} + 2[z^{n-1}u^j]\frac{1}{(X+Y)^3} \\ &\quad + 2[z^{n-1}u^{j-1}]\frac{1}{(X+Y)^3} \\ &= -\frac{2n^2 - 15jn - 5n + 15j^2 - 3}{4(n+1)(n+2)(n+3)}\binom{2n}{n} \quad \text{für } 1 \le j \le n\,. \end{aligned} \tag{3.83}$$

Mit den Beziehungen (3.30), (3.20), (3.72), (3.74), (3.83) folgt aus (3.80) schließlich für $1 \le j \le n$

$$\begin{aligned} [z^n u^j]G_2(z,u) &= \frac{2(2n+1)(6jn + 5n - 6j^2 + 6j + 9)}{(n+1)(n+2)(n+3)}\binom{2n}{n} \\ &- \frac{12jn^2 + 5n^2 - 12j^2n + 32jn + 14n - 20j^2 + 20j + 9}{(n+1)(n+2)(n+3)}\binom{2j}{j}\binom{2n+2-2j}{n+1-j}\,. \end{aligned} \tag{3.84}$$

Mit (3.2) erhält man nun

Hilfssatz 3.4.5. *Das zweite Faktorielle Moment $M_{n,j}^{(2)}$ der Höhe des entsprechend der Inorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist für $1 \leq j \leq n$ gegeben durch*

$$\boxed{\begin{aligned} M_{n,j}^{(2)} = {} & \frac{2(2n+1)(6jn+5n-6j^2+6j+9)}{(n+2)(n+3)} \\ & - \frac{12jn^2+5n^2-12j^2n+32jn+14n-20j^2+20j+9}{(n+2)(n+3)} \frac{\binom{2j}{j}\binom{2n+2-2j}{n+1-j}}{\binom{2n}{n}} . \end{aligned}} \tag{3.85}$$

Daher gilt weiter

Satz 3.4.6. *Die Varianz $V_{n,j}$ der Höhe des entsprechend der Inorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist gegeben durch*

$$\boxed{\begin{aligned} V_{n,j} = {} & \frac{(2n+1)(12jn^2+7n^2-12j^2n+36jn+26n-24j^2+24j+27)}{(n+2)^2(n+3)} \\ & - \frac{4jn^3+5n^3-4j^2n^2+16jn^2+24n^2-12j^2n+44jn+43n-32j^2+32j+24}{2(n+2)^2(n+3)} \frac{\binom{2j}{j}\binom{2n+2-2j}{n+1-j}}{\binom{2n}{n}} \\ & - \frac{(4jn+n-4j^2+4j+1)^2}{4(n+2)^2} \frac{\binom{2j}{j}^2\binom{2n+2-2j}{n+1-j}^2}{\binom{2n}{n}^2} \quad \text{für } 1 \leq j \leq n . \end{aligned}} \tag{3.86}$$

Daraus leitet man die folgenden asymptotischen Aussagen ab.

Korollar 3.4.7. *Die Varianz $V_{n,j}$ der Höhe des entsprechend der Inorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für $j \sim \rho n$ und $0 < \rho < 1$ gegeben durch*

$$\boxed{V_{n,j} \sim 8\frac{\rho(1-\rho)(3\pi-8)}{\pi} n .} \tag{3.87}$$

Korollar 3.4.8. *Die Varianz $V_{n,j}$ der Höhe des entsprechend der Inorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für fixes j gegeben durch*

$$\boxed{V_{n,j} \sim 24j + 14 - 2(4j+5)\binom{2j}{j}4^{-j} - 4(4j+1)^2\binom{2j}{j}^2 4^{-2j} .} \tag{3.88}$$

Korollar 3.4.9. *Die Varianz $V_{n,j}$ der Höhe des entsprechend der Inorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für fixes $l = n - j$ gegeben durch*

$$\boxed{V_{n,j} \sim 24(l+1) + 14 - 2(4(l+1)+5)\binom{2(l+1)}{l+1}4^{-(l+1)} - 4(4(l+1)+1)^2\binom{2(l+1)}{l+1}^2 4^{-2(l+1)} .} \tag{3.89}$$

3.5 Die Anzahl der Nachkommen bei Preorder-Traversierung

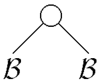
Wir behandeln jetzt Fragen betreffend die Anzahl der Nachkommen von entsprechend der Reihenfolge der Preorder-Traversierung von 1 bis n markierten Knoten in Binärbäumen. Bezeichnen wir also mit $f_{n,j,m}$ die Anzahl der Binärbäume mit n inneren Knoten, bei denen der entsprechend der Preorder-Traversierung markierte Knoten j genau m Nachkommen hat.

Die erzeugende Funktion in 3 Variablen $F(z,u,v)$ der $f_{n,j,m}$, ist also definiert durch

$$F(z,u,v) = \sum_{j\geq 1}\sum_{n\geq j}\sum_{m\geq 0} f_{n,j,m}\, z^n u^j v^m \ .$$

Analog zu der in Abschnitt 3.2 bzw. Abschnitt 3.3 besprochenen Methode erhält man aus der formalen Gleichung (3.1) als Funktionalgleichung für $F(z,u,v)$

$$F(z,u,v) = zuvB^2(zv) + F(z,u,v)zuB(z) + B(zu)zuF(z,u,v) \ , \tag{3.90}$$

wobei sich die rechte Seite dieser Gleichung wieder ergibt, indem man 

aufspaltet in "Wurzel" + "linker Unterbaum" + "rechter Unterbaum".

Dies führt also zu folgender expliziten Gleichung für $F(z,u,v)$:

$$F(z,u,v) = \frac{zuvB^2(zv)}{1 - zuB(z) - zuB(zu)} \ . \tag{3.91}$$

Ableiten von (3.91) nach v und Auswerten an der Stelle $v = 1$ führt dann zu

$$G_1(z,u) = \frac{2(1-Y)}{X(1+X)(X+Y)} = \frac{2}{X(X+Y)} + \frac{2}{1+X} - \frac{2}{X} \ . \tag{3.92}$$

Von den auf der rechten Seite dieser Gleichung stehenden Summanden ergibt beim Ablesen der Koeffizienten für $0 \leq j \leq n$ nur der erste Summand einen Beitrag, es gilt also

$$[z^n u^j]G_1(z,u) = \frac{1}{2}[z^n u^j]\frac{1}{X(X+Y)} \ . \tag{3.93}$$

Mit den Gleichungen (3.72) und (3.55) schreibt sich der gesuchte Koeffizient als Summe

$$[z^n u^j]\frac{1}{X(X+Y)} = \sum_{k=j}^{n}[z^k u^j]\frac{1}{X+Y}[z^{n-k}]\frac{1}{X} = \sum_{k=j}^{n}\frac{1}{2(k+1)}\binom{2k}{k}\binom{2n-2k}{n-k} \ . \tag{3.94}$$

Die Frage ob diese Summe hypergeometrischer Terme einen hypergeometrischen Term als Lösung besitzt, läßt sich nun mittels Gosper's Algorithmus entscheiden, denn dieser Algorithmus gibt eine Antwort auf die Frage, ob eine Summe dieser Bauart indefinit in einfacher Form summierbar ist.

Da Zeilberger's Algorithmus quasi Gosper's Algorithmus beinhaltet, wenden wir jenen auf den Summanden

$$F(n,k) = \frac{1}{2(k+1)}\binom{2k}{k}\binom{2n-2k}{n-k}$$

an und erhalten so

$$\begin{aligned}(-n-1)F(n,k) &= G(n,k+1) - G(n,k)\\ \text{mit}\quad G(n,k) &= (k+1)(2n-2k+1)F(n,k)\ .\end{aligned} \tag{3.95}$$

Das bedeutet, daß $F(n,k)$ indefinit summierbar ist.

Schreiben wir abkürzend für unsere gesuchte Summe

$$f(n,j) = \sum_{k=j}^{n} F(n,k) = [z^n u^j]\frac{1}{X(X+Y)}\ ,$$

dann ergibt sich durch Summieren von (3.95)

$$\begin{aligned}\sum_{k=j}^{n-1}(-n-1)F(n,k) &= (-n-1)\left(f(n,j) - f(n,n)\right)\\ &= \sum_{k=j}^{n-1}\left(G(n,k+1) - G(n,k)\right) = G(n,n) - G(n,j)\end{aligned} \tag{3.96}$$

und weiter

$$\begin{aligned}[z^n u^j]\frac{1}{X(X+Y)} &= f(n,j) = f(n,n) + \frac{G(n,j) - G(n,n)}{n+1}\\ &= \frac{2n+1-2j}{2(n+1)}\binom{2j}{j}\binom{2n-2j}{n-j} \quad \text{für } 1 \le j \le n\ .\end{aligned} \tag{3.97}$$

Also folgt für die Koeffizienten von $G_1(z,u)$:

$$[z^n u^j]G_1(z,u) = \frac{2n+1-2j}{n+1}\binom{2j}{j}\binom{2n-2j}{n-j}\ . \tag{3.98}$$

Mit (3.2) erhält man sofort weiter

Satz 3.5.1. *Der Erwartungswert $E_{n,j}$ der Anzahl der Nachkommen des entsprechend der Preorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist gegeben durch*

$$\boxed{E_{n,j} = (2n+1-2j)\frac{\binom{2j}{j}\binom{2n-2j}{n-j}}{\binom{2n}{n}} \quad \text{für } 1 \le j \le n\,.} \tag{3.99}$$

Daraus ergeben sich die asymptotischen Beziehungen

Korollar 3.5.2. *Der Erwartungswert $E_{n,j}$ der Anzahl der Nachkommen des entsprechend der Preorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für $j \sim \rho n$ und $0 < \rho < 1$ gegeben durch*

$$\boxed{E_{n,j} \sim 2\frac{\sqrt{1-\rho}}{\sqrt{\pi}\sqrt{\rho}}\sqrt{n}\,.} \tag{3.100}$$

Korollar 3.5.3. *Der Erwartungswert $E_{n,j}$ der Anzahl der Nachkommen des entsprechend der Preorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für fixes j gegeben durch*

$$\boxed{E_{n,j} \sim 2\binom{2j}{j}4^{-j}n\,.} \tag{3.101}$$

Korollar 3.5.4. *Der Erwartungswert $E_{n,j}$ der Anzahl der Nachkommen des entsprechend der Preorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für fixes $l = n - j$ gegeben durch*

$$\boxed{E_{n,j} \sim (2l+1)\binom{2l}{l}4^{-l}\,.} \tag{3.102}$$

Weiters erhalten wir für $G_2(z,u)$ aus (3.91) den folgenden Ausdruck

$$\begin{aligned} G_2(z,u) &= \frac{(1-Y)(1-X)(1+3X)}{X^3(1+X)(X+Y)} \\ &= \frac{1}{X^3(X+Y)} + \frac{2}{X^2(X+Y)} - \frac{3}{X(X+Y)} + \frac{4}{X} - \frac{1}{X^2} - \frac{1}{X^3} - \frac{4}{X+1}\,. \end{aligned} \tag{3.103}$$

Um die Koeffizienten dieses Ausdrucks ablesen zu können, benötigen wir noch die Entwicklungen von $\frac{1}{X^2(X+Y)}$ and $\frac{1}{X^3(X+Y)}$.
Aus der Gleichung

$$\frac{1}{X^2(X+Y)} = \frac{1}{X^2Y} + \frac{1}{4z(1-u)Y} - \frac{1}{4z(1-u)X}$$

erhalten wir die Koeffizienten von $\frac{1}{X^2(X+Y)}$:

$$[z^nu^j]\frac{1}{X^2(X+Y)} = 4^{n-j}\binom{2j}{j} - \frac{1}{4}\binom{2(n+1)}{n+1} \quad \text{für } 1 \le j \le n\,. \tag{3.104}$$

Als Koeffizienten von $\frac{1}{X^3(X+Y)}$ erhalten wir zunächst die folgende Summe

$$\begin{aligned} [z^nu^j]\frac{1}{X^3(X+Y)} &= \sum_{k=j}^{n} \frac{1}{2(k+1)}\binom{2k}{k}(2(n-k)+1)\binom{2(n-k)}{n-k} \\ &= (2n+3)\sum_{k=j}^{n}\frac{1}{2(k+1)}\binom{2k}{k}\binom{2(n-k)}{n-k} - \sum_{k=j}^{n}\binom{2k}{k}\binom{2(n-k)}{n-k}\,, \end{aligned}$$

welche unter Verwendung von (3.94) und (3.97) weiter vereinfacht werden kann, und man erhält schließlich für $1 \le j \le n$

$$[z^nu^j]\frac{1}{X^3(X+Y)} = \frac{(2n+3)(2n+1-2j)}{2(n+1)}\binom{2j}{j}\binom{2(n-j)}{n-j} - \sum_{k=j}^{n}\binom{2k}{k}\binom{2(n-k)}{n-k}\,. \tag{3.105}$$

Die verbleibende Summe dieses Ausdrucks kann nicht in eine geschlossene Form gebracht werden, da mit Gosper's Algorithmus gezeigt werden kann, daß dieser Ausdruck nicht indefinit summierbar ist (siehe [34]).

Kombiniert man nun die gesammelten Resultate, erhält man daraus

$$\begin{aligned}[z^n u^j]G_2(z,u) = {}& \frac{n(2n-2j+1)}{n+1}\binom{2j}{j}\binom{2n-2j}{n-j} + 2\cdot 4^{n-j}\binom{2j}{j} - \frac{1}{2}\binom{2n+2}{n+1} \\ & - \sum_{k=j}^{n}\binom{2k}{k}\binom{2n-2k}{n-k} \quad \text{für } 1 \le j \le n\,,\end{aligned} \tag{3.106}$$

und daher

Hilfssatz 3.5.5. *Das zweite Faktorielle Moment $M_{n,j}^{(2)}$ der Anzahl der Nachkommen des entsprechend der Preorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist gegeben durch*

$$\boxed{\begin{aligned}M_{n,j}^{(2)} = {}& n(2n-2j+1)\frac{\binom{2j}{j}\binom{2n-2j}{n-j}}{\binom{2n}{n}} + 2(n+1)\frac{4^{n-j}\binom{2j}{j}}{\binom{2n}{n}} - 2n - 1 \\ & - \frac{n+1}{\binom{2n}{n}}\sum_{k=j}^{n}\binom{2k}{k}\binom{2n-2k}{n-k} \quad \text{für } 1 \le j \le n\,.\end{aligned}} \tag{3.107}$$

Dies führt unmittelbar zu

Satz 3.5.6. *Die Varianz $V_{n,j}$ der Anzahl der Nachkommen des entsprechend der Preorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist für $1 \le j \le n$ gegeben durch*

$$\boxed{\begin{aligned}V_{n,j} = {}& (n+1)(2n-2j+1)\frac{\binom{2j}{j}\binom{2n-2j}{n-j}}{\binom{2n}{n}} + 2(n+1)\frac{4^{n-j}\binom{2j}{j}}{\binom{2n}{n}} - 2n - 1 \\ & - (2n+1-2j)^2\frac{\binom{2j}{j}^2\binom{2n-2j}{n-j}^2}{\binom{2n}{n}^2} - \frac{n+1}{\binom{2n}{n}}\sum_{k=j}^{n}\binom{2k}{k}\binom{2n-2k}{n-k}\,.\end{aligned}} \tag{3.108}$$

Es ist zu bemerken, daß die Summe über alle k einfach ausgewertet werden kann. Es gilt nämlich

$$\sum_{k=0}^{n}\binom{2k}{k}\binom{2n-2k}{n-k} = 4^n\,. \tag{3.109}$$

Wegen der im letzten Ausdruck enthaltenen Summe gestalten sich die asymptotischen Auswertungen etwas komplizierter. Um für ein festes Verhältnis $\rho = j/n$ eine asymptotische Beziehung zu erhalten, benötigt man einen Satz, der in der Wahrscheinlichkeitstheorie als *Arcussinus-Gesetz* bekannt ist.

Satz 3.5.7. *Für $0 < \rho < 1$ gilt*

$$\boxed{\frac{1}{4^n}\sum_{k=0}^{\rho n}\binom{2k}{k}\binom{2(n-k)}{n-k} \sim \frac{2}{\pi}\arcsin\sqrt{\rho}\,.} \tag{3.110}$$

Beweis. Für den Beweis spalten wir für die Summe bei $k = n^{1/4}$ auf in die beiden Anteile

$$\sum_{k=0}^{n^{1/4}}\binom{2k}{k}\binom{2(n-k)}{n-k} \quad \text{und} \quad \sum_{k=n^{1/4}}^{\rho n}\binom{2k}{k}\binom{2(n-k)}{n-k}\,.$$

Da die Summanden der auszuwertenden Summe für $0 \le k < n/2$ monoton fallend sind, schätzen wir die Summanden mit $k < n^{1/4}$ durch $k = 0$ ab. Dies ergibt

$$\frac{1}{4^n}\sum_{k=0}^{n^{1/4}}\binom{2k}{k}\binom{2(n-k)}{n-k} = \mathcal{O}\left(n^{1/4}\frac{\binom{2n}{n}}{4^n}\right) = \mathcal{O}\left(n^{1/4}\frac{1}{\sqrt{\pi}\sqrt{n}}\right) = \mathcal{O}\left(n^{-1/4}\right)\,.$$

Für die Summanden mit $n^{1/4} \le k \le \rho n$ verwenden wir die asymptotischen Beziehungen

$$\binom{2k}{k} = \frac{4^k}{\sqrt{\pi}\sqrt{k}}\left(1 + \mathcal{O}\left(\frac{1}{k}\right)\right) \quad \text{und} \quad \binom{2(n-k)}{n-k} = \frac{4^{n-k}}{\sqrt{\pi}\sqrt{n-k}}\left(1 + \mathcal{O}\left(\frac{1}{n-k}\right)\right)\,.$$

Dies ergibt

$$\begin{aligned}\frac{1}{4^n}\binom{2k}{k}\binom{2(n-k)}{n-k} &= \frac{1}{\pi\sqrt{k(n-k)}}\cdot\left(1 + \mathcal{O}\left(\frac{1}{k}\right) + \mathcal{O}\left(\frac{1}{n-k}\right) + \mathcal{O}\left(\frac{1}{k}\right)\mathcal{O}\left(\frac{1}{n-k}\right)\right)\\ &= \frac{1}{\pi\sqrt{k(n-k)}}\cdot\left(1 + \mathcal{O}\left(n^{-1/4}\right)\right) \quad \text{für } n^{1/4} < k < \rho n\,.\end{aligned}$$

Das liefert weiter

$$\frac{1}{4^n}\sum_{k=n^{1/4}}^{\rho n}\binom{2k}{k}\binom{2(n-k)}{n-k} = \frac{1}{\pi}\sum_{k=n^{1/4}}^{\rho n}\frac{1}{\sqrt{k(n-k)}}\cdot\left(1 + \mathcal{O}\left(n^{-1/4}\right)\right)\,.$$

Die Summe

$$\frac{1}{\pi}\sum_{k=n^{1/4}}^{\rho n}\frac{1}{\sqrt{k(n-k)}} = \frac{1}{\pi}\sum_{\frac{k}{n}=n^{-3/4}}^{\rho}\frac{1}{n\sqrt{\frac{k}{n}(1-\frac{k}{n})}}$$

kann als Riemann'sche Zwischensumme von

$$\frac{1}{\pi}\int_{n^{-3/4}}^{\rho}\frac{1}{\sqrt{x(1-x)}}dx$$

aufgefaßt werden.

Der Fehler der dabei gemacht wird, kann durch die Summanden am Rand des Intervalls abgeschätzt werden. Man erhält daraus

$$\begin{aligned}\frac{1}{\pi}\sum_{k=n^{1/4}}^{\rho n}\frac{1}{\sqrt{k(n-k)}} &= \frac{1}{\pi}\int_{n^{-3/4}}^{\rho}\frac{1}{\sqrt{x(1-x)}}dx+\mathcal{O}\Bigg(\frac{1}{\sqrt{n^{1/4}(n-n^{1/4})}}\Bigg)+\mathcal{O}\Bigg(\frac{1}{\sqrt{\rho n(n-\rho n)}}\Bigg)\\ &= \frac{1}{\pi}\int_{n^{-3/4}}^{\rho}\frac{1}{\sqrt{x(1-x)}}dx+\mathcal{O}\left(n^{-5/8}\right).\end{aligned}$$

Das dabei auftretende Integral läßt sich in geschlossener Form berechnen, und man erhält

$$\frac{1}{\pi}\int\frac{1}{\sqrt{x(1-x)}}dx=\frac{2}{\pi}\arcsin\sqrt{x}+C\,,$$

mit einer Konstanten C.

Es gilt also

$$\frac{1}{\pi}\int_{n^{-3/4}}^{\rho}\frac{1}{\sqrt{x(1-x)}}dx=\frac{2}{\pi}\left(\arcsin\sqrt{\rho}-\arcsin\sqrt{n^{-3/4}}\right)=\frac{2}{\pi}\arcsin\sqrt{\rho}+\mathcal{O}\left(n^{-3/8}\right).$$

Wir bekommen so

$$\frac{1}{4^n}\sum_{k=n^{1/4}}^{\rho n}\binom{2k}{k}\binom{2(n-k)}{n-k}=\frac{2}{\pi}\arcsin\sqrt{\rho}+\mathcal{O}\left(n^{-1/4}\right).$$

Kombiniert man die gesammelten Ergebnisse, erhält man schließlich

$$\frac{1}{4^n}\sum_{k=0}^{\rho n}\binom{2k}{k}\binom{2(n-k)}{n-k}=\frac{2}{\pi}\arcsin\sqrt{\rho}+\mathcal{O}\left(n^{-1/4}\right).$$

□

Nun ergeben sich die asymptotischen Beziehungen

Korollar 3.5.8. *Die Varianz $V_{n,j}$ der Anzahl der Nachkommen des entsprechend der Preorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für $j \sim \rho n$ und $0 < \rho < 1$ gegeben durch*

$$\boxed{V_{n,j}\sim\frac{2}{\sqrt{\pi}}\left(\frac{\sqrt{1-\rho}}{\sqrt{\rho}}-\arcsin\sqrt{1-\rho}\right)n^{3/2}.} \tag{3.111}$$

Korollar 3.5.9. *Die Varianz $V_{n,j}$ der Anzahl der Nachkommen des entsprechend der Preorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für fixes j gegeben durch*

$$\boxed{V_{n,j}\sim\left(2\binom{2j}{j}4^{-j}-\binom{2j}{j}^2 4^{-2j+1}\right)n^2.} \tag{3.112}$$

Korollar 3.5.10. *Die Varianz $V_{n,j}$ der Anzahl der Nachkommen des entsprechend der Preorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für fixes $l = n - j$ gegeben durch*

$$\boxed{V_{n,j} \sim \frac{2}{3} l(l+1) \binom{2(l+1)}{l+1} 4^{-(l+1)} + l + 1 - 4(l+1)^2 \binom{2(l+1)}{l+1}^2 4^{-2(l+1)} .} \tag{3.113}$$

3.6 Die Höhe der Knoten bei Preorder-Traversierung

Es bezeichne $f_{n,j,m}$ die Anzahl der Binärbäume mit n inneren Knoten, bei denen der entsprechend der Preorder-Traversierung markierte Knoten j genau die Höhe m hat.

Für die durch

$$F(z,u,v) = \sum_{j \geq 1} \sum_{n \geq j} \sum_{m \geq 0} f_{n,j,l}\, z^n u^j v^m$$

definierte erzeugende Funktion $F(z,u,v)$ in drei Variablen ergibt sich mit der schon mehrmals verwendeten Methode aus Abschnitt 3.1 die Gleichung

$$F(z,u,v) = zuvB^2(z) + F(z,u,v)zuvB(z) + B(zu)zuvF(z,u,v) . \tag{3.114}$$

Man bekommt also explizit

$$F(z,u,v) = \frac{zuvB^2(z)}{1 - zuvB(z) - zuvB(zu)} . \tag{3.115}$$

Um die Erwartungswerte der Höhe der Knoten zu erhalten, berechnen wir wieder

$$\begin{aligned} G_1(z,u) &= \frac{4(1-Y)}{(1+Y)(X+Y)^2} = \frac{(1-Y)^2}{uz(X+Y)^2} \\ &= \frac{2}{uz(X+Y)^2} - \frac{4}{(X+Y)^2} - \frac{2Y}{uz(X+Y)^2} . \end{aligned} \tag{3.116}$$

Um zu den Koeffizienten obigen Ausdrucks zu gelangen, benötigen wir noch $[z^n u^j]\frac{Y}{(X+Y)^2}$. Wir können dazu folgendermaßen vorgehen. Mittels Gleichung

$$\frac{Y}{(X+Y)^2} = \frac{Y(X+Y)}{(X+Y)^3} = \frac{XY}{(X+Y)^3} + \frac{1-4uz}{(X+Y)^3} \tag{3.117}$$

und den Beziehungen (3.30) und (3.83) erhält man daraus

$$\begin{aligned} [z^n u^j]\frac{Y}{(X+Y)^2} &= [z^n u^j]\frac{XY}{(X+Y)^3} + [z^n u^j]\frac{1}{(X+Y)^3} - 4[z^{n-1}u^{j-1}]\frac{1}{(X+Y)^3} \\ &= \frac{2n+1-3j}{2(n+1)(n+2)}\binom{2n}{n} \quad \text{für } 1 \leq j \leq n . \end{aligned} \tag{3.118}$$

Damit ergibt sich unter Verwendung von (3.20) und (3.118) für $1 \le j \le n$

$$\begin{aligned}[z^n u^j]G_1(z,u) &= \frac{(jn+2n-j)(2n-2j+1)}{(n+1)(n+2)(n+3)}\binom{2j+2}{j+1}\binom{2n-2j}{n-j} \\ &\quad - \frac{2n-3j}{(n+2)(n+3)}\binom{2n+2}{n+1},\end{aligned} \tag{3.119}$$

und weiter mit Beziehung (3.2)

Satz 3.6.1. *Der Erwartungswert $E_{n,j}$ der Höhe des entsprechend der Preorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist gegeben durch*

$$\boxed{\begin{aligned}E_{n,j} = &\frac{(jn+2n-j)(2n-2j+1)}{(n+2)(n+3)}\frac{\binom{2j+2}{j+1}\binom{2n-2j}{n-j}}{\binom{2n}{n}} \\ &- \frac{2(2n+1)(2n-3j)}{(n+2)(n+3)} \quad \text{für } 1 \le j \le n\,.\end{aligned}} \tag{3.120}$$

Daraus ergeben sich die asymptotischen Beziehungen

Korollar 3.6.2. *Der Erwartungswert $E_{n,j}$ der Höhe des entsprechend der Preorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für $j \sim \rho n$ und $0 < \rho < 1$ gegeben durch*

$$\boxed{E_{n,j} \sim 8\frac{\sqrt{\rho(1-\rho)}}{\sqrt{\pi}}\sqrt{n}\,.} \tag{3.121}$$

Korollar 3.6.3. *Der Erwartungswert $E_{n,j}$ der Höhe des entsprechend der Preorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für fixes j gegeben durch*

$$\boxed{E_{n,j} \sim 2(j+2)\binom{2(j+1)}{j+1}4^{-j} - 8\,.} \tag{3.122}$$

Korollar 3.6.4. *Der Erwartungswert $E_{n,j}$ der Höhe des entsprechend der Preorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für fixes $l = n - j$ gegeben durch*

$$\boxed{E_{n,j} \sim 4(2l+1)\binom{2l}{l}4^{-l} + 4\,.} \tag{3.123}$$

Leiten wir (3.115) zweimal nach v ab und evaluieren wir bei $v = 1$, dann ergibt sich

$$\begin{aligned}G_2(z,u) &= 8\frac{(1-Y)^2(X+Y+2)}{(X+Y)^3(1+Y)^2} \\ &= \frac{8(1-4uz+2u^2z^2)}{u^2z^2}\frac{1}{(X+Y)^3} - \frac{8(1-2uz)}{u^2z^2}\frac{Y}{(X+Y)^3} \\ &\quad + \frac{4(1-4uz+2u^2z^2)}{u^2z^2}\frac{1}{(X+Y)^2} - \frac{4(1-2uz)}{u^2z^2}\frac{Y}{(X+Y)^2}\,.\end{aligned} \tag{3.124}$$

Für das Ablesen der Koeffizienten obigen Ausdrucks benötigen wir noch $[z^n u^j]\frac{Y}{(X+Y)^3}$.
Zunächst erhalten wir folgende Summe

$$
\begin{aligned}
[z^n u^j]\frac{Y}{(X+Y)^3} &= \sum_{k=0}^{j}[(uz)^k]Y \cdot [z^{n-k}u^{j-k}]\frac{1}{(X+Y)^3} \\
&= \sum_{k=0}^{j} -\frac{3}{2}\frac{(2k-2)!}{k!(k-1)!}\frac{(2(n-k)+1)(j-k+1)(n-j+1)}{(n-k+1)(n-k+2)(n-k+3)}\binom{2(n-k)}{n-k},
\end{aligned}
\tag{3.125}
$$

die gültig ist, wenn wir $\frac{(-2)!}{(-1)!} = -\frac{1}{2}$ definieren.
Dies ergibt Sinn, denn mit der Reflexionsformel der Gammafunktion

$$(-z)! = \frac{\pi}{(z-1)!\sin(\pi z)}$$

erhält man zunächst

$$\frac{(-2z)!}{(-z)!} = \frac{\pi}{(2z-1)!\sin(2z\pi)}\frac{(z-1)!\sin(z\pi)}{\pi} = \frac{(z-1)!}{2(2z-1)!\cos(z\pi)} \quad \text{für } z \in \mathbb{C}\setminus\mathbb{Z} \tag{3.126}$$

und daher den Grenzwert

$$\lim_{z\to 1}\frac{(-2z)!}{(-z)!} = -\frac{1}{2}\,. \tag{3.127}$$

Unterwerfen wir also den Summanden

$$F(n,k) = -\frac{3}{2}\frac{(2k-2)!}{k!(k-1)!}\frac{(2(n-k)+1)(j-k+1)(n-j+1)}{(n-k+1)(n-k+2)(n-k+3)}\binom{2(n-k)}{n-k} \tag{3.128}$$

Zeilberger's Algorithmus, dann liefert dieser

$$
\begin{aligned}
3(n+1)(n+2)(n+3)F(n,k) &= G(n,k+1) - G(n,k) \\
\text{mit}\quad G(n,k) &= R(n,k)F(n,k)
\end{aligned}
\tag{3.129}
$$

und $R(n,k) = -\frac{k(2n+3-2k)(34j-9k+5n+21jn-12jnk+3jn^2+6nk+13+3n^2k-36kj+2k^2-2nk^2+8k^2j)}{j+1-k}$.
Sei abkürzend

$$f(n,j) = \sum_{k=0}^{j} F(n,k) = [z^n u^j]\frac{Y}{(X+Y)^3}$$

definiert, so erhält man durch Aufsummieren

$$
\begin{aligned}
\sum_{k=1}^{j-1} 3(n+1)(n+2)(n+3)F(n,k) &= 3(n+1)(n+2)(n+3)(f(n,j) - F(n,0) - F(n,j)) \\
&= \sum_{k=1}^{j-1}(G(n,k+1) - G(n,k)) = G(n,j) - G(n,1)\,,
\end{aligned}
\tag{3.130}
$$

woraus sich schließlich

$$[z^n u^j]\frac{Y}{(X+Y)^3} = f(n,j) = F(n,0) + F(n,j) + \frac{G(n,j)-G(n,1)}{3(n+1)(n+2)(n+3)}$$
$$= \frac{1}{4}\frac{(2j+1)(2n+1-2j)(3n-4j+3)}{(n+1)(n+2)(n+3)}\binom{2j}{j}\binom{2n-2j}{n-j} \quad \text{für } 1 \le j \le n \tag{3.131}$$

ergibt.

Aus (3.124) erhält man nun unter Verwendung von (3.20), (3.30), (3.118) und (3.131) für $1 \le j \le n$

$$\begin{aligned}[z^n u^j]G_2(z,u) = \Big[&4(2n+1)(3jn^3+23n^3-3j^2n^2-51jn^2+18n^2+33j^2n-42jn \\ &+19n+30j^2-30j)\Big]\frac{1}{(n+1)(n+2)(n+3)(n+4)(n+5)}\binom{2n}{n} \\ &-\Big[(n-j+1)(11j^2n^3+47jn^3+46n^3-16j^3n^2-68j^2n^2-44jn^2+36n^2 \\ &-35j^2n-35jn+38n+16j^3+20j^2-16j)\Big]\times \\ &\times\frac{1}{(j+2)(n+1)(n+2)(n+3)(n+4)(n+5)}\binom{2j+2}{j+1}\binom{2n+2-2j}{n+1-j},\end{aligned} \tag{3.132}$$

und weiter

Hilfssatz 3.6.5. *Das zweite Faktorielle Moment $M_{n,j}^{(2)}$ der Höhe des entsprechend der Preorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist für $1 \le j \le n$ gegeben durch*

$$\boxed{\begin{aligned}M_{n,j}^{(2)} = \Big[&4(2n+1)(3jn^3+23n^3-3j^2n^2-51jn^2+18n^2+33j^2n-42jn \\ &+19n+30j^2-30j)\Big]\frac{1}{(n+2)(n+3)(n+4)(n+5)} \\ &-\Big[(n-j+1)(11j^2n^3+47jn^3+46n^3-16j^3n^2-68j^2n^2-44jn^2 \\ &+36n^2-35j^2n-35jn+38n+16j^3+20j^2-16j)\Big]\times \\ &\times\frac{1}{(j+2)(n+2)(n+3)(n+4)(n+5)}\frac{\binom{2j+2}{j+1}\binom{2n+2-2j}{n+1-j}}{\binom{2n}{n}}.\end{aligned}} \tag{3.133}$$

Daraus erhält man nun

Satz 3.6.6. *Die Varianz $V_{n,j}^{(2)}$ der Höhe des entsprechend der Preorder-Traversierung mit*

Marke j versehenen Knotens in einem Binärbaum der Größe n ist gegeben durch

$$\boxed{\begin{aligned} V_{n,j} = \Big[& 2n(2n+1)(6n^4 j + 28n^4 - 6n^3 j^2 - 21n^3 j + 86n^3 - 60jn^2 - 40n^2 \\ & + 12j^2 n + 297jn - 62n - 186j^2 + 138j - 12)\Big] \times \\ & \times \frac{1}{(n+2)^2(n+3)^2(n+4)(n+5)} \\ & - \Big[2(2j+1)(2n-2j+1)(5j^2n^5 + 26jn^5 + 24n^5 - 8j^3n^4 - 79j^2n^4 \\ & \quad - 152jn^4 - 132n^4 + 44j^3n^3 - 51j^2n^3 - 554jn^3 - 864n^3 + 200j^3n^2 \\ & \quad + 919j^2n^2 + 1112jn^2 - 444n^2 - 188j^3n - 242j^2n + 528jn - 24n \\ & \quad - 48j^3 - 120j^2 + 48j)\Big] \times \\ & \times \frac{1}{(j+1)(j+2)(n+2)^2(n+3)^2(n+4)(n+5)} \frac{\binom{2j}{j}\binom{2n-2j}{n-j}}{\binom{2n}{n}} \\ & - \frac{4(jn+2n-j)^2(2n-2j+1)^2}{(n+2)^2(n+3)^2} \frac{\binom{2j+2}{j+1}^2 \binom{2n-2j}{n-j}^2}{\binom{2n}{n}^2} \quad \text{für } 1 \le j \le n\,. \end{aligned}} \tag{3.134}$$

Dies führt zu den asymptotischen Beziehungen

Korollar 3.6.7. *Die Varianz $V_{n,j}$ der Höhe des entsprechend der Preorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für $j \sim \rho n$ und $0 < \rho < 1$ gegeben durch*

$$\boxed{V_{n,j} \sim 8\frac{\rho(1-\rho)(3\pi - 8)}{\pi} n\,.} \tag{3.135}$$

Korollar 3.6.8. *Die Varianz $V_{n,j}$ der Höhe des entsprechend der Preorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für fixes j gegeben durch*

$$\boxed{V_{n,j} \sim 24j + 112 - 4(j+2)^2 \binom{2(j+1)}{j+1}^2 4^{-2j} - \frac{2(j+4)(5j+6)}{j+2} \binom{2(j+1)}{j+1} 4^{-j}\,.} \tag{3.136}$$

Korollar 3.6.9. *Die Varianz $V_{n,j}$ der Höhe des entsprechend der Preorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für fixes $l = n - j$ gegeben durch*

$$\boxed{V_{n,j} \sim 24l + 28 - 16(2l+1)^2 \binom{2l}{l}^2 4^{-2l} + 12(2l+1)\binom{2l}{l} 4^{-l}\,.} \tag{3.137}$$

3.7 Die Anzahl der Nachkommen bei Postorder-Traversierung

Nun betrachten wir die Anzahl der Nachkommen von entsprechend der Reihenfolge der Postorder-Traversierung von 1 bis n markierten Knoten in Binärbäumen. Bezeichnen wir also mit $f_{n,j,m}$ die Anzahl der Binärbäume mit n inneren Knoten, bei denen der entsprechend der Postorder-Traversierung markierte Knoten j genau m Nachkommen hat.

Die erzeugende Funktion in 3 Variablen $F(z,u,v)$ der $f_{n,j,m}$, ist dann definiert durch

$$F(z,u,v) = \sum_{j\geq 1}\sum_{n\geq j}\sum_{m\geq 0} f_{n,j,m}\, z^n u^j v^m .$$

Aus der formalen Gleichung (3.1) erhält man als Funktionalgleichung für $F(z,u,v)$

$$F(z,u,v) = zuvB^2(zuv) + F(z,u,v)zB(z) + B(zu)zF(z,u,v) , \tag{3.138}$$

woraus sich $F(z,u,v)$ ergibt als

$$F(z,u,v) = \frac{zuvB^2(zuv)}{1-zB(z)-zB(zu)} . \tag{3.139}$$

Daraus berechnet man zunächst

$$G_1(z,u) = \frac{2(1-Y)}{Y(1+X)(X+Y)} = \frac{1}{2zXY} - \frac{2}{X(X+Y)} - \frac{1}{2zY} . \tag{3.140}$$

Das Ablesen der Koeffizienten dieses Ausdrucks ergibt mit den Beziehungen (3.47) und (3.97) für $1 \leq j \leq n$

$$[z^n u^j]G_1(z,u) = \frac{1}{2}[z^{n+1}u^j]\frac{1}{XY} - 2[z^n u^j]\frac{1}{X(X+Y)} = \frac{1}{2}\frac{j}{n+1}\binom{2j}{j}\binom{2n+2-2j}{n+1-j} . \tag{3.141}$$

Damit ergibt sich sofort

Satz 3.7.1. *Der Erwartungswert $E_{n,j}$ der Anzahl der Nachkommen des entsprechend der Postorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist gegeben durch*

$$\boxed{E_{n,j} = \frac{j}{2}\frac{\binom{2j}{j}\binom{2n+2-j}{n+1-j}}{\binom{2n}{n}} \quad \text{für } 1 \leq j \leq n .} \tag{3.142}$$

Daraus ergeben sich die asymptotischen Beziehungen

Korollar 3.7.2. *Der Erwartungswert $E_{n,j}$ der Anzahl der Nachkommen des entsprechend der Postorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für $j \sim \rho n$ und $0 < \rho < 1$ gegeben durch*

$$\boxed{E_{n,j} \sim 2\frac{\sqrt{\rho}}{\sqrt{\pi}\sqrt{1-\rho}}\sqrt{n} .} \tag{3.143}$$

Korollar 3.7.3. *Der Erwartungswert $E_{n,j}$ der Anzahl der Nachkommen des entsprechend der Postorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für fixes j gegeben durch*

$$\boxed{E_{n,j} \sim 2j\binom{2j}{j}4^{-j}\,.} \tag{3.144}$$

Korollar 3.7.4. *Der Erwartungswert $E_{n,j}$ der Anzahl der Nachkommen des entsprechend der Postorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für fixes $l = n - j$ gegeben durch*

$$\boxed{E_{n,j} \sim 2\binom{2(l+1)}{l+1}4^{-(l+1)}n\,.} \tag{3.145}$$

Weiters erhalten wir für $G_2(z,u)$

$$\begin{aligned} G_2(z,u) &= \frac{(1-Y)^2(1+3Y)}{Y^3(1+X)(X+Y)} \\ &= 3\frac{1}{X(X+Y)} + 2\frac{1}{X^2(X+Y)} - \frac{1}{X^3(X+Y)} - \frac{3}{4}\frac{X}{zY} + \frac{1}{2}\frac{X}{zY^2} + \frac{1}{4}\frac{X}{zY^3} \\ &\quad - 3\frac{1}{XY} - 2\frac{1}{X^2Y} + \frac{1}{X^3Y} + 2\frac{1}{XY^2} - \frac{1}{X^2Y^2} + \frac{1}{XY^3} + \frac{3}{4}\frac{1}{zY} - \frac{1}{2}\frac{1}{zY^2} - \frac{1}{4}\frac{1}{zY^3}\,. \end{aligned} \tag{3.146}$$

Kombination voriger Resultate ergibt für die Koeffizienten

$$\begin{aligned} [z^n u^j]G_2(z,u) &= \frac{nj}{2(n+1)}\binom{2j}{j}\binom{2n+2-2j}{n+1-j} + \frac{1}{2}4^j\binom{2n+2-2j}{n+1-j} - \frac{1}{2}\binom{2n+2}{n+1} \\ &\quad - 4^n + \sum_{k=j}^{n}\binom{2k}{k}\binom{2n-2k}{n-k} \quad \text{für } 1 \le j \le n\,. \end{aligned} \tag{3.147}$$

Dies ergibt

Hilfssatz 3.7.5. *Das zweite Faktorielle Moment $M_{n,j}^{(2)}$ der Anzahl der Nachkommen des entsprechend der Postorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist für $1 \le j \le n$ gegeben durch*

$$\boxed{\begin{aligned} M_{n,j}^{(2)} &= \frac{nj}{2}\frac{\binom{2j}{j}\binom{2n+2-2j}{n+1-j}}{\binom{2n}{n}} + \frac{n+1}{2}\frac{4^j\binom{2n+2-2j}{n+1-j}}{\binom{2n}{n}} - (n+1)\frac{4^n}{\binom{2n}{n}} - 2n - 1 \\ &\quad + \frac{n+1}{\binom{2n}{n}}\sum_{k=j}^{n}\binom{2k}{k}\binom{2n-2k}{n-k}\,. \end{aligned}} \tag{3.148}$$

Daraus erhält man unmittelbar

Satz 3.7.6. *Die Varianz $V_{n,j}$ der Anzahl der Nachkommen des entsprechend der Postorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist gegeben durch*

$$\boxed{\begin{aligned} V_{n,j} = {} & \frac{(n+1)j}{2}\frac{\binom{2j}{j}\binom{2n+2-2j}{n+1-j}}{\binom{2n}{n}} + \frac{n+1}{2}\frac{4^j\binom{2n+2-2j}{n+1-j}}{\binom{2n}{n}} - (n+1)\frac{4^n}{\binom{2n}{n}} - 2n - 1 \\ & - \frac{j^2}{4}\frac{\binom{2j}{j}^2\binom{2n+2-2j}{n+1-j}^2}{\binom{2n}{n}^2} + \frac{n+1}{\binom{2n}{n}}\sum_{k=j}^{n}\binom{2k}{k}\binom{2n-2k}{n-k} \quad \text{für } 1 \le j \le n\,. \end{aligned}} \tag{3.149}$$

Daraus erhält man die asymptotischen Beziehungen

Korollar 3.7.7. *Die Varianz $V_{n,j}$ der Anzahl der Nachkommen des entsprechend der Postorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für $j \sim \rho n$ und $0 < \rho < 1$ gegeben durch*

$$\boxed{V_{n,j} \sim \frac{2}{\sqrt{\pi}}\left(\frac{\sqrt{\rho}}{\sqrt{1-\rho}} + \arcsin\sqrt{1-\rho} - \frac{\pi}{2}\right) n^{3/2}\,.} \tag{3.150}$$

Korollar 3.7.8. *Die Varianz $V_{n,j}$ der Anzahl der Nachkommen des entsprechend der Postorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für fixes j gegeben durch*

$$\boxed{V_{n,j} \sim \frac{2j(j-1)}{3}\binom{2j}{j}4^{-j} + j - 4j^2\binom{2j}{j}^2 4^{-2j}\,.} \tag{3.151}$$

Korollar 3.7.9. *Die Varianz $V_{n,j}$ der Anzahl der Nachkommen des entsprechend der Postorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für fixes $l = n - j$ gegeben durch*

$$\boxed{V_{n,j} \sim \left(2\binom{2(l+1)}{l+1}4^{-(l+1)} - 4\binom{2(l+1)}{l+1}^2 4^{-2(l+1)}\right) n^2\,.} \tag{3.152}$$

3.8 Die Höhe der Knoten bei Postorder-Traversierung

Bezeichnet $f_{n,j,m}$ die Anzahl der Binärbäume mit n inneren Knoten, bei denen der entsprechend der Postorder-Traversierung markierte Knoten j genau die Höhe m hat.

Es ergibt sich dann für die gewöhnliche erzeugende Funktion $F(z,u,v)$ der $f_{n,j,m}$

$$F(z,u,v) = \sum_{j\ge1}\sum_{n\ge j}\sum_{m\ge0} f_{n,j,m}\, z^n u^j v^m\ .$$

die Funktionalgleichung

$$F(z,u,v) = zuvB^2(zu) + F(z,u,v)zvB(z) + B(zu)zvF(z,u,v)\,. \tag{3.153}$$

Daraus erhält man unmittelbar

$$F(z,u,v) = \frac{zuvB^2(zu)}{1 - zvB(z) - zvB(zu)}\,. \tag{3.154}$$

Wir berechnen wieder

$$\begin{aligned} G_1(z,u) &= \frac{4(1-Y)(1+Y)}{(1+X)^2(X+Y)^2} = \frac{u}{z}\frac{(1-X)^2}{(X+Y)^2} \\ &= \frac{2u}{z(X+Y)^2} - \frac{4u}{(X+Y)^2} - \frac{2u}{z}\frac{X}{(X+Y)^2}\,. \end{aligned} \tag{3.155}$$

Das Ablesen der Koeffizienten des bisher noch nicht behandelten Ausdrucks geschieht unter Verwendung von (3.72) und (3.118) leicht durch

$$\begin{aligned} [z^n u^j]\frac{X}{(X+Y)^2} &= [z^n u^j]\frac{1}{X+Y} - [z^n u^j]\frac{Y}{(X+Y)^2} \\ &= -\frac{n-1-3j}{2(n+1)(n+2)}\binom{2n}{n} \quad \text{für } 1 \le j \le n\,. \end{aligned} \tag{3.156}$$

Damit und unter Verwendung von (3.20) erhält man für $1 \le j \le n$

$$\begin{aligned} [z^n u^j]G_1(z,u) &= \frac{n+3-3j}{(n+2)(n+3)}\binom{2n+2}{n+1} \\ &\quad + \frac{j(n^2-jn+2n+j-1)}{2(n+1)(n+2)(n+3)}\binom{2j}{j}\binom{2(n+2-j)}{n+2-j}\,. \end{aligned} \tag{3.157}$$

und daraus sofort

Satz 3.8.1. *Der Erwartungswert $E_{n,j}$ der Höhe des entsprechend der Postorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist gegeben durch*

$$\boxed{\begin{aligned} E_{n,j} &= \frac{2(2n+1)(n+3-3j)}{(n+2)(n+3)} \\ &\quad + \frac{j(n^2-jn+2n+j-1)}{2(n+2)(n+3)}\frac{\binom{2j}{j}\binom{2(n+2-j)}{n+2-j}}{\binom{2n}{n}} \quad \text{für } 1 \le j \le n\,. \end{aligned}} \tag{3.158}$$

Daraus ergeben sich die asymptotischen Beziehungen

Korollar 3.8.2. *Der Erwartungswert $E_{n,j}$ der Höhe des entsprechend der Postorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für $j \sim \rho n$ und $0 < \rho < 1$ gegeben durch*

$$\boxed{E_{n,j} \sim 8\frac{\sqrt{\rho(1-\rho)}}{\sqrt{\pi}}\sqrt{n}\,.} \tag{3.159}$$

Korollar 3.8.3. *Der Erwartungswert $E_{n,j}$ der Höhe des entsprechend der Postorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für fixes j gegeben durch*

$$\boxed{E_{n,j} \sim 8j\binom{2j}{j}4^{-j} + 4\,.} \tag{3.160}$$

Korollar 3.8.4. *Der Erwartungswert $E_{n,j}$ der Höhe des entsprechend der Postorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für fixes $l = n - j$ gegeben durch*

$$\boxed{E_{n,j} \sim 2(l+3)\binom{2(l+2)}{l+2}4^{-(l+1)} - 8\,.} \tag{3.161}$$

Weiters betrachten wir

$$\begin{aligned}
G_2(z,u) &= 8\frac{(1-Y)(1+Y)(1-X)(X+Y+2)}{(1+X)^3(X+Y)^3} \\
&= \frac{u}{z^2}\frac{(1-X)^4}{(X+Y)^3} + \frac{u}{2z^2}\frac{(1-X)^4}{(X+Y)^2} \\
&= -\frac{8(1-2z)u}{z^2}\frac{X}{(X+Y)^3} + \frac{8(1-4z+2z^2)u}{z^2}\frac{1}{(X+Y)^3} \\
&\quad - \frac{4(1-2z)u}{z^2}\frac{X}{(X+Y)^2} + \frac{4(1-4z+2z^2}{z^2)}\frac{1}{(X+Y)^2}\,.
\end{aligned} \tag{3.162}$$

Das Ablesen der Koeffizienten des noch fehlenden Ausdrucks ist mit den schon errechneten Beziehungen (3.20) und (3.131) leicht möglich. Es gilt nämlich für $1 \le j \le n$

$$\begin{aligned}
[z^n u^j]\frac{X}{(X+Y)^3} &= [z^n u^j]\frac{1}{(X+Y)^2} - [z^n u^j]\frac{Y}{(X+Y)^3} \\
&= -\frac{1}{4}\frac{(2j+1)(2n-2j+1)(n-4j-3)}{(n+1)(n+2)(n+3)}\binom{2j}{j}\binom{2(n-j)}{n-j}\,.
\end{aligned} \tag{3.163}$$

Unter Benützung von (3.20), (3.30), (3.156) und (3.163) erhält man dann für $1 \le j \le n$

$$\begin{aligned}
[z^n u^j]G_2(z,u) &= \Big[j(5n^5 - 26jn^4 + 47n^4 + 37j^2n^3 - 163jn^3 + 143n^3 - 16j^3n^2 \\
&\quad + 116j^2n^2 - 250jn^2 + 129n^2 - 13j^2n + 31jn - 40n + 16j^3 - 68j^2 \\
&\quad + 72j - 20)\Big] \times \\
&\quad \times \frac{1}{(n+3-j)(n+1)(n+2)(n+3)(n+4)(n+5)}\binom{2j}{j}\binom{2(n+2-j)}{n+2-j} \\
&\quad + \Big[4(2n+1)(3jn^3 + 2n^3 - 3j^2n^2 - 9jn^2 + 18n^2 + 33j^2n - 84jn \\
&\quad + 40n + 30j^2 - 30j)\Big]\frac{1}{(n+1)(n+2)(n+3)(n+4)(n+5)}\binom{2n}{n}\,.
\end{aligned} \tag{3.164}$$

Damit erhält man

Hilfssatz 3.8.5. *Das zweite Faktorielle Moment $M_{n,j}^{(2)}$ der Höhe des entsprechend der Postorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist für $1 \le j \le n$ gegeben durch*

$$\boxed{\begin{aligned} M_{n,j}^{(2)} = \Big[&4(2n+1)(3jn^3 + 2n^3 - 3j^2n^2 - 9jn^2 + 18n^2 + 33j^2n - 84jn \\ &+ 40n + 30j^2 - 30j)\Big] \frac{1}{(n+2)(n+3)(n+4)(n+5)} \\ + \Big[&j(5n^5 - 26jn^4 + 47n^4 + 37j^2n^3 - 163jn^3 + 143n^3 - 16j^3n^2 \\ &+ 116j^2n^2 - 250jn^2 + 129n^2 - 13j^2n + 31jn - 40n + 16j^3 - 68j^2 \\ &+ 72j - 20)\Big] \frac{1}{(n+3-j)(n+2)(n+3)(n+4)(n+5)} \frac{\binom{2j}{j}\binom{2(n+2-j)}{n+2-j}}{\binom{2n}{n}}. \end{aligned}} \tag{3.165}$$

Als unmittelbare Folgerung ergibt sich daraus

Satz 3.8.6. *Die Varianz $V_{n,j}^{(2)}$ der Höhe des entsprechend der Postorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist gegeben durch*

$$\boxed{\begin{aligned} V_{n,j} = \Big[&2n(2n+1)(6jn^4 + n^4 + 11n^3 + 33jn^3 - 6j^2n^3 + 36jn^2 + 35n^2 + 12j^2n \\ &+ 51jn + 13n - 186j^2 + 234j - 60)\Big] \frac{1}{(n+2)^2(n+3)^2(n+4)(n+5)} \\ + \Big[&j(3n^7 - 14jn^6 + 23n^6 + 19j^2n^5 - 38jn^5 - 2n^5 - 8j^3n^4 - 29j^2n^4 \\ &+ 428jn^4 - 436n^4 + 44j^3n^3 - 681j^2n^3 + 1950jn^3 - 1271n^3 \\ &+ 200j^3n^2 - 955j^2n^2 + 1794jn^2 - 1003n^2 - 188j^3n + 950j^2n \\ &- 1048jn + 262n - 48j^3 + 264j^2 - 336j + 120)\Big] \times \\ \times &\frac{1}{2(n+3-j)(n+2)^2(n+3)^2(n+4)(n+5)} \frac{\binom{2j}{j}\binom{2(n+2-j)}{n+2-j}}{\binom{2n}{n}} \\ - &\frac{j^2(n^2 - jn + 2n + j - 1)^2}{4} \frac{\binom{2j}{j}^2\binom{2(n+2-j)}{n+2-j}^2}{\binom{2n}{n}^2} \quad \text{für } 1 \le j \le n. \end{aligned}} \tag{3.166}$$

Dies führt zu den asymptotischen Beziehungen

Korollar 3.8.7. *Die Varianz $V_{n,j}$ der Höhe des entsprechend der Postorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für $j \sim \rho n$ und $0 < \rho < 1$ gegeben durch*

$$\boxed{V_{n,j} \sim 8\frac{\rho(1-\rho)(3\pi - 8)}{\pi} n.} \tag{3.167}$$

Korollar 3.8.8. *Die Varianz $V_{n,j}$ der Höhe des entsprechend der Postorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für fixes j gegeben durch*

$$\boxed{V_{n,j} \sim 24j + 4 + 24j\binom{2j}{j}4^{-j} - 64j^2\binom{2j}{j}^2 4^{-2j}\,.} \tag{3.168}$$

Korollar 3.8.9. *Die Varianz $V_{n,j}$ der Höhe des entsprechend der Postorder-Traversierung mit Marke j versehenen Knotens in einem Binärbaum der Größe n ist asymptotisch für fixes $l = n - j$ gegeben durch*

$$\boxed{V_{n,j} \sim 24l + 136 - 4(l+3)^2\binom{2(l+2)}{l+2}^2 4^{-2(l+1)} - \frac{2(l+5)(5l+11)}{l+3}\binom{2(l+2)}{l+2}4^{-(l+1)}\,.} \tag{3.169}$$

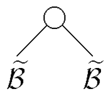
3.9 Binärbäume ohne Unterscheidung von internen und externen Knoten

3.9.1 Einleitung

In manchen Anwendungen erweist es sich als zweckmäßig, wenn man in einem Binärbaum keine Unterscheidung in interne und externe Knoten durchführt. Traversiert man nun solch einen Baum, ist es wiederum interessant, Fragen nach der Höhe bzw. der Anzahl der Nachfolger des Knotens, den man in der Traversierung als j-tes besuchte, zu behandeln.

Wir bestimmen im folgenden die Erwartungswerte der Höhe bzw. der Anzahl der Nachfolger von entsprechend der Reihenfolge der Inorder-Traversierung markierten Knoten in Binärbäumen, wobei nicht in interne und externe Knoten unterschieden wird.

Hier bezeichnen wir mit $\widetilde{\mathcal{B}}$ die Familie der Binärbäume, wobei keine Unterscheidung in interne und externe Knoten durchgeführt wird. Sei $\bigcirc$ das Symbol für einen Knoten, dann ist die Familie $\widetilde{\mathcal{B}}$ durch folgende formale Gleichung gegeben:

$$\widetilde{\mathcal{B}} \quad = \quad \bigcirc \quad + $$


$$. \tag{3.170}$$

Mithilfe der symbolischen Operatormethode übersetzt sich diese formale Gleichung mittels

$$\widetilde{B}(z) = z + z\widetilde{B}^2(z) \tag{3.171}$$

direkt in eine Gleichung für die erzeugende Funktion $\widetilde{B}(z)$ der $\widetilde{B}_n$, wobei $\widetilde{B}_n$ die Anzahl der Binärbäume mit insgesamt n Knoten bezeichnet.

Als Lösung dieser erzeugenden Funktion ergibt sich dann

$$\widetilde{B}(z) = \frac{1 - \sqrt{1 - 4z^2}}{2z}, \tag{3.172}$$

woraus man für die $\widetilde{B}_n$ folgendes erhält:

$$\widetilde{B}_n = \begin{cases} \frac{1}{k+1}\binom{2k}{k} & \text{für } n = 2k+1 \text{ mit } k \in \mathbb{N}, \\ 0 & \text{sonst}. \end{cases} \tag{3.173}$$

Weiters werden wir die Abkürzungen

$$\widetilde{X} = \sqrt{1-4z^2}, \qquad \widetilde{Y} = \sqrt{1-4u^2z^2} \tag{3.174}$$

verwenden.

3.9.2 Die Anzahl der Nachfolger bei Inorder-Traversierung

Bezeichnen wir mit $f_{n,j,m}$ die Anzahl der Binärbäume mit n Knoten, bei denen der entsprechend der Reihenfolge der Inorder-Traversierung markierte Knoten j genau m Nachfolger besitzt. Für die durch

$$F(z,u,v) = \sum_{j \geq 1} \sum_{n \geq j} \sum_{m \geq 1} f_{n,j,m}\, z^n u^j v^m \tag{3.175}$$

definierte erzeugende Funktion $F(z,u,v)$ erhalten wir aus der formalen Gleichung (3.170) die Funktionalgleichung

$$F(z,u,v) = zuv + \widetilde{B}(zuv)zuv\widetilde{B}(zv) + F(z,u,v)z\widetilde{B}(z) + \widetilde{B}(zu)zuF(z,u,v). \tag{3.176}$$

Daraus ergibt sich sofort

$$F(z,u,v) = \frac{zuv + \widetilde{B}(zuv)zuv\widetilde{B}(zv)}{1 - z\widetilde{B}(z) - zu\widetilde{B}(zu)}. \tag{3.177}$$

Für die Bestimmung der Erwartungswerte $E_{n,j}$ definieren wir mittels

$$G(z,u) = \frac{\partial}{\partial v} F(z,u,v)\bigg|_{v=1} \tag{3.178}$$

eine Funktion $G(z,u)$.

Der Erwartungswert $E_{n,j}$ der Anzahl der Nachfolger des entsprechend der Reihenfolge der Inorder-Traversierung markierten Knotens j in einem Binärbaum mit insgesamt n Knoten ist dann für $n = 2k+1$, $k \in \mathbb{N}$ gegeben durch

$$E_{n,j} = \frac{1}{\widetilde{B}_n}[z^n u^j]G(z,u). \tag{3.179}$$

Differenzieren wir (3.177) nach v und werten anschließend bei $v = 1$ aus, erhalten wir unter Verwendung der Abkürzungen (3.174)

$$\begin{aligned} G(z,u) = {} & \frac{2zu}{\widetilde{X}+\widetilde{Y}} - \frac{1}{2z(\widetilde{X}+\widetilde{Y})} + \frac{1}{2z\widetilde{X}\widetilde{Y}} + \frac{1}{2z\widetilde{X}^2(\widetilde{X}+\widetilde{Y})} - \frac{1}{2z\widetilde{X}^2\widetilde{Y}} \\ & - \frac{8z^3u^2}{\widetilde{X}^2(\widetilde{X}+\widetilde{Y})} + \frac{8z^3u^2}{\widetilde{X}^2\widetilde{Y}}. \end{aligned} \tag{3.180}$$

Für das Ablesen der Koeffizienten können wir schon in diesem Kapitel hergeleitete Formeln verwenden, wenn man den Zusammenhang

$$[z^{2k}u^{2i}]f(\widetilde{X},\widetilde{Y}) = [z^k u^i]f(X,Y) \tag{3.181}$$

mit den durch (3.6) definierten abkürzenden Schreibweisen X bzw. Y beachtet.

Man erhält unter Verwendung der Beziehungen (3.72), (3.55) und (3.104) unter Berücksichtigung von (3.179) aus (3.180)

Satz 3.9.1. *Der Erwartungswert $E_{n,j}$ der Anzahl der Nachfolger des entsprechend der Reihenfolge der Inorder-Traversierung markierten Knotens j in einem Binärbaum mit insgesamt n Knoten ist gegeben durch*

$$\boxed{E_{n,j} = \begin{cases} 1 & \text{für } n = 2k+1 \text{ und } j = 2i+1 \text{ mit } 0 \le i \le k\,,\ k,i \in \mathbb{N}\,, \\ 1 + \frac{k+1}{2}\frac{\binom{2i}{i}\binom{2(k+1-i)}{k+1-i}}{\binom{2k}{k}} & \text{für } n = 2k+1 \ \text{ und } j = 2i \text{ mit } 1 \le i \le k\,,\ k,i \in \mathbb{N}\,, \\ 0 & \text{sonst}\,. \end{cases}} \tag{3.182}$$

3.9.3 Die Höhe der Knoten bei Inorder-Traversierung

Seien hier $f_{n,j,m}$ die Anzahl der Binärbäume mit n Knoten, bei denen der entsprechend der Reihenfolge der Inorder-Traversierung markierte Knoten j die Höhe m hat.

Für die durch (3.175) definierte erzeugende Funktion $F(z,u,v)$ erhalten wir mittels der symbolischen Operatormethode aus der formalen Gleichung (3.170) die Funktionalgleichung

$$F(z,u,v) = zuv + \widetilde{B}(zu)zuv\widetilde{B}(z) + F(z,u,v)zv\widetilde{B}(z) + \widetilde{B}(zu)zuvF(z,u,v)\,, \tag{3.183}$$

woraus sich explizit

$$F(z,u,v) = \frac{zuv + \widetilde{B}(zu)zuv\widetilde{B}(z)}{1 - zv\widetilde{B}(z) - zuv\widetilde{B}(zu)} \tag{3.184}$$

ergibt.

Ableiten von (3.184) nach v und Auswerten bei $v = 1$ ergibt unter Verwendung von (3.174) für die durch (3.178) definierte Funktion $G(z,u)$:

$$G(z,u) = \frac{4zu}{(\widetilde{X}+\widetilde{Y})^2} - \frac{1}{z(\widetilde{X}+\widetilde{Y})} + \frac{1}{z(\widetilde{X}+\widetilde{Y})^2} + \frac{\widetilde{X}\widetilde{Y}}{z(\widetilde{X}+\widetilde{Y})^2}\,. \tag{3.185}$$

Man erhält unter Verwendung der Beziehungen (3.72), (3.20) und (3.74) unter Berücksichtigung von (3.179) aus (3.185)

Satz 3.9.2. *Der Erwartungswert $E_{n,j}$ der Höhe des entsprechend der Reihenfolge der Inorder-Traversierung markierten Knotens j in einem Binärbaum mit insgesamt n Knoten ist* *gegeben durch*

$$\boxed{E_{n,j} = \begin{cases} \frac{2(2i+1)(2k-2i+1)}{k+2} \frac{\binom{2i}{i}\binom{2(k-i)}{k-i}}{\binom{2k}{k}} & \text{für } n = 2k+1 \text{ und } j = 2i+1 \\ & \text{mit } 0 \leq i \leq k\,,\ k, i \in \mathbb{N}\,, \\ \frac{k+4ik-4i^2+4i+1}{2(k+2)} \frac{\binom{2i}{i}\binom{2(k+1-i)}{k+1-i}}{\binom{2k}{k}} - \frac{2k+1}{k+2} & \text{für } n = 2k+1 \text{ und } j = 2i \\ & \text{mit } 1 \leq i \leq k\,,\ k, i \in \mathbb{N}\,, \\ 0 & \text{sonst}\,. \end{cases}} \tag{3.186}$$

			j fest	$j \sim \rho n$	$l := n - j$ fest
	Höhe der Blätter	$E_{n,j}$	$(2j+1)\binom{2j}{j}4^{-j+1}$	$8\frac{\sqrt{\rho(1-\rho)}}{\sqrt{\pi}}\sqrt{n}$	$(2l+1)\binom{2l}{l}4^{-l+1}$
		$V_{n,j}$	$24(j+1)$ $-16(2j+1)^2\binom{2j}{j}^2 4^{-2j}$ $-(2j+1)\binom{2j}{j}4^{-j+1}$	$8\frac{\rho(1-\rho)(3\pi-8)}{\pi}n$	$24(l+1)$ $-16(2l+1)^2\binom{2l}{l}^2 4^{-2l}$ $-(2l+1)\binom{2l}{l}4^{-l+1}$
Inorder	Nachfolger	$E_{n,j}$	$\binom{2j}{j}4^{-j}n$	$\frac{1}{\sqrt{\pi}\sqrt{\rho(1-\rho)}}\sqrt{n}$	$\binom{2(l+1)}{l+1}4^{-(l+1)}n$
		$V_{n,j}$	$\left(\binom{2j}{j}4^{-j}-\binom{2j}{j}^2 4^{-2j}\right)n^2$	$\left(\frac{1}{\sqrt{\pi}\sqrt{\rho(1-\rho)}}-\frac{\sqrt{\pi}}{2}\right)n^{3/2}$	$\Big(\binom{2(l+1)}{l+1}4^{-(l+1)}$ $-\binom{2(l+1)}{l+1}^2 4^{-2(l+1)}\Big)n^2$
	Höhe	$E_{n,j}$	$2(4j+1)\binom{2j}{j}4^{-j}-2$	$8\frac{\sqrt{\rho(1-\rho)}}{\sqrt{\pi}}\sqrt{n}$	$2(4(l+1)+1)\binom{2(l+1)}{l+1}4^{-(l+1)}-2$
		$V_{n,j}$	$24j+14$ $-2(4j+5)\binom{2j}{j}4^{-j}$ $-4(4j+1)^2\binom{2j}{j}^2 4^{-2j}$	$8\frac{\rho(1-\rho)(3\pi-8)}{\pi}n$	$24(l+1)+14$ $-2(4(l+1)+5)\binom{2(l+1)}{l+1}4^{-(l+1)}$ $-4(4(l+1)+1)^2\binom{2(l+1)}{l+1}^2 4^{-2(l+1)}$
Preorder	Nachfolger	$E_{n,j}$	$2\binom{2j}{j}4^{-j}n$	$2\frac{\sqrt{1-\rho}}{\sqrt{\pi}\sqrt{\rho}}\sqrt{n}$	$(2l+1)\binom{2l}{l}4^{-l}$
		$V_{n,j}$	$\Big(2\binom{2j}{j}4^{-j}$ $-\binom{2j}{j}^2 4^{-2j+1}\Big)n^2$	$\frac{2}{\sqrt{\pi}}\Big(\frac{\sqrt{1-\rho}}{\sqrt{\rho}}$ $-\arcsin\sqrt{1-\rho}\Big)n^{3/2}$	$\frac{2}{3}l(l+1)\binom{2(l+1)}{l+1}4^{-(l+1)}+l+1$ $-4(l+1)^2\binom{2(l+1)}{l+1}^2 4^{-2(l+1)}$
	Höhe	$E_{n,j}$	$2(j+2)\binom{2(j+1)}{j+1}4^{-j}-8$	$8\frac{\sqrt{\rho(1-\rho)}}{\sqrt{\pi}}\sqrt{n}$	$4(2l+1)\binom{2l}{l}4^{-l}+4$
		$V_{n,j}$	$24j+112$ $-4(j+2)^2\binom{2(j+1)}{j+1}^2 4^{-2j}$ $-\frac{2(j+4)(5j+6)}{j+2}\binom{2(j+1)}{j+1}4^{-j}$	$8\frac{\rho(1-\rho)(3\pi-8)}{\pi}n$	$24l+28$ $-16(2l+1)^2\binom{2l}{l}^2 4^{-2l}$ $+12(2l+1)\binom{2l}{l}4^{-l}$
Postorder	Nachfolger	$E_{n,j}$	$2j\binom{2j}{j}4^{-j}$	$2\frac{\sqrt{\rho}}{\sqrt{\pi}\sqrt{1-\rho}}\sqrt{n}$	$2\binom{2(l+1)}{l+1}4^{-(l+1)}n$
		$V_{n,j}$	$\frac{2j(j-1)}{3}\binom{2j}{j}4^{-j}+j$ $-4j^2\binom{2j}{j}^2 4^{-2j}$	$\frac{2}{\sqrt{\pi}}\Big(\frac{\sqrt{\rho}}{\sqrt{1-\rho}}$ $+\arcsin\sqrt{1-\rho}-\frac{\pi}{2}\Big)n^{3/2}$	$\Big(2\binom{2(l+1)}{l+1}4^{-(l+1)}$ $-4\binom{2(l+1)}{l+1}^2 4^{-2(l+1)}\Big)n^2$
	Höhe	$E_{n,j}$	$8j\binom{2j}{j}4^{-j}+4$	$8\frac{\sqrt{\rho(1-\rho)}}{\sqrt{\pi}}\sqrt{n}$	$2(l+3)\binom{2(l+2)}{l+2}4^{-(l+1)}-8$
		$V_{n,j}$	$24j+4$ $+24j\binom{2j}{j}4^{-j}$ $-64j^2\binom{2j}{j}^2 4^{-2j}$	$8\frac{\rho(1-\rho)(3\pi-8)}{\pi}n$	$24l+136$ $-4(l+3)^2\binom{2(l+2)}{l+2}^2 4^{-2(l+1)}$ $-\frac{2(l+5)(5l+11)}{l+3}\binom{2(l+2)}{l+2}4^{-(l+1)}$

Tabelle 3.1: Asymptotische Resultate zu den expliziten Formeln dieses Kapitels

Kapitel 4

Über die Knotengrade in Binärbäumen

4.1 Einleitung

In den letzten Jahren sind eine Reihe von Artikeln publiziert worden, die sich mit der Verteilung der Knoten mit einer bestimmten Anzahl an Kindern in verschiedenen Baumfamilien beschäftigen. So zeigte 1986 Mahmoud in [27], daß die Erwartungswerte der Anzahl der Knoten mit 0, 1 bzw. 2 Kinder. in Binären Suchbäumen jeweils asymptotisch $\frac{n}{3}$ sind. Durch Anwendung von Grenzwertsätzen für abhängige Zufallsvariable konnte Devroye 1991 in [5] die asymptotischen Verteilungen der Anzahl der Knoten mit 0, 1 bzw. 2 Kindern in Binären Suchbäumen bestimmen.

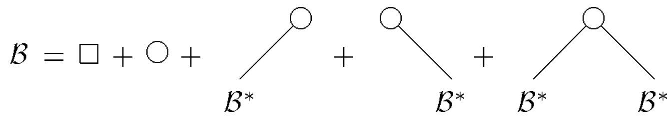
Für die gewöhnlichen Binärbäume bestimmte Mahmoud 1995 in [28] schließlich die exakten als auch die asymptotischen Verteilungen der Anzahl der Knoten der 3 verschiedenen Knotenarten. Obwohl in einer probabilistischen Sichtweise damit auch für die Binärbäume dieses Problem zufriedenstellend gelöst wurde, läßt sich von einem kombinatorischen Standpunkt aus betrachtet, die Fragestellung noch weiter verfolgen. Es zeigt sich nämlich, daß sich die angegebenen Lösungen der exakten Verteilungen noch weiter vereinfachen lassen. So hat Prodinger 1996 in [38] die auftretenden Summen in Mahmoud's Ergebnis mit Zeilberger's Algorithmus in eine geschlossene Form gebracht. Direkte kombinatorische Beweise für die Anzahl der Binärbäume, die eine bestimmte Anzahl an Knoten mit 0, 1 bzw. 2 Kindern besitzen lieferte 1997 Rote in [41]. Weiters brachte er kombinatorische Beweise für die entsprechenden Verallgemeinerungen für d-äre Bäume bzw. Ebene Wurzelbäume.

Eine andere Möglichkeit die Verteilung der Knotengrade in Binärbäumen zu studieren besteht darin, die entsprechend der Reihenfolge von bestimmten Traversierungen (siehe Kapitel 3) numerierten Knoten zu betrachten. So ermittelte Sprugnoli 1990 in [45] unter anderem exakte Formeln für die Anzahl der Binärbäume, bei denen ein entsprechend der Reihenfolge der Inorder-Traversierung markierter Knoten 0, 1 bzw. 2 Kinder besitzt (wobei bei einem Kind auch unterschieden wird, ob es vom Vaterknoten durch eine nach links führende oder eine nach rechts führende Kante erreicht wird).

In diesem Kapitel werden wir die durch unterschiedliche Methoden gewonnenen Ergebnisse, betreffend die exakte Verteilung der Anzahl der Knoten mit 0, 1 bzw. 2 Kindern in

Binärbäumen, durch ein einheitliches Verfahren wiedergewinnen und noch erweitern. So werden wir nämlich nicht nur entsprechend der Reihenfolge der Inorder-Traversierung markierte Knoten betrachten, sondern die Fragestellung auch auf Preorder- und Postorder-traversierte Bäume ausdehnen.

Wir verwenden zur Gewinnung der Resultate wieder die symbolische Operatormethode, starten diesmal aber von einer formalen Gleichung für die Familie der Binärbäume $\mathcal{B}$, welche die Unterscheidung in Knoten mit 0, 1 bzw. 2 Kindern berücksichtigt. Sie lautet



(4.1)

mit $\mathcal{B}^* = \mathcal{B} - \square$.

Diese formale Gleichung übersetzen wir in eine Funktionalgleichung der erzeugenden Funktion der gerade betrachteten Parameter. Lösen der Funktionalgleichung und Koeffizientenablesen liefert dann schon das gewünschte Ergebnis.

Weiters verwenden wir in diesem Kapitel wieder die Bezeichnung B_n für die Anzahl der Binärbäume der Größe n. Die gewöhnliche erzeugende Funktion $B(z) = \sum_{n\geq 0} B_n z^n$ ist dann gegeben durch

$$B(z) = \frac{1-\sqrt{1-4z}}{2z},$$

während $B^*(z)$ durch $B^*(z) = B(z) - 1$ gegeben ist.

Auch sollen wieder die Abkürzungen

$$X = \sqrt{1-4z} \qquad \text{und} \qquad Y = \sqrt{1-4uz}$$

Verwendung finden.

4.2 Die Anzahl der Binärbäume mit einer vorgegebenen Zahl an Knoten mit 0, 1 und 2 Kindern

Hier bezeichnen wir hier mit $F_{n,m}$ die Anzahl der Binärbäume der Größe n, die genau m Knoten mit 2 Kindern besitzen. Sei weiters k die Anzahl der Blätter des Baumes und l die Anzahl der Knoten des Baums mit genau einem Kind, dann gelten die Beziehungen

$$l = n - 1 - 2m \text{ und } k = m + 1. \tag{4.2}$$

(vergleiche mit dem Beweis von Satz 1.8.7). Es gilt dann

Satz 4.2.1. *Die Anzahl der Binärbäume $F_{n,m}$ der Größe n, die genau m Knoten mit 2 Kindern besitzen, ist für $n \geq 1$ gegeben durch*

$$F_{n,m} = \frac{1}{m+1}\binom{2m}{m}\binom{n-1}{n-2m-1}2^{n-2m-1} = 2^l\binom{2m+l}{l}B_m = \frac{1}{n}\binom{n}{k\; l\; m}2^l. \tag{4.3}$$

Beweis. Wir wollen die formale Gleichung (4.1) direkt in eine Funktionalgleichung für die erzeugende Funktion

$$F(z,v) = \sum_{n\geq 0}\sum_{m\geq 0} F_{n,m}\, z^n v^m$$

der $F_{n,m}$ übersetzen. Dazu definieren wir einen Operator $\widetilde{F}$, der jedem Objekt B aus $\mathcal{B}$ einen Term in z und v durch folgende Vorschrift

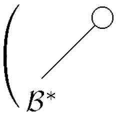
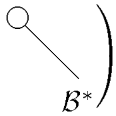
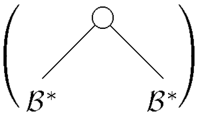
$$\widetilde{F}(B) := z^{w(B)} v^{s(B)}$$

zuweist. Dabei soll $w(B)$ die Größe des Baumes B und $s(B)$ die Anzahl der Knoten mit 2 Kindern in B bezeichnen. Für eine Familie von Bäumen $(B_i)_{i\in I}$, $B_i \in \mathcal{B}$ soll der Operator $\widetilde{F}$ definiert sein durch

$$\widetilde{F}\left((B_i)_{i\in I}\right) := \sum_{i\in I} \widetilde{F}(B_i)\,,$$

sofern gewährleistet ist, daß für jedes n nur endlich viele $B_i \in \mathcal{B}_n$ sind. Dann weist $\widetilde{F}$ jeder Familie $(B_i)_{i\in I}$ eine formale Potenzreihe in z und v zu.

Wenden wir nun $\widetilde{F}$ auf die Gleichung (4.1) an, so bekommen wir aufgrund der Additivität dieser Funktion zunächst

$\widetilde{F}(\mathcal{B}) = \widetilde{F}(\square) + \widetilde{F}(\bigcirc) + \widetilde{F}\Big($  $\Big) + \widetilde{F}\Big($  $\Big) + \widetilde{F}\Big($  $\Big)\,.$ (4.4)

Wegen $\widetilde{F}(\mathcal{B}) = F(z,v)$, $\widetilde{F}(\mathcal{B}^*) = F(z,v) - 1$, $\widetilde{F}(\square) = 1$, $\widetilde{F}(\bigcirc) = z$ und den leicht einzusehenden Faltungen erhalten wir daraus die Funktionalgleichung

$$F(z,v) = 1 + z + 2z(F(z,v)-1) + zv(F(z,v)-1)^2\,. \tag{4.5}$$

Definieren wir $G(z,v) := F(z,v) - 1$, dann schreibt sich diese Gleichung als

$$G(z,v) = z + 2zG(z,v) + zvG(z,v)^2\,, \tag{4.6}$$

woraus sich die Lösung

$$G(z,v) = \frac{1-2z-\sqrt{(1-2z)^2-4z^2v}}{2zv} = \frac{1-2z-(1-2z)\sqrt{1-\left(\frac{2z}{1-2z}\right)^2 v}}{2zv} \tag{4.7}$$

ergibt.

Daraus erhält man für $n \geq 1$ zunächst

$$F_{n,m} = [z^n u^m]G(z,v) = -\frac{1}{2}[z^{n+1}v^{m+1}](1-2z)\sqrt{1-\left(\frac{2z}{1-2z}\right)^2 v}\;. \tag{4.8}$$

Mit der Beziehung

$$[z^n]\sqrt{1-z} = -\frac{2}{4^n}B_{n-1} = -\frac{2}{n\,4^n}\binom{2(n-1)}{n-1} \tag{4.9}$$

ergibt sich weiter

$$F_{n,m} = [z^{n+1}](1-2z)\left(\frac{2z}{1-2z}\right)^{2(m+1)} \frac{1}{(m+1)4^{m+1}}\binom{2m}{m}$$
$$= \frac{1}{(m+1)4^{m+1}}\binom{2m}{m}[z^{n+1}]\frac{(2z)^{2(m+1)}}{(1-2z)^{2m+1}} = \frac{1}{m+1}\binom{2m}{m}[z^{n-2m-1}]\frac{1}{(1-2z)^{2m+1}}\,. \tag{4.10}$$

Berücksichtigt man noch

$$\frac{1}{(1-z)^p} = \sum_{k\geq 0}\binom{p+k-1}{k}z^k\,,$$

dann ergibt sich schließlich

$$F_{n,m} = \frac{1}{m+1}\binom{2m}{m}\binom{n-1}{n-2m-1}2^{n-2m-1} \quad \text{für } n \geq 1\,. \tag{4.11}$$

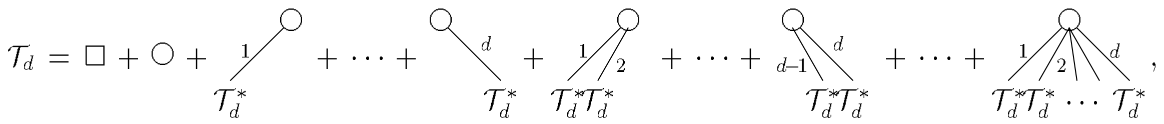
Die beiden anderen Formeln für $F_{n,m}$, die im Satz angegeben wurden, sind unmittelbare Folgerungen aus (4.2). $\square$

Dabei bezeichnet allgemein $\binom{n}{n_0\ n_1\ \cdots\ n_d}$ mit $n_0+n_1+\cdots+n_d = n$ den Multinomialkoeffizienten, ist also definiert durch

$$\binom{n}{n_0\ n_1\ \cdots\ n_d} = \frac{n!}{n_0!\cdot n_1!\cdots n_d!}\,.$$

4.3 Erweiterung der Betrachtungen auf d-äre Bäume

Nun betrachten wir anstelle der Binären Bäume allgemein d-äre Bäume, also Bäume mit höchstens d Kindern, bei denen die wegführenden Kanten mit paarweise verschiedenen Marken aus der Menge $\{1,\ldots,d\}$ bezeichnet sind. Sei also $\mathcal{T}_d$ die Familie der d-ären Bäume, $d \in \mathbb{N}$, dann sind diese beschrieben durch die formale Gleichung



mit $\mathcal{T}_d^* = \mathcal{T}_d - \square\,.$

(4.12)

Bezeichnen wir nun mit $F_{n,n_0,n_1,\ldots,n_d}$ die Anzahl der d-ären Bäume der Größe n, die genau n_i Knoten mit i Kindern besitzen (für $0 \leq i \leq d$). Es gilt dann

Satz 4.3.1. *Die Anzahl der d-ären Bäume $F_{n,n_0,n_1,\ldots,n_d}$ der Größe n, die für $0 \leq i \leq d$ genau n_i Knoten mit i Kindern besitzen, ist für $n \geq 1$ gegeben durch*

$$F_{n,n_0,n_1,\ldots,n_d} = \frac{1}{n}\cdot\binom{n}{n_0\ n_1\ \cdots\ n_d}\cdot\binom{d}{0}^{n_0}\binom{d}{1}^{n_1}\binom{d}{2}^{n_2}\cdots\binom{d}{d}^{n_d}\,,$$
$$\text{wenn } \sum_{i=0}^{d} n_i = n \text{ und } \sum_{i=0}^{d} i\,n_i = n-1 \text{ erfüllt sind und } F_{n,n_0,n_1,\ldots,n_d} = 0 \text{ sonst}\,. \tag{4.13}$$

Beweis. Bezeichnen wir die erzeugende Funktion der $F_{n,n_0,n_1,\dots,n_d}$ mit $F(z,v_0,v_1,\dots,v_d)$, ist diese also gegeben durch

$$F(z,v_0,v_1,\dots,v_d)=\sum_{n\geq 0}\sum_{n_0\geq 0}\cdots\sum_{n_d\geq 0}F_{n,n_0,n_1,\dots,n_d}z^n v_0^{n_0}v_1^{n_1}\cdots v_d^{n_d}\,,$$

dann übersetzt sich die formale Gleichung (4.12) in folgende Funktionalgleichung der $F(z,v_0,v_1,\dots,v_d)$:

$$\begin{aligned}F\left(z,v_0,v_1,\dots,v_d\right)&=1+\binom{d}{0}zv_0+\binom{d}{1}zv_1\Big(F\left(z,v_0,v_1,\dots,v_d\right)-1\Big)\\&\quad+\binom{d}{2}zv_2\Big(F\left(z,v_0,v_1,\dots,v_d\right)-1\Big)^2+\cdots+\binom{d}{d}zv_d\Big(F\left(z,v_0,v_1,\dots,v_d\right)-1\Big)^d\\&=1+z\sum_{i=0}^{d}\binom{d}{i}v_i\Big(F\left(z,v_0,v_1,\dots,v_d\right)-1\Big)^i.\end{aligned}\tag{4.14}$$

Definieren wir uns nun die Funktion $G(z,v_0,v_1,\dots,v_d)$ durch

$$G(z,v_0,v_1,\dots,v_d)=F(z,v_0,v_1,\dots,v_d)-1\,,$$

dann schreibt sich Gleichung (4.14) als

$$\begin{aligned}G(z,v_0,\dots,v_d)&=z\left[\binom{d}{0}v_0+\binom{d}{1}v_1G(z,v_0,\dots,v_d)^1+\cdots+\binom{d}{d}v_dG(z,v_0,\dots,v_d)^d\right]\\&=z\sum_{i=0}^{d}\binom{d}{i}v_iG\left(z,v_0,\dots,v_d\right)^i\end{aligned}\tag{4.15}$$

bzw. als

$$z=\frac{G(z,v_0,\dots,v_d)}{\binom{d}{0}v_0+\binom{d}{1}v_1G(z,v_0,\dots,v_d)^1+\cdots+\binom{d}{d}v_dG(z,v_0,\dots,v_d)^d}\,.\tag{4.16}$$

Um nun zu den gesuchten $F_{n,n_0,\dots,n_d}$ zu gelangen, können wir für $n\geq 1$ die Beziehung

$$F_{n,n_0,\dots,n_d}=[z^nv_0^{n_0}\cdots v_d^{n_d}]G(z,v_0,\dots,v_d)$$

verwenden.

Allerdings kann für $d\geq 5$ die Funktion $G(z,v_0,\dots,v_d)$ nicht explizit aus Gleichung (4.16) erhalten werden. Stattdessen verwenden wir die Lagrange'sche Inversionsformel (siehe beispielsweise [4]):

Satz 4.3.2 (Inversionsformel von Lagrange). *Sei $f(z)$ eine formale Potenzreihe in z und sei weiters $z=\frac{x}{\varphi(x)}$, wobei $\varphi(x)=\sum_{i\geq 0}\varphi_i\,x^i$ eine formale Potenzreihe in x ist mit $\varphi_0\neq 0$. Dann gilt*

$$[z^n]f(x)=\frac{1}{n}[x^{n-1}]f'(x)\,\varphi(x)^n\quad\text{für } n\geq 1\,.\quad\Box\tag{4.17}$$

Fassen wir nämlich G als Potenzreihe in z auf, dann erhalten wir

$$[z^n]G(z,v_0,\dots,v_d) = \frac{1}{n}[G^{n-1}]\left(\binom{d}{0}v_0 + \binom{d}{1}v_1G^1 + \dots + \binom{d}{d}v_dG^d\right)^n \tag{4.18}$$

und weiter

$$\begin{aligned} F_{n,n_0,\dots,n_d} &= [z^n v_0^{n_0}\cdots v_d^{n_d}]G(z,v_0,\dots,v_n) \\ &= \frac{1}{n}[G^{n-1}v_0^{n_0}\cdots v_d^{n_d}]\left(\binom{d}{0}v_0G^0 + \binom{d}{1}v_1G^1 + \dots + \binom{d}{d}v_dG^d\right)^n . \end{aligned} \tag{4.19}$$

Ablesen der Koeffizienten von $v_0^{n_0}$ führt durch einfache Anwendung des Binomischen Lehrsatzes zu

$$F_{n,n_0,\dots,n_d} = \frac{1}{n}\binom{d}{0}^{n_0}\binom{n}{n_0}\cdot[G^{n-1}v_1^{n_1}\dots v_d^{n_d}]G^{0\cdot n_0}\left(\binom{d}{1}v_1G^1 + \dots + \binom{d}{d}v_dG^d\right)^{n-n_0} . \tag{4.20}$$

Liest man nun die Koeffizienten von $v_1^{n_1}$ ab, ergibt sich weiter

$$\begin{aligned} F_{n,n_0,\dots,n_d} &= \frac{1}{n}\binom{d}{0}^{n_0}\binom{n}{n_0}\binom{d}{1}^{n_1}\binom{n-n_0}{n_1} \times \\ &\times [G^{n-1}v_2^{n_2}\dots v_d^{n_d}]G^{0\cdot n_0}G^{1\cdot n_1}\left(\binom{d}{2}v_2G^2 + \dots + \binom{d}{d}v_dG^d\right)^{n-n_0-n_1} . \end{aligned} \tag{4.21}$$

Wenn man nun nacheinander die Koeffizienten von $v_2^{n_2},\cdots,v_{d-1}^{n_{d-1}}$ abliest, wobei man wiederum nur den Binomischen Lehrsatz benötigt, erhält man

$$\begin{aligned} F_{n,n_0,\dots,n_d} &= \frac{1}{n}\binom{d}{0}^{n_0}\binom{n}{n_0}\binom{d}{1}^{n_1}\binom{n-n_0}{n_1}\cdots\binom{d}{i}^{n_i}\binom{n-n_0-\cdots-n_{i-1}}{n_i}\cdots \times \\ &\times \binom{d}{d-1}^{n_{d-1}}\binom{n-n_0-\cdots-n_{d-2}}{n_{d-1}} \times \\ &\times [G^{n-1}v_d^{n_d}]G^{0\cdot n_0}G^{1\cdot n_1}\cdots G^{in_i}\cdots G^{(d-1)n_{d-1}}\left(\binom{d}{d}v_dG^d\right)^{n-n_0-\cdots-n_{d-1}} . \end{aligned} \tag{4.22}$$

Liest man nun die Koeffizienten von $v_d^{n_d}$ ab, erhält man die erste der beiden im Satz angegebenen Bedingungen für eine nichtleere Anzahl an Bäumen mit dieser Verteilung an Knotengraden, nämlich $n_d = n - n_0 - \cdots n_{d-1}$.

Ist diese Bedingung erfüllt erhalten wir

$$\begin{aligned} F_{n,n_0,\dots,n_d} &= \frac{1}{n}\binom{d}{0}^{n_0}\binom{n}{n_0}\binom{d}{1}^{n_1}\binom{n-n_0}{n_1}\cdots\binom{d}{d-1}^{n_{d-1}}\binom{n-n_0-\cdots-n_{d-2}}{n_{d-1}}\binom{d}{d}^{n_d} \times \\ &\times [G^{n-1}]G^{0\cdot n_0+1\cdot n_1+\cdots+dn_d}, . \end{aligned} \tag{4.23}$$

Ablesen der Koeffizienten von G^{n-1} führt zur zweiten im Satz angegebenen Bedingung $n-1 = 0\cdot n_0 + 1\cdot n_1 + \cdots + dn_d$ für eine nichtleere Anzahl an Bäumen.

Bei Erfülltsein dieser eben genannten Bedingung bekommen wir also

$$F_{n,n_0,\ldots,n_d} = \frac{1}{n}\binom{d}{0}^{n_0}\binom{d}{1}^{n_1}\cdots\binom{d}{d}^{n_d}\binom{n}{n_0}\binom{n-n_0}{n_1}\cdots\binom{n-n_0-\cdots-n_{d-1}}{n_d} \tag{4.24}$$

und durch Zusammenfassen schließlich

$$F_{n,n_0,\ldots,n_d} = \frac{1}{n}\binom{n}{n_0\ n_1\ \cdots\ n_d}\binom{d}{0}^{n_0}\binom{d}{1}^{n_1}\cdots\binom{d}{d}^{n_d}. \tag{4.25}$$

□

4.4 Verallgemeinerung auf einfach erzeugte Baumfamilien

Die im letzten Abschnitt verwendete Methode bleibt nicht auf d-äre Bäume beschränkt, sondern kann praktisch unverändert auf eine viel allgemeinere Klasse von Bäumen angewandt werden. Vorausgesetzt werden muß lediglich die Anwendbarkeit der Lagrange'schen Inversionsformel, und für diese Baumfamilien hat sich seit einem Artikel von Meir und Moon [31] der Name "einfach erzeugte Familien von Bäumen" eingebürgert. Genauer soll eine Wurzelbaumfamilie $\mathcal{W}$ eine einfach erzeugte Familie sein, wenn sich die erzeugende Funktion $W(z)$ der Anzahl W_n der Bäume aus $\mathcal{W}$ der Größe n schreiben läßt als

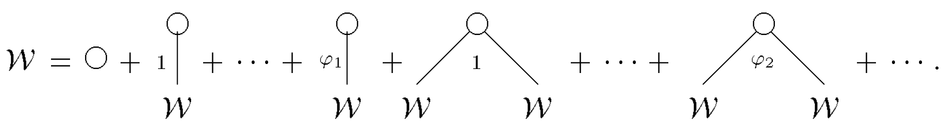
$$W(z) = z\,\Phi\left(W(z)\right) , \tag{4.26}$$

wobei

$$\Phi(W) = 1 + \varphi_1 W + \varphi_2 W^2 + \cdots = \sum_{i\geq 0} \varphi_i W^i$$

eine Potenzreihe mit nichtnegativen Koeffizienten φ_i für $i \geq 1$ und $\varphi_0 = 1$ darstellt.

Für unsere Zwecke wollen wir aus Gründen der kombinatorischen Interpretierbarkeit der Ergebnisse voraussetzen, daß die Koeffizienten φ_i von $\Phi(W)$ nichtnegative ganze Zahlen sein sollen. Dann läßt sich nämlich die Baumfamilie $\mathcal{W}$ durch folgende formale Gleichung beschreiben



(4.27)

Dabei soll es jeweils φ_i verschiedene Arten geben, wie diese i Kanten von einem Knoten wegführen können.

Gesucht ist nun die Anzahl der Bäume aus der Familie $\mathcal{W}$ der Größe n, die n_0 Knoten mit 0 Kindern, n_1 Knoten mit 1 Kind usw. besitzen. Wir bezeichnen diese Größe mit $W_{n,n_0,n_1,n_2,\ldots}$. Dafür gilt nun

Satz 4.4.1. *Sei $\mathcal{W}$ eine einfach erzeugte Familie von Bäumen $\mathcal{W}$, bei der sich die erzeugende Funktion $W(z)$ der Anzahl W_n der Bäume aus $\mathcal{W}$ der Größe n schreiben läßt als*

$$z = \frac{W(z)}{\Phi\left(W(z)\right)}\,, \text{ mit } \Phi(W) = \sum_{i\geq 0}\varphi_i W^i \text{ und } \varphi_0 = 1\,,\ \varphi_i \in \mathbb{N}_0 \text{ für } i \geq 1\,. \tag{4.28}$$

Dann ist die Anzahl der Bäume $W_{n,n_0,n_1,n_2,\ldots}$ aus $\mathcal{W}$ der Größe n, die für $i \geq 0$ genau n_i Knoten mit i Kindern besitzen, für $n \geq 1$ gegeben durch

$$W_{n,n_0,n_1,n_2,\ldots} = \frac{1}{n} \cdot \binom{n}{n_0\ n_1\ n_2\ \cdots} \cdot \varphi_0^{n_0} \varphi_1^{n_1} \varphi_2^{n_2} \cdots, \tag{4.29}$$
$$\text{wenn } \sum_{i \geq 0} n_i = n \text{ und } \sum_{i \geq 0} i\, n_i = n - 1 \text{ erfüllt sind und } W_{n,n_0,n_1,n_2,\ldots} = 0 \text{ sonst}.$$

Beweis. Wir definieren uns einen Operator $\widetilde{F}$, der jedem Baum W aus $\mathcal{W}$ einen Term $z^n v_0^{n_0} v_1^{n_1} v_2^{n_2} \cdots$ aus der Variablenmenge $\{z, v_0, v_1, v_2, \ldots\}$ zuordnet, und zwar derart, daß n die Größe von W und n_i, für $i \geq 0$, die Anzahl der Knoten in W mit genau i Kindern angibt. Es sind dabei natürlich nur endlich viele n_i ungleich 0.

Für eine Familie von Bäumen $(W_i)_{i \in I}$, $W_i \in \mathcal{W}$, soll der Operator $\widetilde{F}$ definiert sein durch

$$\widetilde{F}\left((W_i)_{i \in I}\right) := \sum_{i \in I} \widetilde{F}(B_i),$$

sofern gewährleistet ist, daß für jedes n nur endlich viele W_i der Größe n sind. Dann weist $\widetilde{F}$ jeder Familie $(W_i)_{i \in I}$ eine formale Potenzreihe über dem Alphabet $\{z, v_0, v_1, v_2, \ldots\}$ zu. Dabei sind in jedem Term $z^n v_0^{n_0} v_1^{n_1} v_2^{n_2} \cdots$ dieser formalen Reihe nur endlich viele n_i ungleich 0.

Definieren wir nun $W(z, v_0, v_1, v_2, \cdots)$ durch

$$W(z, v_0, v_1, v_2, \cdots) = \widetilde{F}(\mathcal{W}), \tag{4.30}$$

dann sind die gesuchten $W_{n,n_0,n_1,n_2,\ldots}$ gegeben durch die Beziehung

$$W_{n,n_0,n_1,n_2,\ldots} = [z^n v_0^{n_0} v_1^{n_1} v_2^{n_2} \cdots] W(z, v_0, v_1, v_2, \cdots). \tag{4.31}$$

Nun wenden wir den Operator $\widetilde{W}$ auf die formale Gleichung (4.27) an und erhalten so

$$W(z, v_0, v_1, \cdots) = z \left(\varphi_0 v_0 + \varphi_1 v_1 W(z, v_0, v_1, \cdots) + \varphi_2 v_2 W^2(z, v_0, v_1, \cdots) + \cdots\right) \tag{4.32}$$

bzw.

$$z = \frac{W(z, v_0, v_1, \cdots)}{\varphi_0 v_0 + \varphi_1 v_1 W(z, v_0, v_1, \cdots) + \varphi_2 v_2 W(z, v_0, v_1, \cdots)^2 + \cdots}. \tag{4.33}$$

Für das Ablesen des Koeffizienten $[z^n]$ von $W(z, v_0, v_1, \cdots)$ können wir also die Lagrange'sche Inversionsformel (Satz 4.3.2) verwenden, wenn wir $W(z, v_0, v_1, \cdots)$ als formale Potenzreihe in z auffassen. Damit erhalten wir

$$[z^n] W(z, v_0, v_1, \cdots) = \frac{1}{n} [W^{n-1}] \left(\varphi_0 v_0 + \varphi_1 v_1 G^1 + \varphi_2 v_2 G^2 + \cdots\right)^n \tag{4.34}$$

und weiter

$$\begin{aligned} W_{n,n_0,n_1,\ldots} &= [z^n v_0^{n_0} v_1^{n_1} \cdots] W(z, v_0, v_1, \ldots) \\ &= \frac{1}{n} [W^{n-1} v_0^{n_0} v_1^{n_1} \cdots] \left(\varphi_0 v_0 G^0 + \varphi_1 v_1 G^1 + \varphi_2 v_2 G^2 + \cdots\right)^n. \end{aligned} \tag{4.35}$$

Liest man nun nacheinander im letzten Ausdruck die Koeffizienten von $v_0^{n_0}, \cdots, v_{n-1}^{n_{n-1}}$ ab, bekommt man durch sukzessives Anwenden des Binomischen Lehrsatzes analog wie im Beweis von Satz 4.3.1:

$$\begin{aligned} W_{n,n_0,n_1,\ldots} = &\frac{1}{n}\varphi_0^{n_0}\binom{n}{n_0}\varphi_1^{n_1}\binom{n-n_0}{n_1}\cdots\varphi_i^{n_i}\binom{n-n_0-\cdots-n_{i-1}}{n_i}\cdots \times \\ &\times \varphi_n^{n_n}\binom{n-n_0-\cdots-n_{n-1}}{n_n} \times \\ &\times [W^{n-1-0\cdot n_0-1\cdot n_1-\cdots-n\cdot n_n} v_{n+1}^{n_{n+1}}\cdots]\left(\varphi_{n+1}v_{n+1}W^{n+1}+\cdots\right)^{n-n_0-\cdots-n_n}. \end{aligned} \tag{4.36}$$

Damit dieser Ausdruck ungleich 0 ist, muß zunächst $n_0+\cdots+n_n \leq n$ gelten, denn bei $n_0+\cdots+n_n > n$ ist mindestens ein Binomialkoeffizient $\binom{n-n_0-\cdots-n_{i-1}}{n_i}$ für $0 \leq i \leq n$ gleich 0. Ist diese Bedingung erfüllt, ergibt sich weiters, daß ein Ablesen der Koeffizienten der aufgetretenen Potenz von G nur dann zu einem Ausdruck ungleich 0 führt, wenn $n_0+\cdots+n_n \geq n$ ist. Damit erhalten wir also die Bedingung $n_0+\cdots+n_n = n$. Das heißt also, wir bekommen

$$\begin{aligned} &[W^{n-1-0\cdot n_0-1\cdot n_1-\cdots-n\cdot n_n} v_{n+1}^{n_{n+1}} v_{n+2}^{n_{n+2}}\cdots]\left(\varphi_{n+1}v_{n+1}W^{n+1}+\varphi_{n+2}v_{n+2}W^{n+2}+\cdots\right)^{n-n_0-\cdots-n_n} \\ &\quad = [W^{n-1-0\cdot n_0-1\cdot n_1-\cdots-n\cdot n_n} v_{n+1}^{n_{n+1}}\cdots]\,1\,. \end{aligned} \tag{4.37}$$

Daraus erhalten wir für ein Ergebnis ungleich 0 noch die Bedingungen $n_i = 0$ für $i \geq n+1$ und $0\cdot n_0+1\cdot n_1+\cdots+n\cdot n_n = n-1$, insgesamt also die bei der Definition des Satzes angegebenen Bedingungen, die dem Abzählen der Knoten bzw. der Kanten im Baum entsprechen. Sind die Bedingungen allesamt erfüllt, ergibt sich

$$[W^{n-1-0\cdot n_0-1\cdot n_1-\cdots-n\cdot n_n} v_{n+1}^{n_{n+1}}\cdots]\,1 = 1$$

und somit durch Kombination aller Resultate

$$W_{n,n_0,n_1,\ldots} = \frac{1}{n}\binom{n}{n_0\ n_1\ n_2\ \cdots}\varphi_0^{n_0}\varphi_1^{n_1}\varphi_2^{n_2}\cdots. \tag{4.38}$$

□

Diesen Satz wollen wir nun auf ein paar wichtige Klassen von einfach erzeugten Baumfamilien anwenden.

- Ebene Wurzelbäume

  Hier ergibt sich ein Resultat, das neben Rote's Artikel [41] auch schon in [46] erschienen ist. Es gilt

  $$\Phi(W) = 1+W+W^2+\cdots = \frac{1}{1-W},$$

  also $\varphi_i = 1$ für $i \geq 0$. Damit erhalten wir

Satz 4.4.2. *Die Anzahl der Ebenen Wurzelbäume $P_{n,n_0,n_1,\ldots}$ der Größe n, die für $i \geq 0$ genau n_i Knoten mit i Kindern besitzen, ist gegeben durch*

$$P_{n,n_0,n_1,n_2,\ldots} = \frac{1}{n}\binom{n}{n_0\ n_1\ n_2\ \cdots} \text{ wenn } \sum_{i\geq 0} n_i = n \text{ und } \sum_{i\geq 0} i\, n_i = n-1 \text{ gilt}$$

und

$$P_{n,n_0,n_1,n_2,\ldots} = 0 \text{ sonst}.$$

- Motzkin-Bäume

  Hier gilt

$$\Phi(W) = 1 + W + W^2 .$$

Das ergibt

Satz 4.4.3. *Die Anzahl der Motzkin-Bäume M_{n,n_0,n_1,n_2} der Größe n, die genau n_0 Knoten mit 0 Kindern, n_1 Knoten mit 1 Kind und n_2 Knoten mit 2 Kindern besitzen, ist gegeben durch*

$$M_{n,n_0,n_1,n_2} = \frac{1}{n}\binom{n}{n_0\ n_1\ n_2} \text{ wenn } n_0 + n_1 + n_2 = n \text{ und } n_1 + 2n_2 = n-1 \text{ gilt}$$

und

$$M_{n,n_0,n_1,n_2} = 0 \text{ sonst}.$$

- Binärbäume ohne Unterscheidung von internen und externen Knoten

  Hier gilt

$$\Phi(W) = 1 + W^2 .$$

Anwenden des Satzes ergibt zunächst bei Erfülltsein der Bedingungen $n_0+n_1+n_2 = n$ und $n_1 + 2n_2 = n-1$

$$\widetilde{B}_{n,n_0,n_1,n_2} = \frac{1}{n}\binom{n}{n_0\ n_1\ n_2} \cdot 1^{n_0} \cdot 0^{n_1} \cdot 1^{n_2} , \tag{4.39}$$

wenn $\widetilde{B}_{n,n_0,n_1,n_2}$ die Anzahl der Binärbäume mit n (internen und externen) Knoten, die für $0 \leq i \leq 2$ genau n_i Knoten mit i Kindern haben, bezeichnet.

Als notwendige Bedingung, daß dieser Ausdruck nicht 0 ist, kommt also noch $n_1 = 0$ hinzu. Dies führt schließlich zu den Bedingungen $n = 2m+1$, $n_0 = m+1$, $n_1 = 0$ und $n_2 = m$, mit $m \in \mathbb{N}$, woraus sich

$$\widetilde{B}_{2m+1,m+1,0,m} = \frac{1}{2m+1}\binom{2m+1}{m+1\ 0\ m} = \frac{1}{m+1}\binom{2m}{m} \tag{4.40}$$

ergibt.

Dies führt zu dem natürlich wohlbekannten Resultat

Satz 4.4.4. *Die Anzahl der Binärbäume $\widetilde{B}_{n,n_0,n_1,n_2}$ mit n (internen und externen) Knoten, die genau n_0 Knoten mit 0 Kindern, n_1 Knoten mit 1 Kind und n_2 Knoten mit 2 Kindern besitzen, ist gegeben durch*

$$\widetilde{B}_{n,n_0,n_1,n_2} = \frac{1}{m+1}\binom{2m}{m} \quad \text{wenn} \quad n = 2m+1\,,\; n_0 = m+1\,,\; n_1 = 0 \text{ und } n_2 = m$$
$$\text{mit} \quad m \in \mathbb{N} \text{ gilt}$$

und

$$\widetilde{B}_{n,n_0,n_1,n_2} = 0 \quad \text{sonst}\,.$$

4.5 Die Verteilung der Knotengrade von entsprechend der Inorder-Traversierung markierten Knoten

Wir werden hier Resultate wiedergewinnen, die Sprugnoli in [45] veröffentlicht hat, zeigen aber, wie man diese mithilfe der Operatormethode relativ rasch erhalten kann. Wir bezeichnen im folgenden mit $I_{n,j}$ die Anzahl der Binärbäume der Größe n, bei denen der entsprechend der Reihenfolge der Inorder-Traversierung mit j markierte Knoten ein sogenannter innerer Knoten ist, also ein Knoten mit 2 Kindern.

Weiters sei $F_{n,j}$ die Anzahl der Binärbäume der Größe n, bei denen der entsprechend der Reihenfolge der Inorder-Traversierung mit j markierte Knoten ein Blatt ist, also keine Kinder besitzt.

Abweichend zur Terminologie in Sprugnoli's Artikel werden wir mit $R_{n,j}$ die Anzahl der Binärbäume der Größe n bezeichnen, bei denen der entsprechend der Reihenfolge der Inorder-Traversierung mit j markierte Knoten genau ein Kind besitzt und dieses vom Vaterknoten durch eine nach rechts führende Kante erreicht wird.

Weiters bezeichnen wir mit $L_{n,j}$ die Anzahl der Binärbäume der Größe n, bei denen der entsprechend der Reihenfolge der Inorder-Traversierung mit j markierte Knoten genau ein Kind besitzt und dieses vom Vaterknoten durch eine nach links führende Kante erreicht wird.

Die zugehörigen bivariaten erzeugenden Funktionen $I(z,u)$, $F(z,u)$, $R(z,u)$ und $L(z,u)$ sind dann definiert durch

$$I(z,u) = \sum_{n\geq 0}\sum_{j\geq 0} I_{n,j} z^n u^j\,, \quad F(z,u) = \sum_{n\geq 0}\sum_{j\geq 0} F_{n,j} z^n u^j\,,$$
$$R(z,u) = \sum_{n\geq 0}\sum_{j\geq 0} R_{n,j} z^n u^j\,, \quad L(z,u) = \sum_{n\geq 0}\sum_{j\geq 0} L_{n,j} z^n u^j\,.$$

4.5.1 Die Blätter

Die formale Gleichung der Binärbäume (4.1) übersetzt sich mit der symbolischen Operatormethode in folgende Funktionalgleichung

$$F(z,u) = zu + zF(z,u) + zuF(z,u) + F(z,u)zB^*(z) + B^*(zu)zuF(z,u)\,, \tag{4.41}$$

bzw. die explizite Darstellung

$$F(z,u) = \frac{zu}{1 - z - zu - zB^*(z) - zuB^*(zu)} \,. \tag{4.42}$$

Unter Verwendung der in der Einleitung dieses Kapitels definierten Abkürzungen X und Y vereinfacht sich dies zu

$$F(z,u) = \frac{2zu}{X+Y} \,. \tag{4.43}$$

Ablesen der Koeffizienten dieses Ausdrucks führt unter Benützung der Gleichung (3.72) sofort zu

$$\begin{aligned} F_{n,j} &= [z^n u^j] F(z,u) = [z^n u^j] \frac{2zu}{X+Y} = 2[z^{n-1}u^{j-1}] \frac{1}{X+Y} \\ &= B_{n-1} = \frac{1}{n}\binom{2(n-1)}{n-1} \quad \text{für } n \geq 1 \,. \end{aligned} \tag{4.44}$$

Dies ergibt also

Satz 4.5.1. *Die Anzahl der Binärbäume $F_{n,j}$ der Größe n, bei denen der entsprechend der Reihenfolge der Inorder-Traversierung markierte Knoten j ein Blatt ist, ist für $1 \leq j \leq n$ gegeben durch*

$$\boxed{F_{n,j} = B_{n-1} = \frac{1}{n}\binom{2(n-1)}{n-1} \,.} \tag{4.45}$$

4.5.2 Die Knoten mit 2 Kindern

Wir erhalten hier aus der formalen Gleichung (4.1) folgende Funktionalgleichung für $I(z,u)$:

$$I(z,u) = zI(z,u) + zuI(z,u) + B^*(zu)zuB^*(z) + I(z,u)zB^*(z) + B^*(zu)zuI(z,u) \,, \tag{4.46}$$

aus der man die explizite Darstellung

$$I(z,u) = \frac{zuB^*(z)B^*(zu)}{1 - z - zu - zB^*(z) - zuB^*(zu)} \tag{4.47}$$

gewinnt.

Dies formt man unter Verwendung der oben definierten Abkürzungen um zu

$$\begin{aligned} I(z,u) &= \frac{Y^2 - 2Y + 1 - XY^2 + 2XY - Y}{2(X+Y)(1+X)} = \frac{1-Y^2}{2(X+Y)} - \frac{1-Y}{1+X} \\ &= \frac{2zu}{X+Y} - \frac{XY - X - Y + 1}{4z} \,. \end{aligned} \tag{4.48}$$

Mit

$$[z^n]X = -2B_{n-1} = -\frac{2}{n}\binom{2(n-1)}{n-1} \tag{4.49}$$

und der Beziehung (3.72) erhält man durch Ablesen der Koeffizienten

$$\begin{aligned} I_{n,j} &= [z^n u^j] I(z,u) = 2[z^{n-1}u^{j-1}]\frac{1}{X+Y} - \frac{1}{4}[z^{n+1}u^j]XY \\ &= B_{n-1} - \frac{1}{4}[z^{n+1-j}]X[(uz)^j]Y = B_{n-1} - B_{n-j}B_{j-1} \\ &= \frac{1}{n}\binom{2(n-1)}{n-1} - \frac{1}{j}\binom{2(j-1)}{j-1}\frac{1}{n+1-j}\binom{2(n-j)}{n-j}. \end{aligned} \tag{4.50}$$

Dies ergibt also

Satz 4.5.2. *Die Anzahl der Binärbäume $I_{n,j}$ der Größe n, bei denen der entsprechend der Reihenfolge der Inorder-Traversierung markierte Knoten j 2 Kinder besitzt, ist für $1 \le j \le n$ gegeben durch*

$$\boxed{I_{n,j} = B_{n-1} - B_{n-j}B_{j-1} = \frac{1}{n}\binom{2(n-1)}{n-1} - \frac{1}{j}\binom{2(j-1)}{j-1}\frac{1}{n+1-j}\binom{2(n-j)}{n-j}.} \tag{4.51}$$

4.5.3 Die Knoten mit 1 Kind, das durch eine rechte Kante erreicht wird

Wir erhalten als Funktionalgleichung für $R(z,u)$:

$$R(z,u) = zR(z,u) + zuB^*(z) + zuR(z,u) + R(z,u)zB^*(z) + B^*(zu)zuR(z,u). \tag{4.52}$$

Dies führt zur expliziten Darstellung

$$R(z,u) = \frac{zuB^*(z)}{1 - z - zu - zB^*(z) - zuB^*(zu)}. \tag{4.53}$$

Daraus erhält man durch Vereinfachen

$$\begin{aligned} R(z,u) &= \frac{1 + XY^2 - X - Y^2}{2(X+Y)(1+X)} = \frac{(1+Y)^2}{2(X+Y)} - \frac{1+Y}{1+X} \\ &= \frac{1-2uz}{X+Y} - \frac{X}{X+Y} + \frac{XY}{4z} + 1 + \frac{X-Y-1}{4z}. \end{aligned} \tag{4.54}$$

Für das Koeffizientenablesen dieses Ausdrucks benötigt man außer den Beziehungen (3.72) und (4.49) noch

$$\begin{aligned} [z^n u^j]\frac{X}{X+Y} &= [z^n u^j]\frac{X^2}{X(X+Y)} = [z^n u^j]\frac{1-4z}{X(X+Y)} \\ &= [z^n u^j]\frac{1}{X(X+Y)} - 4[z^{n-1}u^j]\frac{1}{X(X+Y)} = \frac{2j-n}{2n(n+1)}\binom{2j}{j}\binom{2(n-j)}{n-j}, \end{aligned} \tag{4.55}$$

wobei man die letzte Gleichung aus Beziehung (3.94) folgert.

Damit bekommt man

$$\begin{aligned} R_{n,m} &= [z^n u^j] R(z,u) \\ &= [z^n u^j] \frac{1}{X+Y} - 2[z^{n-1}u^{j-1}] \frac{1}{X+Y} - [z^n u^j] \frac{X}{X+Y} + \frac{1}{4}[z^{n+1}u^j] XY \qquad (4.56) \\ &= \frac{1}{2(n+1)} \left[\frac{n-2}{2n-1} \binom{2n}{n} - \frac{2j-n-2}{n} \binom{2(j-1)}{j-1} \binom{2(n+1-j)}{n+1-j} \right]. \qquad (4.57) \end{aligned}$$

Somit ergibt sich

Satz 4.5.3. *Die Anzahl der Binärbäume $R_{n,j}$ der Größe n, bei denen der entsprechend der Reihenfolge der Inorder-Traversierung markierte Knoten j genau 1, mit einer nach rechts führenden Kante verbundenes Kind besitzt, ist für $1 \le j \le n$ gegeben durch*

$$\boxed{R_{n,j} = \frac{1}{2(n+1)} \left[\frac{n-2}{2n-1} \binom{2n}{n} - \frac{2j-n-2}{n} \binom{2(j-1)}{j-1} \binom{2(n+1-j)}{n+1-j} \right].} \qquad (4.58)$$

4.5.4 Die Knoten mit 1 Kind, das durch eine linke Kante erreicht wird

Die Funktionalgleichung für $L(z,u)$ lautet

$$L(z,u) = zuB^*(zu) + zL(z,u) + zuL(z,u) + L(z,u)zB^*(z) + B^*(zu)zuL(z,u). \qquad (4.59)$$

Dies formt man um zur expliziten Darstellung für $L(z,u)$:

$$L(z,u) = \frac{zuB^*(zu)}{1 - z - zu - zB^*(z) - zuB^*(zu)}. \qquad (4.60)$$

Nach Vereinfachungen erhält man daraus

$$L(z,u) = \frac{1 - 2Y + Y^2}{2(X+Y)} = \frac{1-2uz}{X+Y} - 1 + \frac{X}{X+Y}. \qquad (4.61)$$

Das Ablesen der Koeffizienten dieses Ausdrucks führt unter Zuhilfenahme der Beziehungen (3.72) und (4.55) zu

$$\begin{aligned} L_{n,j} = [z^n u^j] L(z,u) &= [z^n u^j] \frac{1}{X+Y} - 2[z^{n-1}u^{j-1}] \frac{1}{X+Y} + [z^n u^j] \frac{X}{X+Y} \\ &= \frac{1}{2(n+1)} \left[\frac{n-2}{2n-1} \binom{2n}{n} + \frac{2j-n}{n} \binom{2j}{j} \binom{2(n-j)}{n-j} \right]. \qquad (4.62) \end{aligned}$$

Das ergibt

Satz 4.5.4. *Die Anzahl der Binärbäume $L_{n,j}$ der Größe n, bei denen der entsprechend der Reihenfolge der Inorder-Traversierung markierte Knoten j genau 1, mit einer nach links führenden Kante verbundenes Kind besitzt, ist für $1 \le j \le n$ gegeben durch*

$$\boxed{L_{n,j} = R_{n,n+1-j} = \frac{1}{2(n+1)} \left[\frac{n-2}{2n-1} \binom{2n}{n} + \frac{2j-n}{n} \binom{2j}{j} \binom{2(n-j)}{n-j} \right].} \qquad (4.63)$$

4.5.5 Der Erwartungswert der Anzahl an Kindern

Bezeichnen wir mit $E_{n,j}$ den Erwartungswert der Anzahl an Kindern des entsprechend der Reihenfolge der Inorder-Traversierung markierten Knotens j in einem Binärbaum der Größe n.

Aus den eben errechneten Größen ist dieser Wert mittels

$$E_{n,j} = \frac{2\,I_{n,j} + L_{n,j} + R_{n,j}}{\frac{1}{n+1}\binom{2n}{n}} \tag{4.64}$$

leicht zu erhalten.

Es ergibt sich so

Satz 4.5.5. *Der Erwartungswert $E_{n,j}$ der Anzahl an Kindern des entsprechend der Reihenfolge der Inorder-Traversierung markierten Knotens j in einem Binärbaum der Größe n ist für $1 \leq j \leq n$ gegeben durch*

$$\boxed{E_{n,j} = 1 - \frac{n+1}{2(n+1-j)(2j-1)} \frac{\binom{2j}{j}\binom{2(n-j)}{n-j}}{\binom{2n}{n}}.} \tag{4.65}$$

4.6 Die Verteilung der Knotengrade von entsprechend der Preorder-Traversierung markierten Knoten

Nun gehen wir daran, die im letzten Abschnitt untersuchten Größen auch für Preorder-traversierte Binärbäume zu gewinnen. Wir bezeichnen nun mit $I_{n,j}$ die Anzahl der Binärbäume der Größe n, bei denen der entsprechend der Reihenfolge der Preorder-Traversierung mit j markierte Knoten ein sogenannter innerer Knoten ist, also ein Knoten mit 2 Kindern.

Dann sei $F_{n,j}$ die Anzahl der Binärbäume der Größe n, bei denen der entsprechend der Reihenfolge der Preorder-Traversierung mit j markierte Knoten ein Blatt ist, also keine Kinder besitzt.

Weiters bezeichnen wir mit $R_{n,j}$ die Anzahl der Binärbäume der Größe n, bei denen der entsprechend der Reihenfolge der Preorder-Traversierung mit j markierte Knoten genau ein Kind besitzt und dieses vom Vaterknoten durch eine nach rechts führende Kante erreicht wird.

Und schließlich soll $L_{n,j}$ die Anzahl der Binärbäume der Größe n darstellen, bei denen der entsprechend der Reihenfolge der Preorder-Traversierung mit j markierte Knoten genau ein Kind besitzt und dieses vom Vaterknoten durch eine nach links führende Kante erreicht wird.

Die zugehörigen bivariaten erzeugenden Funktionen $I(z,u)$, $F(z,u)$, $R(z,u)$ und $L(z,u)$ sind also definiert durch

$$I(z,u) = \sum_{n\geq 0}\sum_{j\geq 0} I_{n,j} z^n u^j\,, \quad F(z,u) = \sum_{n\geq 0}\sum_{j\geq 0} F_{n,j} z^n u^j\,,$$
$$R(z,u) = \sum_{n\geq 0}\sum_{j\geq 0} R_{n,j} z^n u^j\,, \quad L(z,u) = \sum_{n\geq 0}\sum_{j\geq 0} L_{n,j} z^n u^j\,.$$

4.6.1 Die Blätter

Wir erhalten aus der formalen Gleichung (4.1) als Funktionalgleichung für $F(z,u)$

$$F(z,u) = zu + 2zuF(z,u) + F(z,u)zuB^*(z) + B^*(zu)zuF(z,u)\,, \tag{4.66}$$

und daraus sofort

$$F(z,u) = \frac{zu}{1 - 2zu - zuB^*(z) - zuB^*(zu)}\,. \tag{4.67}$$

Unter Verwendung der zu Beginn dieses Kapitels definierten abkürzenden Schreibweisen, vereinfacht sich dies zu

$$F(z,u) = -\frac{1}{2}\frac{XY - X + Y - 1}{X+Y} = \frac{1}{2}\frac{(1-Y)^2}{X+Y} - \frac{Y}{2} + \frac{1}{2}\,. \tag{4.68}$$

Für $1 \le j \le n-1$ führt das Ablesen der Koeffizienten dieses Ausdrucks unter Verwendung von (3.72) und (4.55) zu

$$\begin{aligned} F_{n,j} = [z^n u^j]F(z,u) &= \frac{1}{2}[z^n u^j]\frac{(1-Y)^2}{X+Y} \\ &= [z^n u^j]\frac{1}{X+Y} - 2[z^{n-1}u^{j-1}]\frac{1}{X+Y} + [z^n u^j]\frac{X}{X+Y} \\ &= \frac{1}{2(n+1)}\left[\frac{n-2}{2n-1}\binom{2n}{n} + \frac{2j-n}{n}\binom{2j}{j}\binom{2(n-j)}{n-j}\right]\,, \end{aligned} \tag{4.69}$$

während sich für $n = j$

$$F_{n,n} = \frac{1}{n+1}\binom{2n}{n} \tag{4.70}$$

ergibt.

Alles zusammengefaßt führt dies zu

Satz 4.6.1. *Die Anzahl der Binärbäume $F_{n,j}$ der Größe n, bei denen der entsprechend der Reihenfolge der Preorder-Traversierung markierte Knoten j ein Blatt ist, ist gegeben durch*

$$\boxed{\begin{aligned} &F_{n,j} = \frac{1}{2(n+1)}\left[\frac{n-2}{2n-1}\binom{2n}{n} + \frac{2j-n}{n}\binom{2j}{j}\binom{2(n-j)}{n-j}\right] \\ &\quad \text{für } 1 \le j \le n-1\,, \\ &F_{n,n} = \frac{1}{n}\binom{2(n-1)}{n-1} \quad \text{für } n \ge 1\,. \end{aligned}} \tag{4.71}$$

4.6.2 Die Knoten mit 2 Kindern

Wir bekommen hier für $I(z,u)$ die Funktionalgleichung

$$I(z,u) = 2zuI(z,u) + zu\left(B^*(z)\right)^2 + I(z,u)zuB^*(z) + B^*(zu)zuI(z,u)\,, \tag{4.72}$$

woraus man sofort die explizite Darstellung

$$I(z,u) = \frac{zu\,(B^*(z))^2}{1-2zu-zuB^*(z)-zuB^*(zu)} \tag{4.73}$$

gewinnt.

Dies formt man um zu

$$I(z,u) = -\frac{1}{2}\frac{X^2Y-X^2-2XY+2X+Y-1}{(X+Y)(1+X)} = \frac{1}{2}\frac{(1+Y)^2}{X+Y} - \frac{Y}{2} + \frac{1}{2} - \frac{2}{X+1}\,. \tag{4.74}$$

Das Ablesen der Koeffizienten dieses Ausdrucks führt unter Zuhilfenahme der Gleichungen (3.72) und (4.55) für $1 \le j \le n-1$ zu

$$\begin{aligned} I_{n,j} = [z^n u^j] I(z,u) &= \frac{1}{2}[z^n u^j]\frac{(1+Y)^2}{X+Y} \\ &= [z^n u^j]\frac{1}{X+Y} - 2[z^{n-1}u^{j-1}]\frac{1}{X+Y} - [z^n u^j]\frac{X}{X+Y} \\ &= \frac{1}{2(n+1)}\left[\frac{n-2}{2n-1}\binom{2n}{n} - \frac{2j-n}{n}\binom{2j}{j}\binom{2(n-j)}{n-j}\right] \end{aligned} \tag{4.75}$$

und für $n=j$ zu $I_{n,n}=0$.

Dies faßt man zusammen in

Satz 4.6.2. *Die Anzahl der Binärbäume $I_{n,j}$ der Größe n, bei denen der entsprechend der Reihenfolge der Preorder-Traversierung markierte Knoten j 2 Kinder besitzt, ist gegeben durch*

$$\boxed{\begin{aligned} &I_{n,j} = \frac{1}{2(n+1)}\left[\frac{n-2}{2n-1}\binom{2n}{n} - \frac{2j-n}{n}\binom{2j}{j}\binom{2(n-j)}{n-j}\right] \\ &\quad \text{für } 1 \le j \le n-1\,, \\ &I_{n,n} = 0 \quad \text{für } n \ge 1\,. \end{aligned}} \tag{4.76}$$

4.6.3 Die Knoten mit 1 Kind, das durch eine rechte Kante erreicht wird

Die Funktionalgleichung für $R(z,u)$ lautet:

$$R(z,u) = zuB^*(z) + 2zuR(z,u) + R(z,u)zuB^*(z) + B^*(zu)zuR(z,u)\,. \tag{4.77}$$

Daraus gewinnt man die explizite Darstellung

$$R(z,u) = \frac{zuB^*(z)}{1-2zu-zuB^*(z)-zuB^*(zu)}\,, \tag{4.78}$$

welche sich nach Umformungen schreiben läßt als

$$R(z,u) = \frac{1}{2}\frac{XY-X-Y+1}{X+Y} = \frac{1}{2}\frac{1-Y^2}{X+Y} + \frac{Y}{2} - \frac{1}{2} = \frac{2zu}{X+Y} + \frac{Y}{2} - \frac{1}{2}\,. \tag{4.79}$$

Koeffizientenablesen dieses Ausdrucks führt für $1 \leq j \leq n-1$ zu

$$R_{n,j} = [z^n u^j]R(z,u) = 2[z^{n-1}u^{j-1}]\frac{1}{X+Y} = \frac{1}{n}\binom{2(n-1)}{n-1}, \tag{4.80}$$

während für $n = j$ $R_{n,n} = 0$ gilt.

Somit ergibt sich

Satz 4.6.3. *Die Anzahl der Binärbäume $R_{n,j}$ der Größe n, bei denen der entsprechend der Reihenfolge der Preorder-Traversierung markierte Knoten j genau 1, mit einer nach rechts führenden Kante verbundenes Kind besitzt, ist gegeben durch*

$$\boxed{\begin{aligned} &R_{n,j} = \frac{1}{n}\binom{2(n-1)}{n-1} \quad \text{für } 1 \leq j \leq n-1\,, \\ &R_{n,n} = 0 \quad \text{für } n \geq 1\,. \end{aligned}} \tag{4.81}$$

4.6.4 Die Knoten mit 1 Kind, das durch eine linke Kante erreicht wird

Die Funktionalgleichung für $L(z,u)$ lautet

$$L(z,u) = zuB^*(z) + 2zuL(z,u) + L(z,u)zuB^*(z) + B^*(zu)zuL(z,u)\,, \tag{4.82}$$

woraus man die explizite Darstellung

$$L(z,u) = \frac{zuB^*(z)}{1 - 2zu - zuB^*(z) - zuB^*(zu)}\,, \tag{4.83}$$

erhält.

Das heißt also die erzeugenden Funktionen $L(z,u)$ und $R(z,u)$ sind ident, also sind natürlich $L_{n,j}$ und $R_{n,j}$ gleich und man bekommt

Satz 4.6.4. *Die Anzahl der Binärbäume $L_{n,j}$ der Größe n, bei denen der entsprechend der Reihenfolge der Preorder-Traversierung markierte Knoten j genau 1, mit einer nach links führenden Kante verbundenes Kind besitzt, ist gegeben durch*

$$\boxed{\begin{aligned} &L_{n,j} = \frac{1}{n}\binom{2(n-1)}{n-1} \quad \text{für } 1 \leq j \leq n-1\,, \\ &L_{n,n} = 0 \quad \text{für } n \geq 1\,. \end{aligned}} \tag{4.84}$$

4.6.5 Der Erwartungswert der Anzahl an Kindern

Bezeichnen wir mit $E_{n,j}$ den Erwartungswert der Anzahl an Kindern des entsprechend der Reihenfolge der Preorder-Traversierung markierten Knotens j in einem Binärbaum der Größe n.

Dafür erhält man nun

Satz 4.6.5. *Der Erwartungswert $E_{n,j}$ der Anzahl an Kindern des entsprechend der Reihenfolge der Preorder-Traversierung markierten Knotens j in einem Binärbaum der Größe n ist für $1 \leq j \leq n$ gegeben durch*

$$\boxed{E_{n,j} = 1 + \frac{n-2j}{n} \frac{\binom{2j}{j}\binom{2(n-j)}{n-j}}{\binom{2n}{n}}.} \tag{4.85}$$

4.7 Die Verteilung der Knotengrade von entsprechend der Postorder-Traversierung markierten Knoten

Nun bezeichnen wir mit $I_{n,j}$ die Anzahl der Binärbäume der Größe n, bei denen der entsprechend der Reihenfolge der Postorder-Traversierung mit j markierte Knoten ein sogenannter innerer Knoten ist, also ein Knoten mit 2 Kindern.

Es sei nun $F_{n,j}$ die Anzahl der Binärbäume der Größe n, bei denen der entsprechend der Reihenfolge der Postorder-Traversierung mit j markierte Knoten ein Blatt ist, also keine Kinder besitzt.

Weiters bezeichnen wir mit $R_{n,j}$ die Anzahl der Binärbäume der Größe n, bei denen der entsprechend der Reihenfolge der Postorder-Traversierung mit j markierte Knoten genau ein Kind besitzt und dieses vom Vaterknoten durch eine nach rechts führende Kante erreicht wird.

Und schließlich soll $L_{n,j}$ die Anzahl der Binärbäume der Größe n darstellen, bei denen der entsprechend der Reihenfolge der Postorder-Traversierung mit j markierte Knoten genau ein Kind besitzt und dieses vom Vaterknoten durch eine nach links führende Kante erreicht wird.

Weiters definieren wir uns die bivariaten erzeugenden Funktionen $I(z,u)$, $F(z,u)$, $R(z,u)$ und $L(z,u)$ durch

$$I(z,u) = \sum_{n\geq 0}\sum_{j\geq 0} I_{n,j} z^n u^j\,, \quad F(z,u) = \sum_{n\geq 0}\sum_{j\geq 0} F_{n,j} z^n u^j\,,$$
$$R(z,u) = \sum_{n\geq 0}\sum_{j\geq 0} R_{n,j} z^n u^j\,, \quad L(z,u) = \sum_{n\geq 0}\sum_{j\geq 0} L_{n,j} z^n u^j\,.$$

4.7.1 Die Blätter

Die formale Gleichung (4.1) übersetzt sich in folgende Funktionalgleichung für $F(z,u)$:

$$F(z,u) = zu + 2zF(z,u) + F(z,u)zB^*(z) + B^*(zu)zF(z,u)\,. \tag{4.86}$$

Daraus erhalten wir die explizite Darstellung

$$F(z,u) = \frac{zu}{1-2z-zB^*(z)-zB^*(zu)}\,. \tag{4.87}$$

Unter Verwendung der Abkürzungen läßt sich dies schreiben als

$$\begin{aligned} F(z,u) &= \frac{(1-Y)(1+Y)^2}{2(X+Y)(X+1)} = \frac{(1+Y)^2}{2(X+Y)} - \frac{(1+Y)^2}{2(X+1)} \\ &= \frac{1}{X+Y} - \frac{2zu}{X+Y} - \frac{X}{X+Y} + \frac{XY}{4z} - \frac{1-2zu+Y}{4z} + \frac{(1-2zu)X}{4z} . \end{aligned} \tag{4.88}$$

Für $2 \le j \le n$ führt das Ablesen der Koeffizienten dieses Ausdrucks unter Verwendung von (3.72), (4.49) und (4.55) zu

$$\begin{aligned} F_{n,j} &= [z^n u^j] F(z,u) \\ &= [z^n u^j] \frac{1}{X+Y} - 2[z^{n-1}u^{j-1}] \frac{1}{X+Y} - [z^n u^j] \frac{X}{X+Y} + \frac{1}{4}[z^{n+1}u^j] XY \\ &= \frac{1}{2(n+1)} \left[\frac{n-2}{2n-1} \binom{2n}{n} + \frac{j(2j-n-2)(2j-2n-1)}{n(n+1-j)(2j-1)} \binom{2j}{j} \binom{2(n-j)}{n-j} \right] , \end{aligned} \tag{4.89}$$

während sich für $j = 1$

$$F_{n,1} = \frac{1}{n+1} \binom{2n}{n} \tag{4.90}$$

ergibt.

Alles zusammengefaßt führt dies zu

Satz 4.7.1. *Die Anzahl der Binärbäume $F_{n,j}$ der Größe n, bei denen der entsprechend der Reihenfolge der Postorder-Traversierung markierte Knoten j ein Blatt ist, ist gegeben durch*

$$\boxed{\begin{aligned} &F_{n,j} = \frac{1}{2(n+1)} \left[\frac{n-2}{2n-1} \binom{2n}{n} + \frac{j(2j-n-2)(2j-2n-1)}{n(n+1-j)(2j-1)} \binom{2j}{j} \binom{2(n-j)}{n-j} \right] \\ &\quad \text{für } 2 \le j \le n \, , \\ &F_{n,1} = \frac{1}{n} \binom{2(n-1)}{n-1} \quad \text{für } n \ge 1 \, . \end{aligned}} \tag{4.91}$$

4.7.2 Die Knoten mit 2 Kindern

Für $I(z,u)$ ergibt sich die Funktionalgleichung

$$I(z,u) = 2zI(z,u) + zu\,(B^*(zu))^2 + I(z,u)zB^*(z) + B^*(zu)zI(z,u) \tag{4.92}$$

bzw. die explizite Darstellung

$$I(z,u) = \frac{zu\,(B^*(zu))^2}{1-2z-zB^*(z)-zB^*(zu)} . \tag{4.93}$$

Durch Umformungen erhält man weiter

$$\begin{aligned} I(z,u) &= \frac{(1-Y)^3}{2(X+Y)(X+1)} = \frac{(1-Y)^2}{2(X+Y)} - \frac{(1-Y)^2}{2(X+1)} \\ &= \frac{1}{X+Y} - \frac{2zu}{X+Y} + \frac{X}{X+Y} - \frac{XY}{4z} - \frac{1-2zu-Y}{4z} + \frac{(1-2zu)X}{4z} . \end{aligned} \tag{4.94}$$

Für $2 \leq j \leq n$ ergibt das Ablesen der Koeffizienten unter Zuhilfenahme der Gleichungen (3.72), (4.49) und (4.55)

$$\begin{aligned} I_{n,j} &= [z^n u^j] I(z,u) \\ &= [z^n u^j]\frac{1}{X+Y} - 2[z^{n-1}u^{j-1}]\frac{1}{X+Y} + [z^n u^j]\frac{X}{X+Y} - \frac{1}{4}[z^{n+1}u^j]XY \\ &= \frac{1}{2(n+1)}\left[\frac{n-2}{2n-1}\binom{2n}{n} - \frac{j(2j-n-2)(2j-2n-1)}{n(n+1-j)(2j-1)}\binom{2j}{j}\binom{2(n-j)}{n-j}\right], \end{aligned} \tag{4.95}$$

während für $j = 1$ $I_{n,1} = 0$ gilt.

Dies faßt man zusammen in

Satz 4.7.2. *Die Anzahl der Binärbäume $I_{n,j}$ der Größe n, bei denen der entsprechend der Reihenfolge der Postorder-Traversierung markierte Knoten j 2 Kinder besitzt, ist gegeben durch*

$$\boxed{\begin{aligned} &I_{n,j} = \frac{1}{2(n+1)}\left[\frac{n-2}{2n-1}\binom{2n}{n} - \frac{j(2j-n-2)(2j-2n-1)}{n(n+1-j)(2j-1)}\binom{2j}{j}\binom{2(n-j)}{n-j}\right] \\ &\quad \text{für } 2 \leq j \leq n\,, \\ &I_{n,1} = 0 \quad \text{für } n \geq 1\,. \end{aligned}} \tag{4.96}$$

4.7.3 Die Knoten mit 1 Kind, das durch eine rechte Kante erreicht wird

Hier lautet die Funktionalgleichung für $R(z,u)$

$$R(z,u) = zuB^*(zu) + 2zR(z,u) + R(z,u)zB^*(z) + B^*(zu)zR(z,u)\,, \tag{4.97}$$

was zu

$$R(z,u) = \frac{zuB^*(zu)}{1 - 2z - zB^*(z) - zB^*(zu)}\,, \tag{4.98}$$

führt.

Daraus erhält man nach einigen Umformungen

$$R(z,u) = \frac{(1+Y)(1-Y)^2}{2(X+Y)(X+1)} = \frac{1-Y^2}{2(X+Y)} - \frac{1-Y^2}{2(X+1)} = \frac{2zu}{X+Y} - \frac{2zu}{X+1}\,. \tag{4.99}$$

Koeffizientenablesen dieses Ausdrucks führt für $2 \leq j \leq n$ zu

$$R_{n,j} = [z^n u^j]R(z,u) = 2[z^{n-1}u^{j-1}]\frac{1}{X+Y} = \frac{1}{n}\binom{2(n-1)}{n-1}\,, \tag{4.100}$$

während für $j = 1$ $R_{n,1} = 0$ gilt.

Das ergibt schon

Satz 4.7.3. *Die Anzahl der Binärbäume $R_{n,j}$ der Größe n, bei denen der entsprechend der Reihenfolge der Postorder-Traversierung markierte Knoten j genau 1, mit einer nach rechts führenden Kante verbundenes Kind besitzt, ist gegeben durch*

$$\boxed{\begin{aligned}&R_{n,j} = \frac{1}{n}\binom{2(n-1)}{n-1} \quad \text{für } 2 \le j \le n\,,\\ &R_{n,1} = 0 \quad \text{für } n \ge 1\,.\end{aligned}} \tag{4.101}$$

4.7.4 Die Knoten mit 1 Kind, das durch eine linke Kante erreicht wird

Man bekommt als Funktionalgleichung für $L(z,u)$

$$L(z,u) = zuB^*(z) + 2zL(z,u) + L(z,u)zB^*(z) + B^*(zu)zL(z,u)\,, \tag{4.102}$$

woraus sich die explizite Darstellung

$$L(z,u) = \frac{zuB^*(zu)}{1 - 2z - zB^*(z) - zB^*(zu)}\,, \tag{4.103}$$

ergibt.

Da also die erzeugenden Funktionen $L(z,u)$ und $R(z,u)$ ident sind, müssen auch $L_{n,j}$ und $R_{n,j}$ gleich sein und man bekommt

Satz 4.7.4. *Die Anzahl der Binärbäume $L_{n,j}$ der Größe n, bei denen der entsprechend der Reihenfolge der Postorder-Traversierung markierte Knoten j genau 1, mit einer nach links führenden Kante verbundenes Kind besitzt, ist gegeben durch*

$$\boxed{\begin{aligned}&L_{n,j} = \frac{1}{n}\binom{2(n-1)}{n-1} \quad \text{für } 2 \le j \le n\,,\\ &L_{n,1} = 0 \quad \text{für } n \ge 1\,.\end{aligned}} \tag{4.104}$$

4.7.5 Der Erwartungswert der Anzahl an Kindern

Bezeichnen wir mit $E_{n,j}$ den Erwartungswert der Anzahl an Kindern des entsprechend der Reihenfolge der Postorder-Traversierung markierten Knotens j in einem Binärbaum der Größe n.

Es ergibt sich

Satz 4.7.5. *Der Erwartungswert $E_{n,j}$ der Anzahl an Kindern des entsprechend der Reihenfolge der Postorder-Traversierung markierten Knotens j in einem Binärbaum der Größe n ist für $1 \le j \le n$ gegeben durch*

$$\boxed{E_{n,j} = 1 - \frac{m(2n+1-2j)(n+2-2j)}{n(2j-1)(n+1-j)} \frac{\binom{2j}{j}\binom{2(n-j)}{n-j}}{\binom{2n}{n}}\,.} \tag{4.105}$$

Kapitel 5

Eine Analyse von Multiple Quickselect

5.1 Einleitung

Der schnellste zum Sortieren eines Datenfeldes verwendete Algorithmus, der am Platz abläuft, also außer dem eigentlichen Datenfeld keinen zusätzlichen Hilfsspeicher benötigt, ist das 1961 von Hoare vorgestellte `Quicksort` (siehe [15]). Aufgrund seiner Effizienz und leichten Implementierbarkeit fand dieser Algorithmus in vielen Bereichen Verwendung und kann heute als Standardsortierprogramm angesehen werden.

Das Prinzip des 'divide and conquer' Algorithmus Quicksort kann wie folgt beschrieben werden. Beim Vorliegen eines Datenfeldes mit n Elementen wird ein Element, das sogenannte Pivotelement, ausgewählt und als Aufteilungselement verwendet. Durch Vergleichen mit den $n-1$ übrigen Elementen wird dieses Pivotelement an die, im sortierten Feld richtige, Position gebracht, während die Elemente, welche kleiner als das Pivotelement sind, links von diesem positioniert werden und die Elemente, welche größer als das Pivotelement sind, rechts von diesem positioniert werden. Nachdem diese Partitionierungsphase abgeschlossen ist, wird Quicksort rekursiv auf die beiden Teilfelder links und rechts des Pivotelements angewendet. Ein leeres Teilfeld beendet natürlich die Rekursion. Detaillierte Beschreibungen bzw. Implementationen dieses Algorithmus findet man beispielsweise in [9] oder [43].

Das durchschnittliche Verhalten des Algorithmus unter dem Modell, das jede Permutation der Elemente von 1 bis n als Eingabefile als gleichwahrscheinlich ansieht, war Gegenstand zahlreicher mathematischer Analysen. Stellvertretend seien hier [23, 42, 39, 40, 6] genannt. Das Hauptaugenmerk der jüngeren Untersuchungen liegt auf der Charakterisierung der Grenzverteilung der Anzahl der Vergleiche in Quicksort.

Gleichzeitig mit Quicksort hat Hoare den Algorithmus `Find` (auch bekannt als `Quickselect`) veröffentlicht, den man verwendet, um das j-größte Element in einem Feld mit n Elementen zu finden (siehe [14]). Die Idee hinter diesem Algorithmus ist die gleiche wie bei Quicksort, allerdings wird nach der Partitionierungsphase nur mehr in höchstens einem Teilfeld links oder rechts vom Pivotelement rekursiv weitergesucht. Bezeichnet k die Position des Aufteilungselementes im vollständig sortierten Feld, dann wird die Suche bei $j < k$ auf das Datenfeld links vom Pivotelement und bei $j > k$ auf das Datenfeld rechts vom Pivotelement eingeschränkt. Ist $j = k$, dann ist das gesuchte Element gefunden, und der Algorithmus terminiert.

Die bei diesem Algorithmus interessanten Parameter sind die Anzahl der Durchläufe (also der rekursiven Aufrufe) und die Anzahl der Vergleiche beim Finden des j-größten unter n in einem Feld gespeicherten Elementen. Auch hier wurde der Algorithmus unter dem Modell der gleichwahrscheinlichen Permutationen eingehend studiert (z. B. in [23, 29, 22, 33]).

Die Idee von Hoare's Find Algorithmus kann erweitert werden zur simultanen Suche von p Elementen mit den Rangordnungen $j_1, j_2, \ldots, j_p$, wobei $1 \leq j_1 < j_2 \leq \cdots < j_p \leq n$ eine feste Menge von p Werten darstellt. Nach der Aufteilungsphase wird die Position des Pivotelements verglichen mit den Rangordnungen j_i, $1 \leq i \leq p$. Je nachdem in welches Intervall die Position des Pivotelements fällt, hat man entweder ein gesuchtes Element gefunden (und) oder man sucht die Elemente, welche eine niedrigere Rangordnung als das Pivotelement haben im linken Teilfeld und die Elemente mit einer größeren Rangordnung im rechten Teilfeld.

Etwas detaillierter läßt sich dieser Algorithmus `Multiple Quickselect` wie folgt beschreiben. Global definiert sollen das n-elementige Datenfeld `A[1..n]` und das Feld `S[1..p]`, welches die sortierten Rangordnungen der p gesuchten Elemente beinhaltet, sein. `MQS` ist eine Prozedur, die vier ganzzahlige Parameter L, R, B und T verlangt. Dabei sollen L und R die linke bzw. rechte Position jenes Teilfeldes von `A` angeben, in denen die Rangordnungen `S[B]` bis `S[T]` gesucht werden.

Beim Aufruf von `MQS(L,R,B,T)` wird nun in der Partitionierungsphase ein Pivotelement aus `A[L..R]` gewählt und an die richtige Position, nennen wir sie k, gebracht, während das entstehende Teilfeld `A[L..k-1]` Elemente kleiner als das Pivotelement und das Teilfeld `A[k+1..R]` Elemente größer als das Pivotelement beinhaltet.

Nun sucht `MQS` in `S[B..T]` die Reihenfolge k des Pivotelements. Wenn k nicht in `S[B..T]` gefunden wurde, so gibt es eine ganze Zahl r, sodaß alle Elemente von `S[B..r]` kleiner als k und alle Elemente von `S[r+1..T]` größer als k sind. Anschließend wird Multiple Quickselect rekursiv mittels `MQS(L,k-1,B,r)` und `MQS(k+1,R,r+1,T)` aufgerufen.

Wenn allerdings k innerhalb von `S[B..T]` an der Position r gefunden wurde, dann wird das gefundene Element `A[k]` ausgegeben und Multiple Quickselect anschließend rekursiv mittels `MQS(L,k-1,B,r-1)` und `MQS(k+1,R,r+1,T)` aufgerufen.

Dadurch werden schließlich alle Elemente mit Rangordnungen in `S[B..T]` ausgegeben. (Natürlich müssen im Programm die Abbruchbedingungen $B > T$ und $L > R$ berücksichtigt werden.)

Bei der Analyse dieses Algorithmus sind vor allem die Anzahl der Durchläufe und der Vergleiche bei bestimmten Reihenfolgenstatistiken interessant. So gibt Prodinger in [37] den Erwartungswert der Anzahl Durchläufe $P[n; j_1, \ldots, j_p]$ und den Erwartungswert der Anzahl der Vergleiche $C[n; j_1, \ldots, j_p]$ von Multiple Quickselect beim Suchen von Elementen mit Rangordnungen $j_1, j_2, \ldots, j_p$ in einem Feld von n Elementen an. Als Korollare erhält er auch die Erwartungswerte der Gesamtdurchschnitte der Anzahl der Durchläufe

$$\mathcal{P}_{n,p} = \frac{1}{\binom{n}{p}} \sum_{1 \leq j_1 < j_2 < \cdots < j_p \leq n} P[n; j_1, \ldots, j_p]$$

bzw. der Vergleiche

$$\mathcal{C}_{n,p} = \frac{1}{\binom{n}{p}} \sum_{1 \le j_1 < j_2 < \cdots < j_p \le n} C[n; j_1, \ldots, j_p]$$

über alle möglichen Auswahlen von p festen Rangordnungen aus den n Elementen. Diesen letztgenannten Erwartungswert für die Vergleiche findet man approximativ auch in [25].

Hier wollen wir nach Anregung von Flajolet mittels erzeugender Funktionen die Analyse der Gesamtdurchschnitte der Anzahl der Durchläufe bzw. der Vergleiche vornehmen. Dadurch erhalten wir auch Formeln für die entsprechenden Varianzen.

Wir werden ausgehen von einer Rekursion für die Wahrscheinlichkeiten $\varphi_{n,p,m}$, daß beim Finden von p zufälligen Elementen in einem Datenfeld der Größe n durch Multiple Quickselect genau m Durchläufe bzw. Vergleiche gemacht werden. Durch Einführen der erzeugenden Funktionen

$$\Phi(z,u,v) = \sum_{n \ge 0} \sum_{p \ge 0} \sum_{m \ge 0} \binom{n}{p} \varphi_{n,p,m} z^n u^p v^m \tag{5.1}$$

übersetzen sich die Rekursionen in Differentialgleichungen.

Für die Berechnung der Erwartungswerte $E_{n,p}$, der zweiten Faktoriellen Momente $M_{n,p}$ bzw. Varianzen $V_{n,p}$ definieren wir die Funktionen

$$G_1(z,u) = \frac{\partial}{\partial v} \Phi(z,u,v) \bigg|_{v=1} \quad \text{und} \quad G_2(z,u) = \frac{\partial^2}{\partial v^2} \Phi(z,u,v) \bigg|_{v=1} . \tag{5.2}$$

Es bestehen dann die Beziehungen

$$E_{n,p} = \frac{1}{\binom{n}{p}} [z^n u^p] \, G_1(z,u) \, , \quad M_{n,p} = \frac{1}{\binom{n}{p}} [z^n u^p] \, G_2(z,u) \;\; \text{und} \;\; V_{n,p} = M_{n,p} + E_{n,p} - E_{n,p}^2 \, . \tag{5.3}$$

Es ist für die Analyse dieses Algorithmus vorteilhaft, wenn man den Binären Suchbaum betrachtet, der mit dem Datenfeld durch die Quicksort Prozedur in einem natürlichen Zusammenhang steht. Rekursiv läßt sich dieser Zusammenhang so beschreiben, daß das gewählte Pivotelement als Wurzel fungiert, an dem links der Binäre Suchbaum der dem Teilfeld links des Pivotelements und rechts der Binäre Suchbaum der dem Teilfeld rechts des Pivotelements entspricht, hängt. Diese Beziehung wird in nachfolgender Abbildung veranschaulicht. Die Teilfelder sind durch Rechtecke dargestellt, während die Pivotelemente eingekreist wurden. Als Pivotelement wurde hier jeweils das linke Element des Datenfeldes gewählt.

Beachtet man diesen Zusammenhang, kann man die aus den Rekursionen erhaltenen Differentialgleichungen auch auf eine andere Art erhalten. Man denkt sich die Koeffizienten $[\frac{z^n}{n!} u^p v^m]$ der Funktionen $\Phi(z,u,v)$ als die Anzahl der Paare aus einem Binären Suchbaum der Größe n und einer Markierung von genau p Knoten, sodaß durch den Multiple Quickselect Algorithmus genau m Durchläufe (bzw. Schlüsselvergleiche) gemacht werden um alle diese markierten Knoten zu finden.

Weiters besitzen diese Größen im Baummodell noch eine andere Interpretation. So werden in Multiple Quickselect beim Partitionierungsschritt in einem Datenfeld der Größe n genau

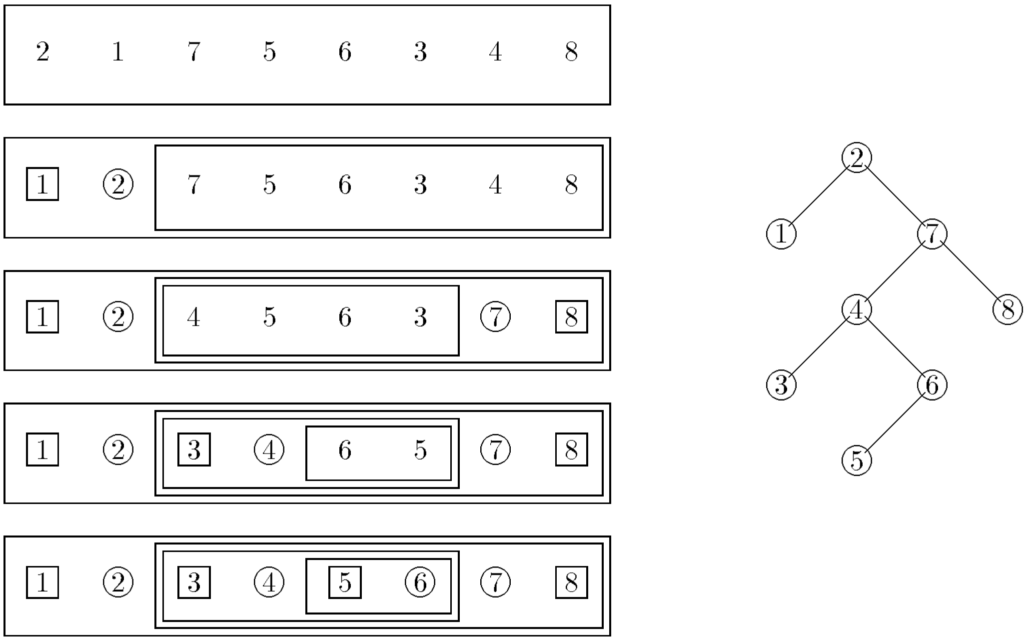


Abbildung 5.1: Der Partitionierungsprozeß des Datenfeldes 2 1 7 5 6 3 4 8 und der entsprechende Binäre Suchbaum.

$n-1$ Vergleiche gemacht, und dies entspricht im Baummodell der Anzahl der echten Nachfolger des Pivotelements. Die Gesamtdurchläufe in Multiple Quickselect beim Suchen von p Elementen lassen sich im Modell der Binären Suchbäume interpretieren als p-Ascendants. Darunter verstehen wir in einem Baum, bei dem p Knoten markiert wurden, die Anzahl der Knoten, die Vorgänger von mindestens einem markierten Knoten sind.

Eine der Methoden die Effizienz von Quicksort bzw. Quickselect noch zu erhöhen besteht darin, daß man das Pivotelement nicht zufällig wählt, sondern das Mittlere von drei gewählten Elementen zur Aufteilung heranzieht (zum Median-of-three Quickselect siehe beispielsweise [21]). Dadurch werden vollständige Entartungen, wo der entsprechende Binäre Suchbaum zu einer linearen Liste degeneriert, vermieden. Dieses Verfahren kann natürlich auch auf Multiple Quickselect angewendet werden, wobei sich nur in der Partitionierungsphase die Aufteilungswahrscheinlichkeiten ändern und die kleinen Datenfelder der Größe $n \leq 2$ nicht mehr rekursiv abgearbeitet werden, die Struktur des Algorithmus aber unverändert bleibt. Bei der mathematischen Analyse dieses Multiple Quickselect mit Median-of-three Partition bekommt man es dann mit Differentialgleichungen höherer Ordnung zu tun. Wir beschränken uns hier auf die Berechnung der Erwartungswerte der Anzahl der Durchläufe bzw. Vergleiche.

Zu Vergleichszwecken berechnen wir die Erwartungswerte der betrachteten Größen auch für das Binärbaummodell (Catalanmodell). Es sollen also hier alle $\frac{1}{n+1}\binom{2n}{n}$ Binärbäume der Größe n gleichwahrscheinlich sein. Die mathematische Analyse gestaltet sich hier etwas einfacher, da man es anstelle von Differentialgleichungen mit Funktionalgleichungen zu tun hat, die etwas leichter zu lösen sind.

5.2 Multiple Quickselect

5.2.1 Die Anzahl der Durchläufe

Bezeichnen wir mit $\varphi_{n,p,m}$ die Wahrscheinlichkeit, daß beim Finden von p zufälligen Elementen in einem Datenfeld der Größe n durch den Multiple Quickselect Algorithmus genau m Durchläufe gemacht werden.

Indem man unterscheidet, ob das Pivotelement ein gesuchtes Element ist (das entspricht im Baummodell einer markierten Wurzel), ergibt sich folgende Rekursion

$$\begin{aligned}\varphi_{n,p,m} = \sum_{k=1}^{n} \frac{1}{n} \Bigg[& \sum_{i=0}^{p} \sum_{l=0}^{m-1} \frac{\binom{k-1}{i}\binom{n-k}{p-i}}{\binom{n}{p}} \varphi_{k-1,i,l}\, \varphi_{n-k,p-i,m-1-l} \\ & + \sum_{i=0}^{p-1} \sum_{l=0}^{m-1} \frac{\binom{k-1}{i}\binom{n-k}{p-i-1}}{\binom{n}{p}} \varphi_{k-1,i,l}\, \varphi_{n-k,p-1-i,m-1-l} \Bigg] \quad \text{für } 1 \le p \le n\,, \\ \varphi_{n,0,m} = & \begin{cases} 1 & \text{wenn } m = 0\,, \\ 0 & \text{wenn } m \neq 0\,. \end{cases}, \quad \varphi_{n,p,m} = 0 \text{ sonst.}\end{aligned} \tag{5.4}$$

Mit der durch Gleichung (5.1) definierten Funktion $\Phi(z,u,v)$ erhält man aus (5.4) durch Multiplikation mit $nz^{n-1}u^p v^m$ und Aufsummieren über alle $n \ge 1$, $p \ge 1$ und $m \ge 0$ die Differentialgleichung

$$\begin{aligned}&\frac{\partial}{\partial z}\Phi(z,u,v) = v(1+u)\Phi(z,u,v)^2 + \frac{1-v}{(1-z)^2}\,, \\ &\Phi(0,u,v) = 1\,.\end{aligned} \tag{5.5}$$

Für die Bestimmung der Erwartungswerte differenzieren wir (5.5) nach v und werten anschließend bei $v = 1$ aus. Dies liefert die lineare Differentialgleichung

$$\begin{aligned}&\frac{\partial}{\partial z}G_1(z,u) = \frac{2(1+u)G_1(z,u)}{1-z(1+u)} + \frac{1+u}{(1-z(1+u))^2} - \frac{1}{(1-z)^2}\,, \\ &G_1(0,u) = 0\,.\end{aligned} \tag{5.6}$$

Dabei verwenden wir die durch (5.2) definierte Funktion $G_1(z,u)$ und die Beziehung

$$\Phi(z,u,1) = \sum_{n\ge 0}\sum_{p\ge 0}\sum_{m\ge 0}\binom{n}{p}\varphi_{n,p,m} z^n u^p = \sum_{n\ge 0}\sum_{p\ge 0}\binom{n}{p} z^n u^p = \frac{1}{1-z(1+u)}\,. \tag{5.7}$$

Lösen von (5.6) führt zu

$$G_1(z,u) = \frac{2u(1+u)}{(1-z(1+u))^2}\log\left(\frac{1}{1-z}\right) + \frac{1+2u}{1-z(1+u)} - \frac{2u}{(1-z(1+u))^2} - \frac{1}{1-z}\,. \tag{5.8}$$

Für das Ablesen der Koeffizienten dieses Ausdrucks, aber vor allem der in den nachfolgenden Abschnitten auftretenden Ausdrücke benötigt man Beziehungen zwischen Vertretern einer Klasse von Funktionen, die sich in ihrer allgemeinsten Form als

$$\frac{1}{(1-z)^{\alpha_1}(1-z(1+u))^{\alpha_2}}\log\left(\frac{1}{1-z}\right)^{\alpha_3}\log\left(\frac{1}{1-z(1+u)}\right)^{\alpha_4}\,,$$

mit $\alpha_i \in \mathbb{N}$ für $1 \le i \le 4$ schreiben läßt und ihren Koeffizienten bei $z^n u^p$. Wir werden hier kurz beschreiben, wie man alle Beziehungen herleiten kann, welche man für das Ablesen der Koeffizienten der bei uns auftretenden Ausdrücke benötigt.

Zunächst betrachten wir Funktionen $F(z,u) = F(z(1+u))$, die aus Funktionen $F(z)$ durch die Variablensubstitution $z \to z(1+u)$ hervorgehen. Zwischen den Koeffizienten $z^n u^p$ von $F(z,u)$ und den Koeffizienten z^n von $F(z)$ besteht dann klarerweise der Zusammenhang

$$[z^n]F(z) = f_n \quad \Rightarrow \quad [z^n u^p]F(z(1+u)) = \sum_{n\ge 0} f_n z^n (1+u)^n = f_n \binom{n}{p}.$$

Dadurch erhalten wir unter Benützung der Formeln (1.19) die wichtigen Beziehungen

$$[z^n u^p]\frac{1}{(1-z(1+u))^{\alpha+1}} = \binom{n+\alpha}{\alpha}\binom{n}{p} \quad \text{mit } \alpha \in \mathbb{N}, \tag{5.9a}$$

$$[z^n u^p]\frac{1}{(1-z(1+u))^{\alpha+1}} \log\left(\frac{1}{1-z(1+u)}\right) = \binom{n+\alpha}{\alpha}(H_{n+\alpha} - H_\alpha)\binom{n}{p} \quad \text{mit } \alpha \in \mathbb{N}, \tag{5.9b}$$

$$\begin{aligned}&[z^n u^p]\frac{1}{(1-z(1+u))^{\alpha+1}} \log^2\left(\frac{1}{1-z(1+u)}\right) \\ &\quad = \binom{n+\alpha}{\alpha}\left((H_{n+\alpha} - H_\alpha)^2 - \left(H^{(2)}_{n+\alpha} - H_{p+\alpha}\right)\right)\binom{n}{p} \quad \text{mit } \alpha \in \mathbb{N}.\end{aligned} \tag{5.9c}$$

Für die Herleitung der weiteren benötigten Beziehungen spaltet man $\frac{1}{1-z(1+u)}$ und $\log\left(\frac{1}{1-z(1+u)}\right)$ auf in $\frac{1}{(1-z)\,(1-\frac{uz}{1-z})}$ bzw. $\log\left(\frac{1}{1-z}\right) + \log\left(\frac{1}{1-\frac{uz}{1-z}}\right)$ und entwickelt. Dies sei exemplarisch an zwei Beispielen ausgeführt.

$$\begin{aligned}&[z^n u^p]\frac{1}{(1-z(1+u))^{\alpha+1}} \log\left(\frac{1}{1-z}\right) = [z^n u^p]\frac{1}{(1-z)^{\alpha+1}\,(1-\frac{uz}{1-z})^{\alpha+1}} \log\left(\frac{1}{1-z}\right) \\ &\quad = [z^n u^p]\sum_{k\ge 0}\binom{k+\alpha}{\alpha}\frac{u^k z^k}{(1-z)^{k+\alpha+1}}\log\left(\frac{1}{1-z}\right) = \binom{p+\alpha}{\alpha}[z^{n-p}]\frac{1}{(1-z)^{p+\alpha+1}}\log\left(\frac{1}{1-z}\right) \\ &\quad = \binom{p+\alpha}{\alpha}\binom{n+\alpha}{p+\alpha}(H_{n+\alpha} - H_{p+\alpha}) \quad \text{mit } \alpha \in \mathbb{N},\end{aligned} \tag{5.9d}$$

$$\begin{aligned}&[z^n u^p]\log\left(\frac{1}{1-z}\right)\log\left(\frac{1}{1-z(1+u)}\right) = [z^n u^p]\log\left(\frac{1}{1-z}\right)\left(\log\left(\frac{1}{1-z}\right) + \log\left(\frac{1}{1-\frac{uz}{1-z}}\right)\right) \\ &\quad = [z^n u^p]\log\left(\frac{1}{1-z}\right)\log\left(\frac{1}{1-\frac{uz}{1-z}}\right) = [z^n u^p]\log\left(\frac{1}{1-z}\right)\sum_{k\ge 1}\frac{z^k u^k}{k(1-z)^k} \\ &\quad = \frac{1}{p}[z^{n-p}]\frac{1}{(1-z)^p}\log\left(\frac{1}{1-z}\right) = \frac{1}{p}\binom{n-1}{p-1}(H_{n-1} - H_{p-1}) \quad \text{für } p \ge 1.\end{aligned} \tag{5.9e}$$

Weitere benötigte Beziehungen seien nachfolgend angeführt.

$$\begin{aligned}&[z^n u^p]\frac{1}{(1-z(1+u))^{\alpha+1}}\log^2\left(\frac{1}{1-z}\right) \\ &\quad = \binom{p+\alpha}{\alpha}\binom{n+\alpha}{p+\alpha}\left((H_{n+\alpha} - H_{p+\alpha})^2 - \left(H^{(2)}_{n+\alpha} - H^{(2)}_{p+\alpha}\right)\right) \quad \text{mit } \alpha \in \mathbb{N},\end{aligned} \tag{5.9f}$$

$$[z^n u^p]\frac{1}{(1-z)\,(1-z(1+u))} = \binom{n+1}{p+1}\,, \tag{5.9g}$$

$$[z^n u^p]\frac{1}{1-z}\log\left(\frac{1}{1-z(1+u)}\right) = \frac{1}{n}\binom{n}{p} \quad \text{für } p \geq 1\,, \tag{5.9h}$$

$$[z^n u^p]\frac{1}{(1-z)\,(1-z(1+u))}\log\left(\frac{1}{1-z}\right) = (H_{n+1} - H_{p+1})\binom{n+1}{p+1}\,, \tag{5.9i}$$

$$[z^n u^p]\frac{1}{(1-z)\,(1-z(1-u))}\log\left(\frac{1}{1-z(1+u)}\right) = \binom{n+1}{p+1}\left(H_{n+1} - \frac{1}{p+1}\right)\,, \tag{5.9j}$$

$$[z^n u^p]\frac{1}{1-z(1+u)}\log\left(\frac{1}{1-z}\right)\log\left(\frac{1}{1-z(1+u)}\right) = \binom{n}{p}\left(H_n^2 - H_nH_p - H_n^{(2)} + H_p^{(2)}\right)\,, \tag{5.9k}$$

$$\begin{aligned}
&[z^n u^p]\frac{1}{(1-z(1+u))^2}\log\left(\frac{1}{1-z}\right)\log\left(\frac{1}{1-z(1+u)}\right) \\
&\quad = \Bigg[(n+1)H_n^2 - (n+1)H_nH_p - \frac{np+2n-p}{p+1}H_n + nH_p - (n+1)H_n^{(2)} + (n+1)H_p^{(2)} \\
&\qquad - \frac{(p+2)(n-p)}{(p+1)^2}\Bigg]\binom{n}{p}\,,
\end{aligned} \tag{5.9l}$$

$$\begin{aligned}
&[z^n u^p]\frac{1}{(1-z(1+u))^3}\log\left(\frac{1}{1-z}\right)\log\left(\frac{1}{1-z(1+u)}\right) \\
&\quad = \Bigg[(H_{n+2} - H_{p+2})\left(\binom{p+2}{2}H_p - (p+\frac{3}{2})p + \frac{1}{2}\binom{p+1}{2}\right) \\
&\qquad + \binom{p+2}{2}\left((H_{n+2} - H_{p+2})^2 - (H_{n+2}^{(2)} - H_{p+2}^{(2)})\right)\Bigg]\binom{n+2}{p+2}\,.
\end{aligned} \tag{5.9m}$$

Liest man nun von (5.8) die Koeffizienten ab, erhält man unter Verwendung der Beziehungen (5.9) schon

Satz 5.2.1. *Der Erwartungswert $E_{n,p}$ der Anzahl der Durchläufe beim Finden von p Elementen in einem Feld der Größe n durch Multiple Quickselect ist gegeben durch*

$$\boxed{\begin{aligned} E_{n,p} = {} & \frac{2(n+1)^2p}{(n+1-p)(n+2-p)}H_n - \frac{2(n+1)^2p}{(n+1-p)(n+2-p)}H_p \\ & - \frac{2n^2p - n^2 - 2np^2 + 2np - n - p^2 - p}{(n+1-p)(n+2-p)} \quad \text{für } p \geq 1\,, \end{aligned}}$$
$$\boxed{E_{n,0} = 0\,.} \tag{5.10}$$

Besonders angeführt seien die Fälle $p = 1$ (Hoare's Find Algorithmus) und $p = n$ (Quicksort):

Korollar 5.2.2.

$$\boxed{E_{n,1} = \frac{2(n+1)}{n}H_n - 3\,,} \tag{5.11a}$$

$$\boxed{E_{n,n} = n\,.} \tag{5.11b}$$

Für den Erhalt der zweiten Faktoriellen Momente und vor allem der Varianz differenzieren wir (5.5) zweimal nach v und werten anschließend bei $v = 1$ aus. Dies liefert unter Benützung der Beziehungen (5.2) und (5.7) die lineare Differentialgleichung

$$\begin{aligned}
\frac{\partial}{\partial z} G_2(z,u) &= \frac{2(1+u)G_2(z,u)}{1-z(1+u)} + \frac{8u^2(1+u)^3}{(1-z(1+u))^4}\log^2\left(\frac{1}{1-z}\right) \\
&+ \frac{8u(1+u)^2(z^2u^2-2zu^2+2z^2u-2zu+1-2z+z^2}{(1-z)(1-z(1+u))^4}\log\left(\frac{1}{1-z}\right) \\
&+ \frac{2zu(1+u)(-1+z-2u+zu)(z^2u^2+3z^2u+2z^2-2zu^2-3zu-4z+2)}{(1-z)^2(1-z(1+u))^4}, \\
G_2(0,u) &= 0\,. \qquad (5.12)
\end{aligned}$$

Diese Differentialgleichung besitzt die Lösung

$$\begin{aligned}
G_2(z,u) = &-\frac{4u(1+u)^2(-1+z-2u+zu)}{(1-z(1+u))^3}\log^2\left(\frac{1}{1-z}\right) \\
&- \frac{4u(1+u)(zu^2+3u+1-z)}{(1-z(1+u))^3}\log\left(\frac{1}{1-z}\right) + \frac{2(1+u)}{1-z} + \frac{8u^2}{(1-z(1+u))^3} \\
&- \frac{2u(-3+2u)}{(1-z(1+u))^2} - \frac{2(1+4u+2u^2)}{1-z(1+u)}\,. \qquad (5.13)
\end{aligned}$$

Das Ablesen der Koeffizienten von (5.13) gelingt leicht unter Verwendung der Beziehungen (5.9) und man erhält

Hilfssatz 5.2.3. *Das zweite Faktorielle Moment $M_{n,p}$ der Anzahl der Durchläufe beim Finden von p Elementen in einem Feld der Größe n durch Multiple Quickselect ist gegeben durch*

$$\boxed{\begin{aligned}
M_{n,p} = &\frac{4(n+1)^2(n+2)(np+p+2)p}{(n-p+4)^{\underline{4}}}\Big((H_n-H_p)^2 - (H_n^{(2)} - H_p^{(2)})\Big) \\
&- \frac{4(n+1)\,r_1(n,p)}{(n-p+4)^{\underline{4}}}(H_n-H_p) + \frac{2\,r_2(n,p)}{(n-p+4)^{\underline{4}}} \quad \text{für } p \ge 2\,,
\end{aligned}}$$

$$\boxed{M_{n,1} = \frac{4(n+1)}{n}H_n^2 - \frac{4(3n+1)}{n}H_n - \frac{4(n+1)}{n}H_n^{(2)} + 16\,,} \qquad (5.14)$$

mit

$$\begin{aligned}
r_1(n,p) &= 5n^2p^2 + 5n^2p - 4n^2 + 18np - 12n - 3np^2 - 6p^2 + 8p - 8 + 2n^3p^2 - 2n^2p^3 \\
&\quad - n^3p - 4np^3 - 2p^3\,, \\
r_2(n,p) &= n^4p + n^3p^2 + 30n^3p - 3n^2p^3 - 49n^2p^2 + 103n^2p + np^4 + 20np^3 - 98np^2 + 98np \\
&\quad - n^4 - 18n^3 - 47n^2 + 13p^3 - 39p^2 + 2n^2p^4 - 4n^3p^3 + 2n^4p^2 - 30n + 18p\,.
\end{aligned}$$

Die Varianz $V_{n,p}$ erhalten wir aus den vorhergehenden Resultaten leicht mit Beziehung (5.3). Es ergibt sich

Satz 5.2.4. *Die Varianz $V_{n,p}$ der Anzahl der Durchläufe beim Finden von p Elementen in einem Feld der Größe n durch Multiple Quickselect ist gegeben durch*

$$\boxed{\begin{aligned}V_{n,p} = &-\frac{4(n+1)^2 p\, r_1(n,p)}{(n-p+4)^{\underline{4}}(n-p+2)^{\underline{2}}}(H_n-H_p)^2 + \frac{2(n+1)\, r_2(n,p)}{(n-p+4)^{\underline{4}}(n-p+2)^{\underline{2}}}(H_n-H_p)\\ &-\frac{4(n+1)^2(n+2)(np+p+2)p}{(n-p+4)^{\underline{4}}}(H_n^{(2)}-H_p^{(2)}) + \frac{2(n+1)(n-p)\, r_3(n,p)}{(n-p+4)^{\underline{4}}(n-p+2)^{\underline{2}}}\,,\end{aligned}}$$
$$\boxed{V_{n,1} = -\frac{4(n+1)}{n^2}H_n^2 + \frac{2(n+5)}{n}H_n - \frac{4(n+1)}{n}H_n^{(2)} + 4\,.} \tag{5.15}$$

mit

$$\begin{aligned}
r_1(n,p) &= 3n^3p - 2n^3 - 2n^2p^2 + 18n^2p - 10n^2 - np^3 - 5np^2 + 33np - 16n\\
&\quad - p^3 - 5p^2 + 20p - 8\,,\\
r_2(n,p) &= n^5p + 10n^4p^2 - 11n^4p + 8n^4 - 20n^3p^3 + 76n^3p^2 - 87n^3p + 48n^3 + 6n^2p^4\\
&\quad - 76n^2p^3 + 160n^2p^2 - 193n^2p + 104n^2 + 3np^5 + 8np^4 - 83np^3 + 116np^2\\
&\quad - 174np + 96n + 3p^5 + 2p^4 - 19p^3 + 14p^2 - 56p + 32\,,\\
r_3(n,p) &= 2n^4p - n^4 - 12n^3p^2 + 38n^3p - 19n^3 + 18n^2p^3 - 98n^2p^2 + 173n^2p - 75n^2\\
&\quad - 8np^4 + 72np^3 - 215np^2 + 280np - 105n - 11p^4 + 65p^3 - 145p^2 + 151p - 48\,.
\end{aligned}$$

Für $p = n$ (Quicksort) ergibt sich insbesondere

Korollar 5.2.5.

$$\boxed{V_{n,n} = 0\,.}$$

5.2.2 Die Anzahl der Vergleiche

Bezeichnen wir nun mit $\varphi_{n,p,m}$ die Wahrscheinlichkeit, daß beim Finden von p zufälligen Elementen in einem Datenfeld der Größe n durch den Multiple Quickselect Algorithmus genau m Vergleiche gemacht werden.

Indem man unterscheidet, ob das Pivotelement eines der p gesuchten Element darstellt oder nicht bzw. im Baummodell ob die Wurzel einer der p markierten Knoten ist, ergibt sich aufgrund der $n-1$ Schlüsselvergleiche im Partitionierungsschritt folgende Rekursion für $\varphi_{n,p,m}$:

$$\begin{aligned}
\varphi_{n,p,m} = \sum_{k=1}^{n}\frac{1}{n}\Bigg[&\sum_{i=0}^{p}\sum_{l=0}^{m-n+1}\frac{\binom{k-1}{i}\binom{n-k}{p-i}}{\binom{n}{p}}\varphi_{k-1,i,l}\,\varphi_{n-k,p-i,m-n+1-l}\\
&+\sum_{i=0}^{p-1}\sum_{l=0}^{m-n+1}\frac{\binom{k-1}{i}\binom{n-k}{p-i-1}}{\binom{n}{p}}\varphi_{k-1,i,l}\,\varphi_{n-k,p-1-i,m-n+1-l}\Bigg] \quad \text{für } 1 \le p \le n\,,\\
\varphi_{n,0,m} = \begin{cases}1 & \text{wenn } m = 0\,,\\ 0 & \text{wenn } m \neq 0\,.\end{cases}&\,,\quad \varphi_{n,p,m} = 0 \text{ sonst.}
\end{aligned} \tag{5.16}$$

Mit der durch Gleichung (5.1) definierten Funktion $\Phi(z,u,v)$ ergibt sich aus (5.16) durch Multiplikation mit $nz^{n-1}u^p v^m$ und Aufsummieren über alle $n \geq 1$, $p \geq 1$ und $m \geq 0$ die Differentialgleichung

$$\begin{aligned} &\frac{\partial}{\partial z}\Phi(z,u,v) = (1+u)\Phi^2(zv,u,v) + \frac{1}{(1-z)^2} - \frac{1}{(1-zv)^2}\,, \\ &\Phi(0,u,v) = 1\,. \end{aligned} \tag{5.17}$$

Um zu den Erwartungswerten zu gelangen, leiten wir zunächst (5.17) nach v ab und evaluieren anschließend bei $v = 1$. Unter Verwendung von (5.2) und (5.7) ergibt dies folgende lineare Differentialgleichung erster Ordnung für $G_1(z,u)$

$$\begin{aligned} &\frac{\partial}{\partial z}G_1(z,u) - \frac{2(1+u)}{1-z(1+u)}G_1(z,u) = \frac{2z(1+u)^2}{(1-z(1+u))^3} - \frac{2z}{(1-z)^3}\,, \\ &G_1(0,u) = 0\,. \end{aligned} \tag{5.18}$$

Diese Differentialgleichung besitzt die Lösung

$$\begin{aligned} G_1(z,u) = &-\frac{2(1+u)(1+3u)}{(1-z(1+u))^2}\log\left(\frac{1}{1-z}\right) + \frac{2}{(1-z(1+u))^2}\log\left(\frac{1}{1-uz}\right) \\ &+ \frac{2(1+2u)}{u(1-z)} - \frac{1}{(1-z)^2} - \frac{2(1+3u+3u^2)}{u(1-z(1+u))} + \frac{3(1+2u)}{(1-z(1+u))^2}\,. \end{aligned} \tag{5.19}$$

Das Ablesen der Koeffizienten ergibt unter Verwendung der Beziehungen (5.9) sofort

Satz 5.2.6. *Der Erwartungswert $E_{n,p}$ der Anzahl der Vergleiche beim Finden von p Elementen in einem Feld der Größe n durch Multiple Quickselect ist gegeben durch*

$$\boxed{\begin{aligned} E_{n,p} = &-\frac{2p(n+1)(4n-p+5)}{(n+2-p)(n+1-p)}H_{n+1} + \frac{2(n+1)^2(n+2p+2)}{(n+2-p)(n+1-p)}H_p \\ &+ \frac{n^2+(5p-1)n+3p}{n+2-p} \quad \text{für } p \geq 1\,, \end{aligned}}$$
$$\boxed{E_{n,0} = 0\,.} \tag{5.20}$$

Von besonderem Interesse sind wiederum die Fälle $p = 1$ (Hoare's Find Algorithmus) und $p = n$ (Quicksort). Für diese ergibt sich

Korollar 5.2.7.

$$\boxed{E_{n,1} = -\frac{8(n+1)}{n}H_n + 3n + 13\,,} \tag{5.21a}$$

$$\boxed{E_{n,n} = 2(n+1)H_n - 4n\,.} \tag{5.21b}$$

Zur Berechnung des zweiten Faktoriellen Moments und insbesondere der Varianz differenzieren wir (5.17) zweimal nach v und werten an der Stelle $v = 1$ aus. Dies liefert unter

Berücksichtigung der Beziehungen (5.2) und (5.7) folgende lineare Differentialgleichung erster Ordnung für $G_2(z,u)$

$$
\begin{aligned}
&\frac{\partial}{\partial z} G_2(z,u) - \frac{2(1+u)}{1-z(1+u)} G_2(z,u) \\
&\quad = \frac{8(1+3u)^2(1+u)^3}{(1-z(1+u))^4} \log^2\left(\frac{1}{1-z}\right) - \frac{16(1+3u)(1+u)^2}{(1-z(1+u))^4} \log\left(\frac{1}{1-z}\right) \log\left(\frac{1}{1-z(1+u)}\right) \\
&\qquad + \frac{8(1+u)}{(1-z(1+u))^4} \log^2\left(\frac{1}{1-z(1+u)}\right) \\
&\qquad - \frac{8z(1+3u)(1+u)^2(6u^2-9zu^2+2z^2u^2-14zu+5z^2u+9u-6z+3+3z^2)}{(1-z)^2(1-z(1+u))^4} \times \\
&\qquad\quad \times \log\left(\frac{1}{1-z}\right) \\
&\qquad + \frac{8z(1+u)(6u^2-9zu^2+2z^2u^2-14zu+5z^2u+9u-6z+3+3z^2)}{(1-z)^2(1-z(1+u))^4} \times \\
&\qquad\quad \times \log\left(\frac{1}{1-z(1+u)}\right) \\
&\qquad + \Big[2z^2u(53+174z^2+592z^2u+766z^2u^2-164z+165u+5z^4-534zu-700zu^2 \\
&\qquad\quad + 223u^2-108zu^4-438u^3z+453z^2u^3+105z^2u^4-68z^3+144u^3-316z^3u^2 \\
&\qquad\quad - 242z^3u-178z^3u^3+17z^4u^3+27z^4u^2+19z^4u+4z^4u^4-36z^3u^4+36u^4)\Big] \times \\
&\qquad\quad \times \frac{1}{(1-z)^4(1-z(1+u))^4} \\
&G_2(0,u) = 0\,.
\end{aligned}
\tag{5.22}
$$

Mittels Variation der Konstanten erhält man daraus die Lösung

$$
\begin{aligned}
G_2(z,u) = {} & \frac{24(1+u)^3}{(1-z(1+u))^2} \int_0^z \frac{1}{1-t(1+u)} \log\left(\frac{1}{1-t}\right) dt \\
& + \frac{16(1+u)(uz-3u+z-2)}{(1-z(1+u))^3} \log\left(\frac{1}{1-z}\right) \log\left(\frac{1}{1-z(1+u)}\right) \\
& + \frac{4(z+uz+1)}{(1-z(1+u))^3} \log^2\left(\frac{1}{1-z(1+u)}\right) \\
& - \frac{8(1+3u)(1+u)^2(uz-3u+z-2)}{(1-z(1+u))^3} \log^2\left(\frac{1}{1-z}\right) \\
& - \frac{8(8z^2u^3+16u^2z^2+z^2+9uz^2-20zu^2-9u^3z-14uz-2z+3u^2+1+5u)}{(1-z)(1-z(1+u))^3u} \times \\
& \quad \times \log\left(\frac{1}{1-z(1+u)}\right) \\
& + \Big[4(1+u)(10z^2u^4-16zu^4-19u^3z-20u^3+31z^2u^3+35u^2z^2-30zu^2-7u^2 \\
& \qquad + 16uz^2-24uz+8u+2z^2-4z+2)\Big] \frac{1}{(1-z)(1-z(1+u))^3u} \log\left(\frac{1}{1-z}\right)
\end{aligned}
$$

$$
\begin{aligned}
&+\Big[\frac{2}{3}z(-36z-237uz-519zu^2+12u^4z^4+12-12z^3+90u+198u^2+120u^3\\
&\quad -330u^3z-12zu^4+36z^2+478u^2z^2+219uz^2+346z^2u^3-173u^3z^3\\
&\quad -199u^2z^3-87uz^3-49z^3u^4+48z^2u^4+15uz^4+42u^2z^4+39u^3z^4)\Big]\times\\
&\quad\times\frac{1}{(1-z)^3(1-z(1+u))^3}\,. \qquad (5.23)
\end{aligned}
$$

bzw. durch partielle Integration in Hinblick auf ein leichteres Ablesen der Koeffizienten

$$
\begin{aligned}
G_2(z,u) = &-\frac{24(1+u)^2}{(1-z(1+u))^2}\int_0^z \frac{1}{1-t}\log\left(\frac{1}{1-t(1+u)}\right)dt\\
&-\frac{8(1+u)(1+3u)(1+z+uz)}{(1-z(1+u))^3}\log\left(\frac{1}{1-z}\right)\log\left(\frac{1}{1-z(1+u)}\right)\\
&+\frac{4(1+z+uz)}{(1-z(1+u))^3}\log^2\left(\frac{1}{1-z(1+u)}\right)\\
&-\frac{8(1+3u)(1+u)^2(uz-3u+z-2)}{(1-z(1+u))^3}\log^2\left(\frac{1}{1-z}\right)\\
&-\frac{8(8u^3z^2+16u^2z^2+z^2+9uz^2-20u^2z-9u^3z-14uz-2z+3u^2+1+5u)}{u(1-z)(1-z(1+u))^3}\times\\
&\quad\times\log\left(\frac{1}{1-z(1+u)}\right)\\
&+\Big[4(1+u)(10u^4z^2-16zu^4-19u^3z-20u^3+31u^3z^2+35u^2z^2-30u^2z-7u^2\\
&\quad +16uz^2-24uz+8u+2z^2-4z+2)\Big]\frac{1}{u(1-z)(1-z(1+u))^3}\log\left(\frac{1}{1-z}\right)\\
&+\Big[\frac{2}{3}z(-36z-237uz-519u^2z+12u^4z^4+12-12z^3+90u+198u^2+120u^3\\
&\quad -330u^3z-12zu^4+36z^2+478u^2z^2+219uz^2+346u^3z^2-173u^3z^3\\
&\quad -199u^2z^3-87uz^3-49u^4z^3+48u^4z^2+15uz^4+42u^2z^4+39u^3z^4)\Big]\times\\
&\quad\times\frac{1}{(1-z)^3(1-z(1+u))^3}\,. \qquad (5.24)
\end{aligned}
$$

Für das Ablesen der Koeffizienten von (5.24) benötigt man neben den Gleichungen (5.9) noch eine Beziehung für den nicht geschlossen integrierbaren Bestandteil dieses Ausdrucks. Er lautet

$$
\begin{aligned}
&[z^nu^p]\frac{1}{(1-z(1+u))^2}\int_0^z\frac{1}{1-t}\log\left(\frac{1}{1-t(1+u)}\right)dt\\
&=(n+1)\sum_{k=1}^{n}\frac{1}{k^2}\binom{n-k}{p}+\Big[\frac{n+1}{2}\left(H_n^2-2H_nH_p+H_p^2+3H_p^{(2)}-3H_n^{(2)}\right)\\
&\quad -\frac{n-p}{p+1}H_n+\frac{n-p}{(p+1)^2}\Big]\binom{n}{p}\,. \qquad (5.25)
\end{aligned}
$$

In nachfolgend angeführten Ergebnissen für die zweiten Faktoriellen Momente bzw. Varianzen tritt also eine Summe auf, die nicht in geschlossene Form zu bringen ist. Schreiben wir für diese Summe abkürzend $\lambda_{n,p}$, dann ist sie gegeben durch

$$\lambda_{n,p} = \frac{1}{\binom{n}{p}} \sum_{k=1}^{n} \frac{1}{k^2} \binom{n-k}{p} = \frac{1}{\binom{n}{p}} [z^n] \frac{z^p}{(1-z)^{p+1}} \, \mathrm{L}_2(z) \,,$$

wobei im letzten Ausdruck die durch (1.94) definierten Polylogarithmen $\mathrm{L}_m(z)$ verwendet wurden.

Für die asymptotischen Entwicklungen werden wir noch Beziehungen dieser Summe zu vertrauteren Ausdrücken herstellen.

Mit dieser Schreibweise liefert das Koeffizientenablesen von (5.24)

Hilfssatz 5.2.8. *Das zweite Faktorielle Moment $M_{n,p}$ der Anzahl der Vergleiche beim Finden von p Elementen in einem Feld der Größe n durch Multiple Quickselect ist gegeben durch*

$$\boxed{\begin{aligned} M_{n,p} = & -\frac{24(n+1)^2(n+2)}{(n+1-p)(n+2-p)} \lambda_{n+2,p} + \frac{4(n+1)^2 p\, \Psi_1(n,p)}{(n+4-p)^{\underline{4}}} H_n^2 \\ & + \frac{8(n+1)^2 p\, \Psi_2(n,p)}{(n+4-p)^{\underline{4}}} H_n H_p + \frac{4(n+2)(n+1)^2\, \Psi_3(n,p)}{(n+4-p)^{\underline{4}}} H_p^2 \\ & - \frac{4(n+1)\, \Psi_4(n,p)}{(n+4-p)^{\underline{4}}} H_n - \frac{4(n+1)\, \Psi_5(n,p)}{(n+4-p)^{\underline{4}}} H_p \\ & - \frac{4(n+1)^2\, \Psi_6(n,p)}{(n+4-p)^{\underline{4}}} H_n^{(2)} - \frac{4(n+1)^2\, \Psi_7(n,p)}{(n+4-p)^{\underline{4}}} H_p^{(2)} \\ & + \frac{2}{3} \frac{\Psi_8(n,p)}{(p+1)^2(p+2)(n+4-p)^{\underline{4}}} \quad \text{für } p \geq 2 \,, \end{aligned}}$$

$$\boxed{M_{n,1} = \frac{48(n+1)}{n} H_n^2 - \frac{4(12n^2+35n-1)}{n} H_n - \frac{48(n+1)}{n} H_n^{(2)} + 10n^2 + 65n + 109 \,,}$$

$$\boxed{M_{n,0} = 0 \,,} \tag{5.26}$$

wobei die Polynome $\Psi_i(n,p)$, $1 \leq i \leq 8$ gegeben sind durch

$$\begin{aligned} \Psi_1(n,p) &= 16n^2p - 4n^2 - 8np^2 + 69np - n + p^3 - 14p^2 + 61p + 24 \,, \\ \Psi_2(n,p) &= -4n^3 - 7n^2p - 29n^2 + 2np^2 - 24np - 82n + 2p^2 - 14p - 84 \,, \\ \Psi_3(n,p) &= n^3 + 4n^2p + 8n^2 + 4np^2 + 24np + 19n + p^2 + 47p + 12 \,, \\ \Psi_4(n,p) &= 4n^4p + 15n^3p^2 - 8n^3p + 14n^3 - 24n^2p^3 + 65n^2p^2 - 91n^2p + 58n^2 + 5np^4 \\ &\quad - 59np^3 + 30np^2 - 167np + 68n + 2p^4 - 17p^3 - 47p^2 - 124p + 24 \,, \\ \Psi_5(n,p) &= -n^5 - 6n^4p - 6n^4 - 3n^3p^2 - 18n^3p - 15n^3 + 10n^2p^3 + 7n^2p^2 + 11n^2p - 16n^2 \\ &\quad + 17np^3 + 61np^2 + 127np - 6n + 7p^3 + 45p^2 + 158p \,, \\ \Psi_6(n,p) &= -6n^3 + 16n^2p^2 + 8n^2p - 54n^2 - 8np^3 + 63np^2 + 65np - 156n + p^4 - 14p^3 \\ &\quad + 49p^2 + 108p - 144 \,, \end{aligned}$$

$$
\begin{aligned}
\Psi_7(n,p) &= n^4 - 4n^3p + 16n^3 - 10n^2p^2 - 38n^2p + 89n^2 + 4np^3 - 33np^2 - 135np + 206n \\
&\quad + 4p^3 - 14p^2 - 158p + 168\,, \\
\Psi_8(n,p) &= 12n^6p^3 + 46n^6p^2 + 48n^6p + 2n^6 - 30n^5p^4 + 44n^5p^3 + 486n^5p^2 + 634n^5p + 30n^5 \\
&\quad + 72n^4p^5 - 168n^4p^4 - 455n^4p^3 + 1760n^4p^2 + 3563n^4p + 388n^4 - 102n^3p^6 \\
&\quad + 116n^3p^5 - 330n^3p^4 - 2068n^3p^3 + 3268n^3p^2 + 9998n^3p + 1490n^3 + 48n^2p^7 \\
&\quad - 80n^2p^6 + 36n^2p^5 - 1172n^2p^4 - 4069n^2p^3 + 4534n^2p^2 + 16009n^2p + 2754n^2 \\
&\quad + 42np^7 + 212np^6 + 672np^5 - 1744np^4 - 4886np^3 + 4816np^2 + 15536np \\
&\quad + 3352n + 3p^7 + 172p^6 + 644p^5 - 710p^4 - 2551p^3 + 2194p^2 + 7160p + 2160\,.
\end{aligned}
$$

Aus den zuletzt errechneten Resultaten für den Erwartungswert bzw. das zweite Faktorielle Moment bekommen wir mit Beziehung (5.3) unmittelbar

Satz 5.2.9. *Die Varianz $V_{n,p}$ der Anzahl der Vergleiche beim Finden von p Elementen in einem Feld der Größe n durch Multiple Quickselect ist gegeben durch*

$$
\boxed{
\begin{aligned}
V_{n,p} = &-\frac{24(n+1)^2(n+2)}{(n+1-p)(n+2-p)}\lambda_{n+2,p} + \frac{4(n+1)^2p\,\Psi_1(n,p)}{(n+2-p)^{\underline{2}}(n+4-p)^{\underline{4}}}(H_n-H_p)^2 \\
&- \frac{2(n+1)\,\Psi_2(n,p)}{(n+2-p)^{\underline{2}}(n+4-p)^{\underline{4}}}H_n - \frac{2(n+1)\,\Psi_3(n,p)}{(n+2-p)^{\underline{2}}(n+4-p)^{\underline{4}}}H_p \\
&- \frac{4(n+1)^2\,\Psi_4(n,p)}{(n+4-p)^{\underline{4}}}H_n^{(2)} - \frac{4(n+1)^2\,\Psi_5(n,p)}{(n+4-p)^{\underline{4}}}H_p^{(2)} \\
&+ \frac{1}{3}\frac{\Psi_6(n,p)}{(p+1)^2(p+2)(n+2-p)^{\underline{2}}(n+4-p)^{\underline{4}}} \quad \text{für } p \ge 2\,,
\end{aligned}}
$$

$$
\boxed{V_{n,1} = n^2 - 10n - \frac{16(n+1)(n+4)}{n^2}H_n^2 + \frac{12(9n+17)}{n}H_n - \frac{48(n+1)}{n}H_n^{(2)} - 47\,,}
$$

$$
\boxed{V_{n,0} = 0\,,} \tag{5.27}
$$

wobei die Polynome $\Psi_i(n,p)$, $1 \le i \le 6$ gegeben sind durch

$$
\begin{aligned}
\Psi_1(n,p) &= -4n^4 - 27n^3p - 13n^3 + 30n^2p^2 - 183n^2p + 13n^2 + np^3 + 108np^2 - 379np \\
&\quad + 70n - 2p^3 + 108p^2 - 250p + 48\,, \\
\Psi_2(n,p) &= -54n^5p + 28n^5 + 29n^4p^2 - 383n^4p + 200n^4 + 92n^3p^3 + 202n^3p^2 - 1092n^3p \\
&\quad + 540n^3 - 54n^2p^4 + 252n^2p^3 + 937n^2p^2 - 1693n^2p + 688n^2 - 14np^5 - 78np^4 \\
&\quad - 50np^3 + 1820np^2 - 1450np + 416n + p^6 - 21p^5 + 5p^4 - 347p^3 + 1170p^2 \\
&\quad - 520p + 96\,, \\
\Psi_3(n,p) &= n^6 + 48n^5p - 15n^5 - 14n^4p^2 + 318n^4p - 133n^4 - 112n^3p^3 - 72n^3p^2 + 824n^3p \\
&\quad - 365n^3 + 69n^2p^4 - 382n^2p^3 - 535n^2p^2 + 1168n^2p - 444n^2 + 8np^5 + 143np^4 \\
&\quad - 218np^3 - 1295np^2 + 962np - 244n + 8p^5 + 62p^4 + 172p^3 - 926p^2 + 348p - 48\,, \\
\Psi_4(n,p) &= -6n^3 + 16n^2p^2 + 8n^2p - 54n^2 - 8np^3 + 63np^2 + 65np - 156n + p^4 - 14p^3 \\
&\quad + 49p^2 + 108p - 144\,,
\end{aligned}
$$

$$\begin{aligned}
\Psi_5(n,p) &= n^4 - 4n^3p + 16n^3 - 10n^2p^2 - 38n^2p + 89n^2 + 4np^3 - 33np^2 - 135np + 206n \\
&\quad + 4p^3 - 14p^2 - 158p + 168\,, \\
\Psi_6(n,p) &= 21n^8p^3 + 80n^8p^2 + 81n^8p - 2n^8 - 126n^7p^4 - 114n^7p^3 + 894n^7p^2 + 1422n^7p \\
&\quad + 36n^7 + 315n^6p^5 - 717n^6p^4 - 3199n^6p^3 + 3089n^6p^2 + 10684n^6p + 964n^6 \\
&\quad - 420n^5p^6 + 2123n^5p^5 + 1121n^5p^4 - 15987n^5p^3 + 387n^5p^2 + 42412n^5p + 5716n^5 \\
&\quad + 315n^4p^7 - 2166n^4p^6 + 4763n^4p^5 + 12391n^4p^4 - 37633n^4p^3 - 22649n^4p^2 \\
&\quad + 100085n^4p + 16378n^4 - 126n^3p^8 + 960n^3p^7 - 4892n^3p^6 + 5177n^3p^5 \\
&\quad + 28939n^3p^4 - 55751n^3p^3 - 67233n^3p^2 + 149358n^3p + 29080n^3 + 21n^2p^9 \\
&\quad - 205n^2p^8 + 1075n^2p^7 - 7036n^2p^6 + 4467n^2p^5 + 34141n^2p^4 - 63881n^2p^3 \\
&\quad - 102176n^2p^2 + 140134n^2p + 35076n^2 + 39np^9 + 157np^8 + 656np^7 - 7286np^6 \\
&\quad + 4747np^5 + 28013np^4 - 49394np^3 - 88868np^2 + 73328np + 26224n + 290p^8 \\
&\quad + 406p^7 - 3864p^6 + 1764p^5 + 13338p^4 - 15114p^3 - 33956p^2 + 14672p + 8640\,.
\end{aligned}$$

Besonders erwähnt werden soll noch der Fall $p = n$:

Korollar 5.2.10.

$$\boxed{V_{n,n} = (7n+13)n - 2(n+1)H_n - 4(n+1)^2H_n^{(2)}\,.} \tag{5.28}$$

5.2.3 Asymptotische Betrachtungen

Wir untersuchen nun das Verhalten der Größen dieses Abschnittes für $n \to \infty$ und unterscheiden dabei die Fälle p fest (also nahe dem klassischen Find-Algorithmus mit $p = 1$) , $l = n - p$ fest (also nahe Quicksort) und $p = \rho n$ mit $0 < \rho < 1$.

Unter Verwendung der Entwicklungen der Harmonischen Zahlen (1.92) erhalten wir aus (5.10) für die Erwartungswerte der Anzahl der Durchläufe

Korollar 5.2.11. *Der Erwartungswert $E_{n,p}$ der Anzahl der Durchläufe beim Finden von p Elementen in einem Feld der Größe n durch Multiple Quickselect ist asymptotisch gegeben durch*

$$\boxed{E_{n,p} = 2p\log n + 2\gamma p - 2p - 2pH_p + 1 + \mathcal{O}\left(\frac{\log n}{n}\right) \quad \text{für } p \text{ fix}\,,} \tag{5.29a}$$

$$\boxed{E_{n,p} = n - \frac{1}{3}l + \mathcal{O}\left(\frac{\log n}{n}\right) \quad \text{für } l = n - p \text{ fix}\,,} \tag{5.29b}$$

$$\boxed{\begin{aligned} E_{n,p} = -\frac{2\rho(1-\rho+\log\rho)}{(1-\rho)^2}n + \frac{\rho(4\rho\log\rho + 2\log\rho - \rho^2 - 4\rho + 5)}{(1-\rho)^3} + \mathcal{O}\left(\frac{\log n}{n}\right) \\ \text{für } p = \rho n \text{ und } 0 < \rho < 1\,, \end{aligned}} \tag{5.29c}$$

bzw. aus (5.15) für die Varianz der Anzahl der Durchläufe

Korollar 5.2.12. *Die Varianz $V_{n,p}$ der Anzahl der Durchläufe beim Finden von p Elementen in einem Feld der Größe n durch Multiple Quickselect ist asymptotisch gegeben durch*

$$\boxed{V_{n,p} = 2p\log n + 4p^2 H_p^{(2)} - 2pH_p + 2p\gamma - \frac{2}{3}p^2\pi^2 + \mathcal{O}\left(\frac{\log n}{n}\right) \quad \text{für } p \text{ fix}\,,} \tag{5.30a}$$

$$\boxed{V_{n,p} = \frac{2}{9}l + \mathcal{O}\left(\frac{1}{n}\right) \quad \text{für } l = n - p \text{ fix}\,,} \tag{5.30b}$$

$$\boxed{\begin{aligned} V_{n,p} = &\left(-\frac{4\rho^2(\rho+3)\log^2\rho}{(1-\rho)^5} - \frac{2\rho(3\rho^2+12\rho+1)\log\rho}{(1-\rho)^4} - \frac{16\rho^2}{(1-\rho)^3}\right) n + \mathcal{O}(1) \\ &\text{für } p = \rho n \text{ und } 0 < \rho < 1. \end{aligned}} \tag{5.30c}$$

Für die asymptotischen Entwicklungen des Erwartungswertes der Anzahl der Vergleiche benötigt man ebenfalls nur die Beziehungen (1.92), und man erhält aus (5.20)

Korollar 5.2.13. *Der Erwartungswert $E_{n,p}$ der Anzahl der Vergleiche beim Finden von p Elementen in einem Feld der Größe n durch Multiple Quickselect ist asymptotisch gegeben durch*

$$\boxed{\begin{aligned} E_{n,p} = &(2H_p + 1)n - 8p\log n + 2(4p+1)H_p + 3(2p-1) - 8p\gamma + \mathcal{O}\left(\frac{\log n}{n}\right) \\ &\text{für } p \text{ fix}\,, \end{aligned}} \tag{5.31a}$$

$$\boxed{E_{n,p} = 2n\log n + 2(\gamma+2)n + 2\log n + 1 + 2\gamma + \mathcal{O}\left(\frac{\log n}{n}\right) \quad \text{für } l = n - p \text{ fix}\,,} \tag{5.31b}$$

$$\boxed{\begin{aligned} E_{n,p} = &2n\log n + \frac{(4\rho+2)\log\rho + (2\rho^2 - 4\rho + 2)\gamma - 5\rho^2 + 4\rho + 1}{(1-\rho)^2} n + 2\log n + \mathcal{O}(1) \\ &\text{für } p = \rho n \text{ und } 0 < \rho < 1. \end{aligned}} \tag{5.31c}$$

Für die asymptotischen Entwicklungen der Varianz der Anzahl der Vergleiche benötigt man das Verhalten der Summe $\lambda_{n,p}$ für $n \to \infty$. Dafür werden wir $\lambda_{n,p}$ umformen in einen Ausdruck, den wir leichter behandeln können.

Es gilt nämlich

Hilfssatz 5.2.14.

$$\lambda_{n,p} = -\frac{1}{2}\left(H_n^2 - H_n^{(2)} - H_{n-p}^2 - H_{n-p}^{(2)}\right) + \sum_{j=1}^{p-1} \frac{H_j}{n-p+1+j} \quad \textit{für } p \geq 1\,. \tag{5.32}$$

Beweis. Für $p \geq 1$ erhalten wir durch Aufspalten des Binomialkoeffizienten

$$\binom{n-k}{p} = \frac{n-k}{p}\binom{n-1-k}{p-1} \tag{5.33}$$

für $\lambda_{n,p}$ zunächst

$$\lambda_{n,p} = \sum_{k=1}^{n} \frac{1}{k^2} \frac{\binom{n-k}{p}}{\binom{n}{p}} = \sum_{k=1}^{n-1} \frac{1}{k^2} \frac{\binom{n-k}{p}}{\binom{n}{p}} = \sum_{k=1}^{n-1} \frac{1}{k^2} \frac{\binom{n-k-1}{p-1}}{\binom{n}{p}} \frac{n-k}{p}$$
$$= \sum_{k=1}^{n-1} \frac{1}{k^2} \frac{\binom{n-1-k}{p-1}}{\binom{n-1}{p-1}} - \frac{1}{p} \sum_{k=1}^{n-1} \frac{1}{k} \frac{\binom{n-1-k}{p-1}}{\binom{n}{p}} . \tag{5.34}$$

Mit

$$\sum_{k=1}^{n-1} \frac{1}{k} \binom{n-1-k}{p-1} = [z^{n-1} u^{p-1}] \frac{1}{1-z(1+u)} \log\left(\frac{1}{1-z}\right) = \binom{n-1}{p-1} (H_{n-1} - H_{p-1}) \tag{5.35}$$

ergibt sich die Rekursion

$$\lambda_{n,p} = \lambda_{n-1,p-1} - \frac{1}{n}(H_{n-1} - H_{p-1}) \text{ für } p \geq 1 . \tag{5.36}$$

Damit bekommt man zuerst

$$\lambda_{n,p} = \lambda_{n-p,0} - \sum_{j=1}^{p} \frac{1}{n-p+j} (H_{n-p+j-1} - H_{j-1}) , \tag{5.37}$$

woraus wir mit

$$\lambda_{n-p,0} = \sum_{k=1}^{n-p} \frac{1}{k^2} = H_{n-p}^{(2)} \tag{5.38}$$

die Beziehungen

$$\lambda_{n,p} = H_{n-p}^{(2)} - \sum_{j=1}^{p} \frac{H_{n-p+j}}{n-p+j} + \sum_{j=1}^{p} \frac{1}{(n-p+j)^2} + \sum_{j=1}^{p} \frac{H_{j-1}}{n-p+j} \tag{5.39}$$

und schließlich

$$\lambda_{n,p} = -\frac{1}{2} \left(H_n^2 - H_n^{(2)} - H_{n-p}^2 - H_{n-p}^{(2)} \right) + \sum_{j=1}^{p-1} \frac{H_j}{n-p+1+j} \tag{5.40}$$

erhalten. □

Für die im letzten Ausdruck auftretende Summe gelten nachfolgend angeführte Beziehungen, welche sich für die asymptotischen Berechnungen als hilfreich erweisen.

$$\sum_{j=1}^{p-1} \frac{H_j}{n-p+1+j} = H_n H_p - \sum_{j=1}^{p} \frac{H_{n-p+j}}{j} , \tag{5.41a}$$

$$\sum_{j=1}^{p-1} \frac{H_j}{n-p+1+j} = \sum_{j=1}^{p} \frac{H_{n-j}}{j} - \frac{1}{n+1-p}(H_n - H_p - H_{n+1-p}) - H_p H_{n+1-p} + H_n H_{n+1-p}$$
$$- \frac{1}{2}(H_n^2 - H_n^{(2)}) - \frac{1}{2}(H_{n+1-p}^2 + H_{n+1-p}^{(2)}) , \tag{5.41b}$$

$$\begin{aligned}\sum_{j=1}^{p-1}\frac{H_j}{n-p+1+j} &= -\sum_{j=p+1}^{n}\frac{H_{j-p}}{j}-\frac{1}{2}(H_{p-1}^2+H_{p-1}^{(2)})+H_nH_{p-1}+\frac{H_{n-p}}{p}\\ &\quad-\frac{1}{n+1-p}(H_n-H_p-H_{n+1-p})-H_pH_{n+1-p}+H_nH_{n+1-p}\\ &\quad-\frac{1}{2}(H_{n+1-p}^2+H_{n+1-p}^{(2)})\,, \end{aligned} \tag{5.41c}$$

Nun wollen wir die Summe $\sum_{j=1}^{p-1}\frac{H_j}{n-p+1+j}$ für festes p, festes $l=n-p$ bzw. für ein festes Verhältnis $\rho=\frac{p}{n}$ mit $0<\rho<1$ asymptotisch für $n\to\infty$ auswerten.

Für fixes p erhalten wir direkt

$$\begin{aligned}\sum_{j=1}^{p-1}\frac{H_j}{n-p+1+j} &= \sum_{j=1}^{p-1}\frac{H_j}{n}\frac{1}{1-\frac{p-1-j}{n}}=\sum_{j=1}^{p-1}\frac{H_j}{n}\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right)\\ &= \frac{p(H_p-1)}{n}+\mathcal{O}\left(\frac{1}{n^2}\right)\quad\text{für } n\to\infty\,. \end{aligned} \tag{5.42}$$

Für festes $l=n-p$ betrachten wir zunächst

$$\sum_{j=p+1}^{n}\frac{H_{j-p}}{j}=\sum_{j=1}^{l}\frac{H_j}{n+j-l}=\sum_{j=1}^{l}\frac{H_j}{n}\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right)=\frac{(l+1)H_l-l}{n}+\mathcal{O}\left(\frac{1}{n^2}\right)\,, \tag{5.43}$$

setzen dies in (5.41c) ein und entwickeln. So erhalten wir

$$\begin{aligned}\sum_{j=1}^{p-1}\frac{H_j}{n-p+1+j} &= \frac{1}{2}\log^2 n+\gamma\log n-\frac{\pi^2}{12}+\frac{H_{l+1}}{l+1}-\frac{1}{2}H_{l+1}^2+\frac{\gamma^2}{2}-\frac{1}{2}H_{l+1}^{(2)}+\frac{1}{2}\frac{\log n}{n}+\frac{1}{2}\frac{\gamma}{n}\\ &\quad-\frac{lH_l}{n}-\frac{l}{(l+1)n}+\frac{1}{2n}+\frac{lH_{l+1}}{n}+\frac{l}{n}+\mathcal{O}\left(\frac{\log n}{n^2}\right)\quad\text{für } n\to\infty\,. \end{aligned}$$

Für die Asymptotischen Betrachtungen für ein festes Verhältnis $\rho=\frac{p}{n}$ mit $0<\rho<n$ greifen wir auf (1.104) zurück. Damit und mit Gleichung (5.41b) erhalten wir

$$\begin{aligned}\sum_{j=1}^{p-1}\frac{H_j}{n-p+1+j} &= -\log(1-\rho)\log n-\frac{1}{2}\log^2(1-\rho)-\gamma\log(1-\rho)-\log\rho\log(1-\rho)\\ &\quad-\mathrm{L}_2(\rho)+o(1)\quad\text{für } n\to\infty\,. \end{aligned} \tag{5.44}$$

Aus diesen Beziehungen und mit (5.40) bekommen wir

Hilfssatz 5.2.15. *Die Größe $\lambda_{n,p}$ ist asymptotisch gegeben durch*

$$\begin{aligned}\lambda_{n,p} &= \frac{\pi^2}{6}+\mathcal{O}\left(\frac{\log n}{n}\right)\quad\text{für } p \text{ fix}\,,\\ \lambda_{n,p} &= \frac{l}{n}+\mathcal{O}\left(\frac{1}{n^2}\right)\quad\text{für } l=n-p \text{ fix}\,,\\ \lambda_{n,p} &= \mathrm{L}_2(1-\rho)+o(1)\quad\text{für } p=\rho n \text{ und } 0<\rho<1\,. \end{aligned} \tag{5.45}$$

Damit erhält man aus (5.27)

Korollar 5.2.16. *Die Varianz $V_{n,p}$ der Anzahl der Vergleiche beim Finden von p Elementen in einem Feld der Größe n durch Multiple Quickselect ist asymptotisch gegeben durch*

$$\boxed{V_{n,p} = \left(\frac{1}{3}\frac{21p^3+80p^2+81p-2}{(p+1)^2(p+2)} - 4H_p^{(2)}\right) n^2 + \mathcal{O}(n) \quad \text{für } p \text{ fix}\,,} \tag{5.46a}$$

$$\boxed{V_{n,p} = \left(\frac{7l^4+70l^3+245l^2+368l+168}{(l+4)^{\underline{4}}} - \frac{2}{3}\pi^2\right) n^2 + \mathcal{O}(n\log n) \quad \text{für } l=n-p \text{ fix}\,,} \tag{5.46b}$$

$$\boxed{V_{n,p} = \left(7 - \frac{2}{3}\pi^2\right) n^2 + \mathcal{O}(n\log n) \quad \text{für } p=\rho n \text{ und } 0<\rho<1.} \tag{5.46c}$$

5.3 Multiple Quickselect mit Median-of-Three Partition

5.3.1 Die Anzahl der Durchläufe

Bezeichnen wir nun mit $\varphi_{n,p,m}$ die Wahrscheinlichkeit, daß beim Finden von p zufälligen Elementen in einem Datenfeld der Größe n durch den Multiple Quickselect Algorithmus mit Median-of-Three Partition genau m Durchläufe gemacht werden. Analog zu (5.4) erhalten wir hier wieder eine Rekursion für die $\varphi_{n,p,m}$. Sie unterscheidet sich gegenüber der dort erhaltenen Beziehung nur durch die geänderten Wahrscheinlichkeiten für die Wahl des Pivotelements und durch die Tatsache, daß kleine Datenfelder der Größe $n\le 2$ nicht mehr rekursiv abgearbeitet werden.

Es ergibt sich also folgende Rekursion

$$\begin{aligned}
\varphi_{n,p,m} = \sum_{k=1}^{n} \frac{(k-1)(n-k)}{\binom{n}{3}} &\left[\sum_{i=0}^{p}\sum_{l=0}^{m-1} \frac{\binom{k-1}{i}\binom{n-k}{p-i}}{\binom{n}{p}} \varphi_{k-1,i,l}\,\varphi_{n-k,p-i,m-1-l}\right.\\
&\left.+\sum_{i=0}^{p-1}\sum_{l=0}^{m-1} \frac{\binom{k-1}{i}\binom{n-k}{p-i-1}}{\binom{n}{p}} \varphi_{k-1,i,l}\,\varphi_{n-k,p-1-i,m-1-l}\right] \quad \text{für } n\ge 3 \text{ und } 1\le p\le n\,,\\
\varphi_{n,0,m} = \begin{cases} 1 & \text{wenn } m=0\,,\\ 0 & \text{wenn } m\neq 0\,. \end{cases}, &\quad \varphi_{1,1,1}=1\,,\ \varphi_{2,1,1}=1\,,\ \varphi_{2,2,1}=1\,,\ \varphi_{n,p,m}=0 \text{ sonst.}
\end{aligned} \tag{5.47}$$

Durch Multiplikation mit $\binom{n}{3}z^{n-3}u^p v^m$ und Aufsummieren über alle $n\ge 3$, $p\ge 1$ und $m\ge 0$ erhält man aus (5.47) folgende Differentialgleichung für die in (5.1) definierte Funktion $\Phi(z,u,v)$:

$$\begin{aligned}
&\frac{1}{6}\frac{\partial^3}{\partial z^3}\Phi(z,u,v) = v(1+u)\left(\frac{\partial}{\partial z}\Phi(z,u,v)\right)^2 + \frac{1-v}{(1-z)^4}\,,\\
&\Phi(0,u,v)=1\,,\quad \left.\frac{\partial}{\partial z}\Phi(z,u,v)\right|_{z=0} = 1+uv\,,\quad \left.\frac{\partial^2}{\partial z^2}\Phi(z,u,v)\right|_{z=0} = 2(1+2uv+u^2v)\,.
\end{aligned} \tag{5.48}$$

Sei weiters

$$\Psi(z,u,v) = \frac{\partial}{\partial z}\Phi(z,u,v)\,, \tag{5.49}$$

dann schreibt sich diese Differentialgleichung als

$$\begin{aligned} &\frac{1}{6}\frac{\partial^2}{\partial z^2}\Psi(z,u,v) = v(1+u)\Psi(z,u,v)^2 + \frac{1-v}{(1-z)^4}\,, \\ &\Psi(0,u,v) = 1+uv\,, \quad \frac{\partial}{\partial z}\Psi(z,u,v)\bigg|_{z=0} = 2(1+2uv+u^2v)\,. \end{aligned} \tag{5.50}$$

Um die Erwartungswerte $E_{n,p}$ zu bestimmen, definieren wir zunächst

$$G(z,u) = \frac{\partial}{\partial v}\Psi(z,u,v)\bigg|_{v=1}\,. \tag{5.51}$$

Es besteht dann der Zusammenhang

$$E_{n,p} = \frac{1}{n\binom{n}{p}}[z^{n-1}u^p]G(z,u)\,. \tag{5.52}$$

Differenzieren wir (5.50) nach v und werten anschließend bei $v=1$ aus, liefert dies die lineare Differentialgleichung zweiter Ordnung

$$\begin{aligned} &\frac{1}{6}\frac{\partial^2}{\partial z^2}G(z,u) = \frac{2(1+u)^2}{(1-z(1+u))^2}G(z,u) + \frac{(1+u)^3}{(1-z(1+u))^4} - \frac{1}{(1-z)^4}\,, \\ &G(0,u) = u\,, \quad \frac{\partial}{\partial z}G(z,u)\bigg|_{z=0} = 4u+2u^2\,. \end{aligned} \tag{5.53}$$

Lösen derselben führt zu

$$\begin{aligned} G(z,u) = {} & \frac{12}{7}\frac{(u+1)^2\,p_1(z,u)}{(1-z(1+u))^3\,u^6}\log\left(\frac{1}{1-z}\right) \\ & + \frac{60}{7}\frac{(u+1)^2\,(1-z(1+u))^4}{u^6}\log\left(\frac{1}{1-z(1+u)}\right) - \frac{1}{7}\frac{p_2(z,u)}{(1-z(1+u))^3\,(1-z)^2\,u^5}\,, \end{aligned} \tag{5.54}$$

mit

$$\begin{aligned} p_1(z,u) = {} & 5\,z^7\,u^7 + 35\,z^7\,u^6 + 105\,z^7\,u^5 + 175\,z^7\,u^4 + 175\,z^7\,u^3 + 105\,z^7\,u^2 + 35\,z^7\,u + 5\,z^7 \\ & - 35\,z^6\,u^6 - 210\,z^6\,u^5 - 525\,z^6\,u^4 - 700\,z^6\,u^3 - 525\,z^6\,u^2 - 210\,z^6\,u - 35\,z^6 \\ & + 105\,z^5\,u^5 + 525\,z^5\,u^4 + 1050\,z^5\,u^3 + 1050\,z^5\,u^2 + 525\,z^5\,u + 105\,z^5 \\ & - 175\,z^4\,u^4 - 700\,z^4\,u^3 - 1050\,z^4\,u^2 - 700\,z^4\,u - 175\,z^4 + 175\,z^3\,u^3 \\ & + 525\,z^3\,u^2 + 525\,z^3\,u + 175\,z^3 - 105\,z^2\,u^2 - 210\,z^2\,u - 105\,z^2 + 35\,z\,u \\ & + 35\,z + 2\,u^7 - 5\,, \end{aligned}$$

$$\begin{aligned} p_2(z,u) = {} & 2\,z^9\,u^{13} + 13\,z^9\,u^{12} + 37\,z^9\,u^{11} + 58\,z^9\,u^{10} + 62\,z^9\,u^9 + 202\,z^9\,u^8 + 998\,z^9\,u^7 \\ & + 2872\,z^9\,u^6 + 5018\,z^9\,u^5 + 5637\,z^9\,u^4 + 4125\,z^9\,u^3 + 1910\,z^9\,u^2 + 510\,z^9\,u \end{aligned}$$

$$
\begin{aligned}
&+ 60\,z^9 - 4\,z^8\,u^{13} - 40\,z^8\,u^{12} - 151\,z^8\,u^{11} - 298\,z^8\,u^{10} - 348\,z^8\,u^9 - 554\,z^8\,u^8 \\
&- 2720\,z^8\,u^7 - 9846\,z^8\,u^6 - 20998\,z^8\,u^5 - 27878\,z^8\,u^4 - 23545\,z^8\,u^3 \\
&- 12360\,z^8\,u^2 - 3690\,z^8\,u - 480\,z^8 + 2\,z^7\,u^{13} + 41\,z^7\,u^{12} + 233\,z^7\,u^{11} \\
&+ 611\,z^7\,u^{10} + 867\,z^7\,u^9 + 847\,z^7\,u^8 + 2611\,z^7\,u^7 + 12483\,z^7\,u^6 \\
&+ 34481\,z^7\,u^5 + 56212\,z^7\,u^4 + 56240\,z^7\,u^3 + 34110\,z^7\,u^2 + 11550\,z^7\,u \\
&+ 1680\,z^7 - 14\,z^6\,u^{12} - 161\,z^6\,u^{11} - 630\,z^6\,u^{10} - 1183\,z^6\,u^9 - 1170\,z^6\,u^8 \\
&- 1259\,z^6\,u^7 - 6966\,z^6\,u^6 - 27635\,z^6\,u^5 - 58808\,z^6\,u^4 - 72400\,z^6\,u^3 \\
&- 52060\,z^6\,u^2 - 20370\,z^6\,u - 3360\,z^6 + 42\,z^5\,u^{11} + 329\,z^5\,u^{10} + 917\,z^5\,u^9 \\
&+ 1195\,z^5\,u^8 + 745\,z^5\,u^7 + 1557\,z^5\,u^6 + 10729\,z^5\,u^5 + 33377\,z^5\,u^4 \\
&+ 53725\,z^5\,u^3 + 47550\,z^5\,u^2 + 22050\,z^5\,u + 4200\,z^5 - 70\,z^4\,u^{10} - 385\,z^4\,u^9 \\
&- 740\,z^4\,u^8 - 605\,z^4\,u^7 - 180\,z^4\,u^6 - 1595\,z^4\,u^5 - 9610\,z^4\,u^4 - 22625\,z^4\,u^3 \\
&- 26160\,z^4\,u^2 - 14910\,z^4\,u - 3360\,z^4 + 70\,z^3\,u^9 + 267\,z^3\,u^8 + 324\,z^3\,u^7 \\
&+ 127\,z^3\,u^6 + 1070\,z^3\,u^4 + 4870\,z^3\,u^3 + 8210\,z^3\,u^2 + 6090\,z^3\,u + 1680\,z^3 \\
&- 78\,z^2\,u^8 - 149\,z^2\,u^7 - 71\,z^2\,u^6 - 390\,z^2\,u^3 - 1260\,z^2\,u^2 - 1350\,z^2\,u - 480\,z^2 \\
&+ 24\,z\,u^8 + 55\,z\,u^7 + 31\,z\,u^6 + 60\,z\,u^2 + 120\,z\,u + 60\,z - 7\,u^6\,.
\end{aligned}
$$

Liest man unter Verwendung der Beziehungen (5.9) von (5.54) die Koeffizienten ab, führt das schon zu

Satz 5.3.1. *Der Erwartungswert $E_{n,p}$ der Anzahl der Durchläufe beim Finden von p Elementen in einem Feld der Größe n durch Multiple Quickselect mit Median-of-Three Partition ist gegeben durch*

$$
\boxed{
\begin{aligned}
E_{n,p} = {}& \frac{12}{7}\frac{(n+1)^2\,p}{(n-p+1)\,(n-p+2)}H_n - \frac{12}{7}\frac{(n+1)^2\,p}{(n-p+1)\,(n-p+2)}H_p \\
&- \frac{1}{7}\frac{r(n,p)}{n^{\underline{6}}\,(p+6)^{\underline{6}}\,(n+1-p)\,(n+2-p)}\,, \quad \text{für } n \geq 6 \text{ und } 1 \leq p \leq n\,,
\end{aligned}
}
$$

$$
\boxed{
\begin{aligned}
&E_{1,1} = 1\,,\ E_{2,1} = 1\,,\ E_{2,2} = 1\,,\ E_{3,1} = \frac{5}{3}\,,\ E_{3,2} = \frac{7}{3}\,,\ E_{3,3} = 3\,, \\
&E_{4,1} = \frac{7}{4}\,,\ E_{4,2} = \frac{7}{3}\,,\ E_{4,3} = \frac{11}{4}\,,\ E_{4,4} = 3\,, \\
&E_{5,1} = \frac{51}{25}\,,\ E_{5,2} = \frac{141}{50}\,,\ E_{5,3} = \frac{17}{5}\,,\ E_{5,4} = \frac{96}{25}\,,\ E_{5,5} = \frac{21}{5}\,.
\end{aligned}
}
\tag{5.55}
$$

mit

$$
\begin{aligned}
r(n,p) = {}& 12\,n^8\,p^7 + 247\,n^8\,p^6 + 1995\,n^8\,p^5 + 7945\,n^8\,p^4 + 15813\,n^8\,p^3 + 13048\,n^8\,p^2 \\
&- 180\,n^8\,p - 5040\,n^8 - 12\,n^7\,p^8 - 422\,n^7\,p^7 - 5598\,n^7\,p^6 - 37058\,n^7\,p^5 \\
&- 131838\,n^7\,p^4 - 244328\,n^7\,p^3 - 191592\,n^7\,p^2 + 14688\,n^7\,p + 60480\,n^7 \\
&+ 175\,n^6\,p^8 + 4536\,n^6\,p^7 + 48328\,n^6\,p^6 + 271362\,n^6\,p^5 + 850885\,n^6\,p^4 \\
&+ 1429134\,n^6\,p^3 + 1000692\,n^6\,p^2 - 97272\,n^6\,p - 272160\,n^6 - 945\,n^5\,p^8 \\
&- 21560\,n^5\,p^7 - 200550\,n^5\,p^6 - 977060\,n^5\,p^5 - 2648205\,n^5\,p^4
\end{aligned}
$$

$$
\begin{aligned}
&- 3784340\, n^5\, p^3 - 2220540\, n^5\, p^2 + 236880\, n^5\, p + 544320\, n^5 + 2275\, n^4\, p^8 \\
&+ 48048\, n^4\, p^7 + 403333\, n^4\, p^6 + 1708875\, n^4\, p^5 + 3752980\, n^4\, p^4 \\
&+ 3919377\, n^4\, p^3 + 1317372\, n^4\, p^2 - 477540\, n^4\, p - 347760\, n^4 - 2163\, n^3\, p^8 \\
&- 42098\, n^3\, p^7 - 309372\, n^3\, p^6 - 991802\, n^3\, p^5 - 1085637\, n^3\, p^4 \\
&+ 989548\, n^3\, p^3 + 2813412\, n^3\, p^2 + 1168272\, n^3\, p - 362880\, n^3 + 70\, n^2\, p^8 \\
&- 2196\, n^2\, p^7 - 73908\, n^2\, p^6 - 601272\, n^2\, p^5 - 2321130\, n^2\, p^4 - 4860324\, n^2\, p^3 \\
&- 5002312\, n^2\, p^2 - 1360368\, n^2\, p + 624960\, n^2 + 600\, n\, p^8 + 15120\, n\, p^7 \\
&+ 127440\, n\, p^6 + 574560\, n\, p^5 + 1625400\, n\, p^4 + 2756880\, n\, p^3 \\
&+ 2229600\, n\, p^2 + 342720\, n\, p - 241920\, n + 1440\, p^7 + 10080\, p^6 + 10080\, p^5 \\
&- 50400\, p^4 - 80640\, p^3 + 40320\, p^2 + 69120\, p\,.
\end{aligned}
$$

Die Fälle $p = 1$ (Hoare's Find Algorithmus) und $p = n$ (Quicksort) seien gesondert notiert. Es ergibt sich

Korollar 5.3.2.

$$\boxed{E_{n,1} = \frac{12}{7}\frac{n+1}{n}H_n - \frac{131n-2}{49n}\,,} \tag{5.56a}$$

$$\boxed{E_{n,n} = \frac{6}{7}n - \frac{1}{7}\,.} \tag{5.56b}$$

5.3.2 Die Anzahl der Vergleiche

Wir bezeichnen hier mit $\varphi_{n,p,m}$ die Wahrscheinlichkeit, daß beim Finden von p zufälligen Elementen in einem Datenfeld der Größe n durch den Multiple Quickselect Algorithmus mit Median-of-Three Partition genau m Vergleiche gemacht werden. Man erhält eine Rekursion, die sich gegenüber (5.16) nur durch die geänderten Wahrscheinlichkeiten für die Wahl des Pivotelements und die geänderten Startwerte für $n \leq 2$ unterscheidet. Sie lautet

$$
\begin{aligned}
\varphi_{n,p,m} = \sum_{k=1}^{n} \frac{(k-1)(n-k)}{\binom{n}{3}} \Bigg[&\sum_{i=0}^{p} \sum_{l=0}^{m-n+1} \frac{\binom{k-1}{i}\binom{n-k}{p-i}}{\binom{n}{p}} \varphi_{k-1,i,l}\, \varphi_{n-k,p-i,m-n+1-l} \\
&+ \sum_{i=0}^{p-1} \sum_{l=0}^{m-n+1} \frac{\binom{k-1}{i}\binom{n-k}{p-i-1}}{\binom{n}{p}} \varphi_{k-1,i,l}\, \varphi_{n-k,p-1-i,m-n+1-l} \Bigg] \\
&\text{für } n \geq 3 \text{ und } 1 \leq p \leq n\,,
\end{aligned}
$$

$$\varphi_{n,0,m} = \begin{cases} 1 & \text{wenn } m = 0\,, \\ 0 & \text{wenn } m \neq 0\,. \end{cases}, \quad \varphi_{1,1,0} = 1\,, \quad \varphi_{2,1,1} = 1\,, \quad \varphi_{2,2,1} = 1\,, \quad \varphi_{n,p,m} = 0 \text{ sonst.} \tag{5.57}$$

Aus (5.57) erhalten wir durch Multiplikation mit $\binom{n}{3} z^{n-3} u^p v^m$ und Aufsummieren über alle $n \geq 3$, $p \geq 1$ und $m \geq 0$ die Differentialgleichung

$$
\begin{aligned}
&\frac{1}{6}\frac{\partial^3}{\partial z^3}\Phi(z,u,v) = (1+u)\left(\frac{\partial}{\partial z}\Phi(zv,u,v)\right)^2 + \frac{1}{(1-z)^4} - \frac{v^2}{(1-zv)^4}\,, \\
&\Phi(0,u,v) = 1\,, \quad \frac{\partial}{\partial z}\Phi(z,u,v)\bigg|_{z=0} = 1 + u\,, \quad \frac{\partial^2}{\partial z^2}\Phi(z,u,v)\bigg|_{z=0} = 2(1 + 2uv + u^2 v)\,,
\end{aligned} \tag{5.58}
$$

wobei die erzeugende Funktion $\Phi(z,u,v)$ wie in (5.1) definiert sein soll.

Sei weiters $\Psi(z,u,v)$ definiert wie in (5.49), dann schreibt sich diese Differentialgleichung als

$$\begin{aligned} &\frac{1}{6}\frac{\partial^2}{\partial z^2}\Psi(z,u,v) = (1+u)v^2\Psi(zv,u,v)^2 + \frac{1}{(1-z)^4} - \frac{v^2}{(1-zv)^4}\,, \\ &\Psi(0,u,v) = 1+u\,,\quad \frac{\partial}{\partial z}\Psi(z,u,v)\bigg|_{z=0} = 2(1+2uv+u^2v)\,. \end{aligned} \tag{5.59}$$

Ableiten von (5.59) nach v und anschließendes Auswerten bei $v=1$ liefert folgende lineare Differentialgleichung zweiter Ordnung für die wie in (5.51) definierte Funktion $G(z,u)$:

$$\begin{aligned} &\frac{1}{6}\frac{\partial^2}{\partial z^2}G(z,u) = \frac{2(1+u)^2}{(1-z(1+u))^2}G(z,u) + \frac{2(1+u)^3(1+z(1+u))}{(1-z(1+u))^5} - \frac{2(1+z)}{(1-z)^5}\,, \\ &G(0,u) = 0\,,\quad \frac{\partial}{\partial z}G(z,u)\bigg|_{z=0} = 4u+2u^2\,. \end{aligned} \tag{5.60}$$

Die Lösung dieser Differentialgleichung ist nun gegeben durch

$$\begin{aligned} G(z,u) = {} & \frac{24}{7}\frac{(u+1)^2\,p_1(z,u)}{(1-z(1+u))^3\,u^7}\log\left(\frac{1}{1-z}\right) - \frac{24}{7}\frac{(u+1)\,p_2(z,u)}{(1-z(1+u))^3\,u^7}\log\left(\frac{1}{1-z(1+u)}\right) \\ & - \frac{2}{49}\frac{z\,g_3(z,u)}{(1-z(1+u))^3\,(1-z)^3\,u^6}\,, \end{aligned} \tag{5.61}$$

mit

$$\begin{aligned} p_1(z,u) = {} & 10\,z^7\,u^8 + 85\,z^7\,u^7 + 315\,z^7\,u^6 + 665\,z^7\,u^5 + 875\,z^7\,u^4 + 735\,z^7\,u^3 + 385\,z^7\,u^2 \\ & + 115\,z^7\,u + 15\,z^7 - 70\,z^6\,u^7 - 525\,z^6\,u^6 - 1680\,z^6\,u^5 - 2975\,z^6\,u^4 \\ & - 3150\,z^6\,u^3 - 1995\,z^6\,u^2 - 700\,z^6\,u - 105\,z^6 + 210\,z^5\,u^6 + 1365\,z^5\,u^5 \\ & + 3675\,z^5\,u^4 + 5250\,z^5\,u^3 + 4200\,z^5\,u^2 + 1785\,z^5\,u + 315\,z^5 - 350\,z^4\,u^5 \\ & - 1925\,z^4\,u^4 - 4200\,z^4\,u^3 - 4550\,z^4\,u^2 - 2450\,z^4\,u - 525\,z^4 + 350\,z^3\,u^4 \\ & + 1575\,z^3\,u^3 + 2625\,z^3\,u^2 + 1925\,z^3\,u + 525\,z^3 - 210\,z^2\,u^3 - 735\,z^2\,u^2 \\ & - 840\,z^2\,u - 315\,z^2 + 70\,z\,u^2 + 175\,z\,u + 105\,z - 3\,u^8 - u^7 - 10\,u - 15\,, \end{aligned}$$

$$\begin{aligned} p_2(z,u) = {} & 10\,z^7\,u^9 + 95\,z^7\,u^8 + 400\,z^7\,u^7 + 980\,z^7\,u^6 + 1540\,z^7\,u^5 + 1610\,z^7\,u^4 + 1120\,z^7\,u^3 \\ & + 500\,z^7\,u^2 + 130\,z^7\,u + 15\,z^7 - 70\,z^6\,u^8 - 595\,z^6\,u^7 - 2205\,z^6\,u^6 - 4655\,z^6\,u^5 \\ & - 6125\,z^6\,u^4 - 5145\,z^6\,u^3 - 2695\,z^6\,u^2 - 805\,z^6\,u - 105\,z^6 + 210\,z^5\,u^7 \\ & + 1575\,z^5\,u^6 + 5040\,z^5\,u^5 + 8925\,z^5\,u^4 + 9450\,z^5\,u^3 + 5985\,z^5\,u^2 \\ & + 2100\,z^5\,u + 315\,z^5 - 350\,z^4\,u^6 - 2275\,z^4\,u^5 - 6125\,z^4\,u^4 - 8750\,z^4\,u^3 \\ & - 7000\,z^4\,u^2 - 2975\,z^4\,u - 525\,z^4 + 350\,z^3\,u^5 + 1925\,z^3\,u^4 + 4200\,z^3\,u^3 \\ & + 4550\,z^3\,u^2 + 2450\,z^3\,u + 525\,z^3 - 210\,z^2\,u^4 - 945\,z^2\,u^3 - 1575\,z^2\,u^2 \\ & - 1155\,z^2\,u - 315\,z^2 + 70\,z\,u^3 + 245\,z\,u^2 + 280\,z\,u + 105\,z - u^7 - 10\,u^2 \\ & - 25\,u - 15\,, \end{aligned}$$

$$p_3(z,u) = 2\,z^9\,u^{14} + 16\,z^9\,u^{13} + 49\,z^9\,u^{12} + 91\,z^9\,u^{11} + 14\,z^9\,u^{10} - 2450\,z^9\,u^9 - 17374\,z^9\,u^8$$

$$
\begin{aligned}
&- 60926\, z^9\, u^7 - 130534\, z^9\, u^6 - 184296\, z^9\, u^5 - 176127\, z^9\, u^4 - 113365\, z^9\, u^3 \\
&- 47250\, z^9\, u^2 - 11550\, z^9\, u - 1260\, z^9 - 6\, z^8\, u^{14} - 62\, z^8\, u^{13} - 245\, z^8\, u^{12} \\
&- 518\, z^8\, u^{11} - 434\, z^8\, u^{10} + 6804\, z^8\, u^9 + 60998\, z^8\, u^8 + 253940\, z^8\, u^7 \\
&+ 631002\, z^8\, u^6 + 1015546\, z^8\, u^5 + 1091195\, z^8\, u^4 + 780850\, z^8\, u^3 \\
&+ 358470\, z^8\, u^2 + 95760\, z^8\, u + 11340\, z^8 + 6\, z^7\, u^{14} + 90\, z^7\, u^{13} + 483\, z^7\, u^{12} \\
&+ 1260\, z^7\, u^{11} + 1701\, z^7\, u^{10} - 5439\, z^7\, u^9 - 78855\, z^7\, u^8 - 415647\, z^7\, u^7 \\
&- 1244265\, z^7\, u^6 - 2342319\, z^7\, u^5 - 2882880\, z^7\, u^4 - 2326065\, z^7\, u^3 \\
&- 1189230\, z^7\, u^2 - 350280\, z^7\, u - 45360\, z^7 - 2\, z^6\, u^{14} - 58\, z^6\, u^{13} - 469\, z^6\, u^{12} \\
&- 1652\, z^6\, u^{11} - 2989\, z^6\, u^{10} - 742\, z^6\, u^9 + 43617\, z^6\, u^8 + 332316\, z^6\, u^7 \\
&+ 1274453\, z^6\, u^6 + 2914716\, z^6\, u^5 + 4216380\, z^6\, u^4 + 3907820\, z^6\, u^3 \\
&+ 2256450\, z^6\, u^2 + 740880\, z^6\, u + 105840\, z^6 + 14\, z^5\, u^{13} + 224\, z^5\, u^{12} \\
&+ 1211\, z^5\, u^{11} + 2961\, z^5\, u^{10} + 3640\, z^5\, u^9 - 7161\, z^5\, u^8 - 128646\, z^5\, u^7 \\
&- 709779\, z^5\, u^6 - 2096234\, z^5\, u^5 - 3701635\, z^5\, u^4 - 4043165\, z^5\, u^3 \\
&- 2685690\, z^5\, u^2 - 996660\, z^5\, u - 158760\, z^5 - 42\, z^4\, u^{12} - 462\, z^4\, u^{11} \\
&- 1743\, z^4\, u^{10} - 2898\, z^4\, u^9 - 2205\, z^4\, u^8 + 18648\, z^4\, u^7 + 201453\, z^4\, u^6 \\
&+ 860622\, z^4\, u^5 + 1971627\, z^4\, u^4 + 2635290\, z^4\, u^3 + 2065350\, z^4\, u^2 \\
&+ 882000\, z^4\, u + 158760\, z^4 + 70\, z^3\, u^{11} + 560\, z^3\, u^{10} + 1484\, z^3\, u^9 \\
&+ 1568\, z^3\, u^8 + 574\, z^3\, u^7 - 22330\, z^3\, u^6 - 183015\, z^3\, u^5 - 609210\, z^3\, u^4 \\
&- 1058575\, z^3\, u^3 - 1015770\, z^3\, u^2 - 511560\, z^3\, u - 105840\, z^3 - 70\, z^2\, u^{10} \\
&- 868\, z^2\, u^9 - 1526\, z^2\, u^8 - 728\, z^2\, u^7 + 14980\, z^2\, u^5 + 96110\, z^2\, u^4 \\
&+ 243040\, z^2\, u^3 + 303030\, z^2\, u^2 + 186480\, z^2\, u + 45360\, z^2 + 672\, z\, u^9 \\
&+ 1491\, z\, u^8 + 819\, z\, u^7 - 5460\, z\, u^4 - 26670\, z\, u^3 - 48300\, z\, u^2 - 38430\, z\, u \\
&- 11340\, z - 252\, u^9 - 553\, u^8 - 350\, u^7 + 840\, u^3 + 2940\, u^2 + 3360\, u + 1260\,.
\end{aligned}
$$

Ablesen der Koeffizienten führt mittels (5.52) unter Verwendung von (5.9) nun zu

Satz 5.3.3. *Der Erwartungswert $E_{n,p}$ der Anzahl der Vergleiche beim Finden von p Elementen in einem Feld der Größe n durch Multiple Quickselect mit Median-of-Three Partition ist gegeben durch*

$$
\boxed{
\begin{aligned}
E_{n,p} = &-\frac{12}{7}\,\frac{(n+1)\,(4n-p+5)\,p}{(n-p+1)\,(n-p+2)}\,H_n + \frac{12}{7}\,\frac{(n+1)^2\,(n+2p+2)}{(n-p+2)\,(n-p+1)}\,H_p \\
&+ \frac{1}{49}\,\frac{r(n,p)}{n^{\underline{6}}\,(p+7)^{\underline{7}}\,(n+1-p)\,(n+2-p)}\,, \quad \text{für } n \geq 6 \text{ und } 1 \leq p \leq n\,,
\end{aligned}
}
$$

$$
\boxed{
\begin{aligned}
&E_{1,1} = 0\,,\ \ E_{2,1} = 1\,,\ \ E_{2,2} = 1\,,\ \ E_{3,1} = 2\,,\ \ E_{3,2} = 2\,,\ \ E_{3,3} = 2\,, \\
&E_{4,1} = \frac{7}{2}\,,\ \ E_{4,2} = \frac{23}{6}\,,\ \ E_{4,3} = 4\,,\ \ E_{4,4} = 4\,, \\
&E_{5,1} = \frac{126}{25}\,,\ \ E_{5,2} = \frac{141}{25}\,,\ \ E_{5,3} = \frac{148}{25}\,,\ \ E_{5,4} = 6\,,\ \ E_{5,5} = 6\,.
\end{aligned}
}
\tag{5.62}
$$

mit

$$\begin{aligned}
r(n,p) = {} & 37\,n^9\,p^7 + 1036\,n^9\,p^6 + 11914\,n^9\,p^5 + 72520\,n^9\,p^4 + 250453\,n^9\,p^3 + 485884\,n^9\,p^2 \\
& + 483516\,n^9\,p + 246960\,n^9 + 178\,n^8\,p^8 + 4423\,n^8\,p^7 + 41608\,n^8\,p^6 \\
& + 168238\,n^8\,p^5 + 105322\,n^8\,p^4 - 1459913\,n^8\,p^3 - 5040948\,n^8\,p^2 \\
& - 6877548\,n^8\,p - 3210480\,n^8 - 215\,n^7\,p^9 - 8651\,n^7\,p^8 - 139688\,n^7\,p^7 \\
& - 1178702\,n^7\,p^6 - 5578475\,n^7\,p^5 - 14341019\,n^7\,p^4 - 15631422\,n^7\,p^3 \\
& + 8099412\,n^7\,p^2 + 31066920\,n^7\,p + 16299360\,n^7 + 3192\,n^6\,p^9 \\
& + 103600\,n^6\,p^8 + 1417654\,n^6\,p^7 + 10600072\,n^6\,p^6 + 46767364\,n^6\,p^5 \\
& + 121653280\,n^6\,p^4 + 168818566\,n^6\,p^3 + 84897288\,n^6\,p^2 - 38913336\,n^6\,p \\
& - 40007520\,n^6 - 17780\,n^5\,p^9 - 529760\,n^5\,p^8 - 6609827\,n^5\,p^7 \\
& - 44872436\,n^5\,p^6 - 179988074\,n^5\,p^5 - 428908760\,n^5\,p^4 - 581222523\,n^5\,p^3 \\
& - 387896964\,n^5\,p^2 - 53621316\,n^5\,p + 43711920\,n^5 + 45570\,n^4\,p^9 \\
& + 1288672\,n^4\,p^8 + 15030547\,n^4\,p^7 + 93440452\,n^4\,p^6 + 331748032\,n^4\,p^5 \\
& + 679515928\,n^4\,p^4 + 772514743\,n^4\,p^3 + 431025588\,n^4\,p^2 \\
& + 83709108\,n^4\,p + 740880\,n^4 - 51485\,n^3\,p^9 - 1380029\,n^3\,p^8 \\
& - 14864922\,n^3\,p^7 - 80529498\,n^3\,p^6 - 228035325\,n^3\,p^5 - 283648701\,n^3\,p^4 \\
& + 46816532\,n^3\,p^3 + 407511188\,n^3\,p^2 + 191861280\,n^3\,p - 48404160\,n^3 \\
& + 16758\,n^2\,p^9 + 376590\,n^2\,p^8 + 2608656\,n^2\,p^7 + 2675148\,n^2\,p^6 \\
& - 62468994\,n^2\,p^5 - 381548370\,n^2\,p^4 - 945477876\,n^2\,p^3 - 1039134408\,n^2\,p^2 \\
& - 367655184\,n^2\,p + 42477120\,n^2 + 3960\,n\,p^9 + 169560\,n\,p^8 \\
& + 2311200\,n\,p^7 + 18975600\,n\,p^6 + 99068760\,n\,p^5 + 310910040\,n\,p^4 \\
& + 552004560\,n\,p^3 + 497069280\,n\,p^2 + 161881920\,n\,p - 11854080\,n \\
& + 20160\,p^8 + 161280\,p^7 + 282240\,p^6 - 564480\,p^5 - 1834560\,p^4 \\
& - 564480\,p^3 + 1532160\,p^2 + 967680\,p\,.
\end{aligned}$$

Von besonderem Interesse sind wiederum die Fälle $p = 1$ (Hoare's Find Algorithmus mit Median-of-Three Partition) und $p = n$ (Quicksort mit Median-of-Three Partition). Dafür ergibt sich

Korollar 5.3.4.

$$\boxed{E_{n,1} = \frac{245n^2 + 1111n - 2}{98n} - \frac{48}{7}\frac{n+1}{n}H_n\,,} \tag{5.63a}$$

$$\boxed{E_{n,n} = \frac{12}{7}(n+1)H_n - \frac{173}{49}n + \frac{9}{49}\,.} \tag{5.63b}$$

Es sei hier angemerkt, daß es selbst unter Verwendung eines Formelmanipulationsprogramms nicht mehr selbstverständlich ist, wie man von solchen Funktionen

$$G(z,u) = p(z,u)\,F(z,u)\,,$$

noch die Koeffizienten ablesen kann, wenn das Polynom $p(z, u)$ beispielsweise aus über Hundert Summanden besteht, obwohl man eine explizite Formel für die Koeffizienten von $F(z, u)$ kennt. Es bieten sich hierzu aber folgende zwei Möglichkeiten an.

Einerseits kann man eine Prozedur schreiben, welche das Polynom $p(z, u)$ einliest, jeden Summanden dieses Polynoms analysiert und schließlich aufgrund der Kenntnis der Koeffizienten von $F(z, u)$ ein Ergebnis für die Koeffizienten von $G(z, u)$ liefert.

Andererseits kann man ausnützen, daß man die Struktur der Koeffizienten von $G(z, u)$ kennt, da ja die Koeffizienten von $F(z, u)$ bekannt sind. Man weiß so, daß sich die Koeffizienten von $G(z, u)$ schreiben lassen als

$$[z^n u^p]G(z, u) = q_{n,p}\, h_{n,p}\,,$$

wobei $q_{n,p}$ ein Polynom mit vorerst noch unbestimmten Koeffizienten ist, dessen Grad durch den Grad von $p(z, u)$ bestimmt wird und $h_{n,p}$ eine bekannte Funktion darstellt. Man kann nun das Polynom $q_{n,p}$ bestimmen, indem man genügend viele Koeffizienten von $G(z, u)$ bestimmt und damit die Koeffizienten von $q_{n,p}$ anpaßt.

Beide hier beschriebenen Methoden fanden in dieser Arbeit Verwendung.

5.3.3 Asymptotische Betrachtungen

Wir untersuchen nun das Verhalten der Größen dieses Abschnitts für $n \to \infty$ und unterscheiden dabei die Fälle p fest, $l = n - p$ fest und ein festes Verhältnis $\rho = \frac{p}{n}$ mit $0 < \rho < 1$.

Allein unter Verwendung der asymptotischen Entwicklungen (1.92) für die Harmonischen Zahlen erhalten wir aus (5.55) für die Erwartungswerte der Anzahl der Durchläufe

Korollar 5.3.5. *Der Erwartungswert $E_{n,p}$ der Anzahl der Durchläufe beim Finden von p Elementen in einem Feld der Größe n durch Multiple Quickselect mit Median-of-Three Partition ist asymptotisch gegeben durch*

$$\boxed{\begin{aligned} E_{n,p} = {} & \frac{12}{7} p \log n - \frac{12}{7} p H_p + \frac{12}{7} p\gamma - \tfrac{12p^7+247p^6+1995p^5+7945p^4+15813p^3+13048p^2-180p-5040}{7\,(p+6)^{\underline{6}}} \\ & + \mathcal{O}\left(\frac{\log n}{n}\right) \quad \text{für } p \text{ fix}\,, \end{aligned}} \tag{5.64a}$$

$$\boxed{E_{n,p} = \frac{6}{7} n - \frac{2}{7} l - \frac{1}{7} + \mathcal{O}\left(\frac{\log n}{n}\right) \quad \text{für } l = n - p \text{ fix}\,,} \tag{5.64b}$$

$$\boxed{E_{n,p} = \frac{12}{7} \frac{\rho(\rho - 1 - \log \rho)}{(1-\rho)^2} n + \mathcal{O}(1) \ \ \text{für festes } \rho = \frac{p}{n} \text{ mit } 0 < \rho < 1\,.} \tag{5.64c}$$

bzw. aus (5.62) für die Erwartungswerte der Anzahl der Vergleiche

Korollar 5.3.6. *Der Erwartungswert $E_{n,p}$ der Anzahl der Vergleiche beim Finden von p Elementen in einem Feld der Größe n durch Multiple Quickselect mit Median-of-Three Partition ist asymptotisch gegeben durch*

$$\boxed{\begin{aligned} E_{n,p} = {} & \left(\frac{12}{7}H_p + \frac{37p^7+1036p^6+11914p^5+72520p^4+250453p^3+485884p^2+483516p+246960}{49\,(p+7)^{\underline{7}}}\right) n \\ & + \mathcal{O}\left(\log n\right) \quad \text{für } p \text{ fix}\,, \end{aligned}} \tag{5.65a}$$

$$\boxed{E_{n,p} = \frac{12}{7} n \log n + \frac{84\gamma - 173}{49} n + \frac{12}{7}\log n + \frac{51}{49} + \frac{12}{7}\gamma + \mathcal{O}\left(\frac{\log n}{n}\right) \quad \text{für } l = n - p \text{ fix}\,,} \tag{5.65b}$$

$$\boxed{\begin{aligned} E_{n,p} = {} & \frac{12}{7} n \log n + \frac{168\rho\log\rho + 84\log\rho + 84\gamma\rho^2 - 168\gamma\rho + 84\gamma - 215\rho^2 + 178\rho + 37}{49(1-\rho)^2} n \\ & + \mathcal{O}(\log n) \quad \text{für festes } \rho = \frac{p}{n} \text{ mit } 0 < \rho < 1\,. \end{aligned}} \tag{5.65c}$$

5.4 Bestimmung der betrachteten Größen im Binärbaummodell

5.4.1 Die Anzahl der Durchläufe

Bezeichnen wir mit $\widetilde{\varphi}_{n,p,m}$ die Anzahl der Paare aus einem Binärbaum der Größe n und einer Markierung von genau p Knoten, sodaß durch den Multiple Quickselect Algorithmus genau m Durchläufe gemacht werden um alle diese markierten Knoten zu finden.

Aufteilung des Baumes durch die Wurzel k und Unterscheidung in die Fälle, bei denen die Wurzel einen der p markierten Knoten darstellt oder nicht, führt zu folgender Rekursion für $\widetilde{\varphi}_{n,p,m}$:

$$\begin{aligned} \widetilde{\varphi}_{n,p,m} &= \sum_{k=1}^{n} \left(\sum_{i=0}^{p} \sum_{l=0}^{m-1} \widetilde{\varphi}_{k-1,i,l}\, \widetilde{\varphi}_{n-k,p-i,m-1-l} \right. \\ & \qquad \left. + \sum_{i=0}^{p-1} \sum_{l=0}^{m-1} \widetilde{\varphi}_{k-1,i,l}\, \widetilde{\varphi}_{n-k,p-1-i,m-1-l} \right) \quad \text{für } n \geq 1 \text{ und } p \geq 1\,, \\ \widetilde{\varphi}_{n,0,m} &= \begin{cases} B_n & \text{wenn } m = 0\,, \\ 0 & \text{wenn } m \neq 0\,. \end{cases} \end{aligned} \tag{5.66}$$

Dabei bezeichnet B_n wieder die Anzahl der Binärbäume der Größe n, ist also gegeben durch

$$B_n = \frac{1}{n+1}\binom{2n}{n}\,. \tag{5.67}$$

Für die gewöhnliche erzeugende Funktion $B(z) = \sum_{n\geq 0} B_n z^n$ der B_n gilt

$$B(z) = \frac{1-\sqrt{1-4z}}{2z}\,. \tag{5.68}$$

Weiters verwenden wir im folgenden noch die Abkürzungen

$$X = \sqrt{1-4z}\,, \quad Y = \sqrt{1-4z(1+u)}\,. \tag{5.69}$$

Definieren wir die erzeugende Funktion $\Phi(z,u,v)$ mittels

$$\Phi(z,u,v) = \sum_{n\ge 0}\sum_{p\ge 0}\sum_{m\ge 0} \widetilde{\varphi}_{n,p,m} z^n u^p v^m\,, \tag{5.70}$$

dann ergibt sich aus (5.66) durch Multiplikation mit $z^n u^p v^m$ und Aufsummieren über $n \ge 1$ und $p \ge 1$ die Funktionalgleichung

$$\Phi(z,u,v) = zv(1+u)\Phi^2(z,u,v) + v + (1-v)B(z)\,. \tag{5.71}$$

Zum Erhalt der gesuchten Erwartungswerte $E_{n,p}$ definieren wir uns

$$G(z,u) = \frac{\partial}{\partial v}\Phi(z,u,v)\Big|_{v=1}\,, \tag{5.72}$$

denn es gilt dann die Beziehung

$$E_{n,p} = \frac{1}{B_n \binom{n}{p}}[z^n u^p]G(z,u)\,. \tag{5.73}$$

Differenzieren von Gleichung (5.71) nach v und Auswerten bei $v = 1$ ergibt unter Berücksichtigung der Beziehung

$$\Phi(z,u,1) = \sum_{n\ge 0}\sum_{p\ge 0}\sum_{m\ge 0} \widetilde{\varphi}_{n,p,m} z^n u^p = \sum_{n\ge 0}\sum_{p\ge 0}\binom{n}{p} B_n z^n z^p = B(z(1+u)) \tag{5.74}$$

folgende Funktionalgleichung für $G(z,u)$:

$$G(z,u) = 2z(1+u)B(z(1+u))G(z,u) + z(1+u)B^2(z(1+u)) + 1 - B(z)\,. \tag{5.75}$$

Lösen dieser Funktionalgleichung ergibt unter Verwendung der in (5.69) definierten abkürzenden Schreibweisen

$$G(z,u) = \frac{2(X-Y)}{(X+1)(Y+1)Y} = \frac{2}{Y} - \frac{2}{Y+1} - \frac{2}{Y(X+1)}\,. \tag{5.76}$$

Die für das Ablesen dieses Ausdrucks notwendigen Beziehungen wollen wir im folgenden herleiten.

Sofort erhält man

$$[z^n u^p]\frac{1}{Y} = [z^n u^p]\left(1-4z(1+u)\right)^{-\frac{1}{2}} = [z^n u^p]\sum_{k\ge 0}\binom{2k}{k}(z(1+u))^k = \binom{2n}{n}\binom{n}{p}\,. \tag{5.77}$$

Weiters erhalten wir

$$[z^n u^p]\frac{1}{Y+1} = [z^n u^p]\frac{1-\sqrt{1-4z(1+u)}}{4z(1+u)} = -[z^n u^p]\frac{\sqrt{1-4z(1+u)}}{4z(1+u)}$$
$$= \frac{1}{2}[z^{n+1}u^p]\sum_{k\geq 1}\frac{1}{k}\binom{2(k-1)}{k-1}z^k(1+u)^{k-1} = \frac{1}{2(n+1)}\binom{2n}{n}\binom{n}{p}. \tag{5.78}$$

Um zu den Koeffizienten des Ausdrucks $\frac{1}{(X+1)Y}$ zu gelangen, können wir wie folgt vorgehen. Zunächst spalten wir den Ausdruck auf in

$$[z^n u^p]\frac{1}{(X+1)Y} = [z^n u^p]\frac{1-\sqrt{1-4z}}{4z\sqrt{1-4z(1+u)}}$$
$$= \frac{1}{4}[z^{n+1}u^p]\frac{1}{\sqrt{1-4z(1+u)}} - \frac{1}{4}[z^{n+1}u^p]\frac{\sqrt{1-4z}}{\sqrt{1-4z(1+u)}}. \tag{5.79}$$

Für den einen Ausdruck ergibt sich unmittelbar aus (5.77)

$$\frac{1}{4}[z^{n+1}u^p]\frac{1}{\sqrt{1-4z(1+u)}} = \frac{1}{4}\binom{2(n+1)}{n+1}\binom{n+1}{p}, \tag{5.80}$$

während sich für den anderen

$$\frac{1}{4}[z^{n+1}u^p]\frac{\sqrt{1-4z}}{\sqrt{1-4z(1+u)}} = \frac{1}{4}[z^{n+1}u^p]\frac{1}{\sqrt{1-\frac{4uz}{1-4z}}} = \frac{1}{4}[z^{n+1}u^p]\sum_{k\geq 0}\binom{2k}{k}\frac{u^k z^k}{(1-4z)^k}$$
$$= \frac{1}{4}\binom{2p}{p}[z^{n+1-p}]\frac{1}{(1-4z)^p} = \frac{1}{4}\binom{2p}{p}4^{n+1-p}\binom{n}{p-1} \tag{5.81}$$

ergibt.

Daher erhält man

$$[z^n u^p]\frac{1}{(X+1)Y} = \frac{1}{4}\binom{2(n+1)}{n+1}\binom{n+1}{p} - \frac{1}{4}\binom{2p}{p}4^{n+1-p}\binom{n}{p-1}. \tag{5.82}$$

Kombination von (5.77), (5.78) und (5.82) ergibt aus (5.76) mittels Beziehung (5.73) schließlich

Satz 5.4.1. *Der Erwartungswert $E_{n,p}$ der Anzahl der Durchläufe beim Finden von p Elementen im Binärbaummodell der Größe n durch Multiple Quickselect ist gegeben durch*

$$\boxed{E_{n,p} = \frac{p(n+1)4^{n+1-p}}{2(n+1-p)}\frac{\binom{2p}{p}}{\binom{2n}{n}} - \frac{p(2n+1)}{n+1-p} \quad \text{für } 1 \leq p \leq n\,.} \tag{5.83}$$

5.4.2 Die Anzahl der Vergleiche

Bezeichne nun $\widetilde{\varphi}_{n,p,m}$ die Anzahl der Paare aus einem Binärbaum der Größe n und einer Markierung von genau p Knoten, sodaß durch den Multiple Quickselect Algorithmus genau m Schlüsselvergleiche gemacht werden um alle diese markierten Knoten zu finden.

Aufteilung des Baumes durch die Wurzel k und Unterscheidung in die Fälle in der die Wurzel einen der p markierten Knoten darstellt oder nicht, führt wegen der $n-1$ Schlüsselvergleiche bei jedem Partitionierungsschritt des Algorithmus zu folgender Rekursion für $\widetilde{\varphi}_{n,p,m}$:

$$\begin{aligned}\widetilde{\varphi}_{n,p,m} &= \sum_{k=1}^{n}\Bigg(\sum_{i=0}^{p}\sum_{l=0}^{m-n+1}\widetilde{\varphi}_{k-1,i,l}\,\widetilde{\varphi}_{n-k,p-i,m-n+1-l} \\ &\qquad + \sum_{i=0}^{p-1}\sum_{l=0}^{m-n+1}\widetilde{\varphi}_{k-1,i,l}\,\widetilde{\varphi}_{n-k,p-1-i,m-n+1-l}\Bigg) \quad \text{für } n \geq 1 \text{ und } p \geq 1\,, \\ \widetilde{\varphi}_{n,0,m} &= \begin{cases} B_n & \text{wenn } m = 0\,, \\ 0 & \text{wenn } m \neq 0\,. \end{cases}\end{aligned} \tag{5.84}$$

Sei die erzeugende Funktion $\Phi(z,u,v)$ wie in (5.70) definiert, dann ergibt sich aus (5.84) durch Multiplikation mit $z^n u^p v^m$ und Aufsummieren über $n \geq 1$ und $p \geq 1$ die Funktionalgleichung

$$\Phi(z,u,v) = z(1+u)\Phi^2(zv,u,v) + B(z) - zB^2(zv)\,. \tag{5.85}$$

Zum Erhalt der gesuchten Erwartungswerte $E_{n,p}$ sei die Funktion $G(z,u)$ definiert wie in (5.72).

Differenzieren wir nun Gleichung (5.85) nach v und werten bei $v = 1$ aus, dann erhalten wir eine einfache Funktionalgleichung für $G(z,u)$, deren Lösung

$$G(z,u) = \frac{4}{Y+1} + \frac{1}{Y^2} - \frac{4}{Y} + \frac{4}{Y(X+1)} - \frac{1}{XY} \tag{5.86}$$

lautet.

Für das Ablesen der Koeffizienten benötigt man neben den Beziehungen (5.77), (5.78) und (5.82) noch

$$[z^n u^p]\frac{1}{Y^2} = [z^n u^p]\frac{1}{1-4z(1+u)} = 4^n\binom{n}{p} \tag{5.87}$$

und

$$\begin{aligned}[z^n u^p]\frac{1}{XY} &= [z^n u^p]\frac{1}{(1-4z)\sqrt{1-\frac{4uz}{1-4z}}} = [z^n u^p]\frac{1}{1-4z}\sum_{k\geq 0}\binom{2k}{k}\frac{z^k u^k}{(1-4z)^k} \\ &= \binom{2p}{p}[z^n]\frac{z^p}{(1-4z)^{p+1}} = \binom{2p}{p}[z^{n-p}]\frac{1}{(1-4z)^{p+1}} = \binom{2p}{p}4^{n-p}\binom{n}{p}\,.\end{aligned} \tag{5.88}$$

Damit erhält man aus (5.86) mittels Beziehung (5.73)

Satz 5.4.2. *Der Erwartungswert $E_{n,p}$ der Anzahl der Vergleiche beim Finden von p Elementen im Binärbaummodell der Größe n durch Multiple Quickselect ist gegeben durch*

$$\boxed{E_{n,p} = (n+1)4^n \frac{1}{\binom{2n}{n}} - \frac{(n+1)(n+3p+1)4^{n-p}}{n+1-p}\frac{\binom{2p}{p}}{\binom{2n}{n}} + \frac{2p(2n+1)}{n+1-p} \quad \text{für } 1 \leq p \leq n\,.} \tag{5.89}$$

5.4.3 Asymptotische Betrachtungen

Wir untersuchen nun das Verhalten der Größen dieses Abschnittes für $n \to \infty$ und unterscheiden dabei die Fälle p fest, $l = n - p$ fest und ein festes Verhältnis $\rho = \frac{p}{n}$ mit $0 < \rho < 1$.

Unter Verwendung von Stirling's Approximation für die Faktoriellen (3.22) bekommen wir aus (5.83) für den Erwartungswert der Anzahl der Durchläufe

Korollar 5.4.3. *Der Erwartungswert $E_{n,p}$ der Anzahl der Durchläufe beim Finden von p Elementen im Binärbaummodell der Größe n durch Multiple Quickselect ist asymptotisch gegeben durch*

$$\boxed{E_{n,p} = \frac{2p\binom{2p}{p}\sqrt{\pi}}{4^p}\sqrt{n} + \mathcal{O}(1) \quad \text{für } p \text{ fix}\,,} \tag{5.90a}$$

$$\boxed{E_{n,p} = n + \mathcal{O}(1) \quad \text{für } l = n-p \ \text{ fix}\,,} \tag{5.90b}$$

$$\boxed{E_{n,p} = \frac{2\sqrt{\rho}}{1+\sqrt{\rho}} n + \mathcal{O}(1) \quad \text{für festes } \rho = \frac{p}{n} \ \text{ mit } 0 < \rho < 1\,.} \tag{5.90c}$$

bzw. aus (5.89) für den Erwartungswert der Anzahl der Vergleiche

Korollar 5.4.4. *Der Erwartungswert $E_{n,p}$ der Anzahl der Vergleiche beim Finden von p Elementen im Binärbaummodell der Größe n durch Multiple Quickselect ist asymptotisch gegeben durch*

$$\boxed{E_{n,p} = \left(1 - \frac{\binom{2p}{p}}{4^p}\right)\sqrt{\pi}\, n^{3/2} + \mathcal{O}\left(\sqrt{n}\right) \quad \text{für } p \text{ fix}\,,} \tag{5.91a}$$

$$\boxed{E_{n,p} = \sqrt{\pi}\, n^{3/2} - 3n + \mathcal{O}\left(\sqrt{n}\right) \quad \text{für } l = n-p \text{ fix}\,,} \tag{5.91b}$$

$$\boxed{E_{n,p} = \sqrt{\pi}\, n^{3/2} - \frac{1+\sqrt{\rho}+4\rho}{(1+\sqrt{\rho})\sqrt{\rho}}\, n + \mathcal{O}\left(\sqrt{n}\right) \quad \text{für festes } \rho = \frac{p}{n} \ \text{ mit } 0 < \rho < 1\,.} \tag{5.91c}$$

<table>
<tr><th colspan="3"></th><th>p fest</th><th>$p \sim \rho n$</th><th>$l := n - p$ fest</th></tr>
<tr><td rowspan="4">Multiple quickselect</td><td rowspan="2">Durchläufe</td><td>$E_{n,p}$</td><td>$2p \log n$</td><td>$\frac{2\rho(\rho-1-\log\rho)}{(1-\rho)^2}$</td><td>$n$</td></tr>
<tr><td>$V_{n,p}$</td><td>$2p \log n$</td><td>$\left(-\frac{4\rho^2(\rho+3)\log^2\rho}{(1-\rho)^5} - \frac{2\rho(3\rho^2+12\rho+1)\log\rho}{(1-\rho)^4} - \frac{16\rho^2}{(1-\rho)^3}\right)n$</td><td>$\frac{2}{9}l$</td></tr>
<tr><td rowspan="2">Vergleiche</td><td>$E_{n,p}$</td><td>$(2H_p+1)n$</td><td>$2n\log n$</td><td>$2n\log n$</td></tr>
<tr><td>$V_{n,p}$</td><td>$\left(\frac{21p^3+80p^2+81p-2}{3(p+1)^2(p+2)} - 4H_p^{(2)}\right)n^2$</td><td>$\left(7-\frac{2}{3}\pi^2\right)n^2$</td><td>$\left(\frac{7l^4+70l^3+245l^2+368l+168}{(l+4)^{\underline{4}}} - \frac{2}{3}\pi^2\right)n^2$</td></tr>
<tr><td rowspan="2">MQS mit Median-of-Three Partition</td><td>Durchläufe</td><td>$E_{n,p}$</td><td>$\frac{12}{7}p\log n$</td><td>$\frac{12}{7}\frac{\rho(\rho-1-\log\rho)}{(1-\rho)^2}n$</td><td>$\frac{6}{7}n$</td></tr>
<tr><td>Vergleiche</td><td>$E_{n,p}$</td><td>$\left(\frac{12}{7}H_p+\frac{r(p)}{49\,(p+7)^{\underline{7}}}\right)n$</td><td>$\frac{12}{7}n\log n$</td><td>$\frac{12}{7}n\log n$</td></tr>
<tr><td rowspan="2">Binärbaum-modell</td><td>Durchläufe</td><td>$E_{n,p}$</td><td>$\frac{2p\binom{2p}{p}\sqrt{\pi}}{4^p}\sqrt{n}$</td><td>$\frac{2\sqrt{\rho}}{1+\sqrt{\rho}}n$</td><td>$n$</td></tr>
<tr><td>Vergleiche</td><td>$E_{n,p}$</td><td>$\left(1-\frac{\binom{2p}{p}}{4^p}\right)\sqrt{\pi}n^{3/2}$</td><td>$\sqrt{\pi}\,n^{3/2}$</td><td>$\sqrt{\pi}\,n^{3/2}$</td></tr>
</table>

mit $r(p) := 37p^7+1036p^6+11914p^5+72520p^4+250453p^3+485884p^2+483516p+246960$

Tabelle 5.1: Asymptotische Resultate zu allen expliziten Formeln dieses Kapitels

Literaturverzeichnis

[1] Milton Abramowitz und Irene A. Stegun, Herausgeber. *Handbook of Mathematical Functions*. Dover Publications, Inc., New York, neunte Auflage, 1970.

[2] Heinz Bauer. *Wahrscheinlichkeitstheorie*. Walter de Gruyter, Berlin, vierte Auflage, 1991.

[3] Kai Lai Chung. *A course in Probability Theory*. Academic Press, San Diego, 1974.

[4] Louis Comtet. *Advanced Combinatorics*. Reidel, Dordrecht, 1974.

[5] Luc Devroye. Limit Laws for Local Counters in Random Binary Search Trees. *Random Structures and Algorithms*, 2:303–315, 1991.

[6] William F. Eddy und Mark J. Schervish. How many comparisons does Quicksort need ? *Journal of Algorithms*, 19:402–431, 1995.

[7] Philippe Flajolet und Andrew Odlyzko. Singularity Analysis of Generating Functions. *SIAM Journal of Discrete Mathematics*, 3:216–240, 1990.

[8] Philippe Flajolet und Michèle Soria. General combinatorial schemas: Gaussian limit distributions and exponential tails. *Discrete Mathematics*, 114:159–180, 1993.

[9] Gaston H. Gonnet und Ricardo Baeza-Yates. *Handbook of Algorithms and Data Structures*. Addison-Wesley, Reading, Massachusetts, zweite Auflage, 1991.

[10] Ronald L. Graham, Donald E. Knuth, und Oren Patashnik. *Concrete Mathematics*. Addison-Wesley, Reading, Massachusetts, 1989.

[11] Daniel H. Greene und Donald E. Knuth. *Mathematics for the Analysis of Algorithms*. Birkhäuser, Boston, zweite Auflage, 1982.

[12] Heinz-Richard Halder und Werner Heise. *Einführung in die Kombinatorik*. Carl Hanser Verlag, München, 1976.

[13] P. Hennequin. Combinatorial analysis of Quicksort algorithm. *RAIRO, Theoretical Informatics and Applications*, 23:317–333, 1989.

[14] C. A. R. Hoare. Find (algorithm 65). *Communications of the ACM*, 4:321,322, 1961.

[15] C. A. R. Hoare. Quicksort (algorithm 64). *Communications of the ACM*, 4:321, 1961.

[16] Hsien-Kuei Hwang. On the average shape of Binary Search Trees. Unveröffentlichtes Manuskript, 1990.

[17] Arne Jonassen und Ole-Johan Dahl. Analysis of an Algorithm for Priority Queue Administration. *BIT*, 15:409–422, 1975.

[18] Erich Kamke. *Differentialgleichungen - Lösungsmethoden und Lösungen*, Band 1. Teubner, Stuttgart, zehnte Auflage, 1983.

[19] Peter Kirschenhofer. On the Height of Leaves in Binary Trees. *Journal of Combinatorics, Information & System Sciences*, 8(1):44–60, 1983.

[20] Peter Kirschenhofer. Some new Results on the Average Height of Binary Trees. *Ars Combinatoria*, 16 A:255–260, 1983.

[21] Peter Kirschenhofer, Conrado Martínez, und Helmut Prodinger. Analysis of Hoare's Find Algorithm with Median-of-three partition. *Random Structures & Algorithms*, 10:143–156, 1997.

[22] Peter Kirschenhofer und Helmut Prodinger. Comparisons in Hoare's Find Algorithm. wird erscheinen in *Combinatorics, Probability & Computing*, 1996.

[23] Donald E. Knuth. *The Art of Computer Programming*, Band 3: Sorting and Searching. Addison-Weslay, Reading, Massachusetts, 1973.

[24] Janice Lent. *Probabilistic Analysis of some Searching and Sorting Algorithms*. Dissertation, George Washington University, 1996.

[25] Janice Lent und Hosam Mahmoud. Average-case analysis of multiple Quickselect: An Algorithm for finding order statistics. *Statistics & Probability Letters*, 28:299–310, 1996.

[26] Leonard Lewin, Herausgeber. *Structural Properties of Polylogarithms*. American Mathematical Society, 1991.

[27] Hosam Mahmoud. The Expected Distribution of Degrees in Random Binary Search Trees. *The Computer Journal*, 29:36–37, 1986.

[28] Hosam Mahmoud. The Joint Distribution of the Three Types of Nodes in Uniform Binary Trees. *Algorithmica*, 13:313–323, 1995.

[29] Hosam Mahmoud, Reza Modarres, und Robert Smythe. Analysis of Quickselect: An Algorithm for Order Statistics. *RAIRO, Theoretical Informatics and Applications*, 29:255–276, 1995.

[30] Conrado Martínez und Helmut Prodinger. On the Number of Descendants and Ascendants in Random Search Trees. zur Veröffentlichung vorgelegt, 1996.

[31] A. Meir und J. W. Moon. On the Altitude of Nodes in Random Trees. *Canadian Journal of Mathematics*, 30:997–1015, 1978.

[32] Olli Nevalainen und Jukka Teuhola. The Efficiency of two indexed Priority Queue Algorithms. *BIT*, 18:320–333, 1978.

[33] Alois Panholzer. Average-Case-Analyse verallgemeinerter QUICKSORT und FIND Algorithmen. Diplomarbeit, Technische Universität Wien, 1995.

[34] Marko Petkovšek, Herbert S. Wilf, und Doron Zeilberger. *A=B*. A. K. Peters, Wellesley, 1996.

[35] Patricio V. Poblete. The analysis of heuristics for search trees. *Acta Informatica*, 30:233–248, 1993.

[36] Patricio V. Poblete und J. Ian Munro. The Analysis of a Fringe Heuristic for Binary Search Trees. *Journal of Algorithms*, 6:336–350, 1985.

[37] Helmut Prodinger. Multiple Quickselect - Hoare's Find algorithm for several elements. *Information Processing Letters*, 56:123–129, 1995.

[38] Helmut Prodinger. A Note of the Distribution of the three Types of Nodes in Uniform Binary Trees. `http://info.tuwien.ac.at/theoinf/proding.htm`, 1996.

[39] Mireille Régnier. A limiting distribution for Quicksort. *RAIRO, Theoretical Informatics and Applications*, 23:335–343, 1989.

[40] Uwe Rösler. A limit theorem for Quicksort. *RAIRO, Theoretical Informatics and Applications*, 25:85–100, 1991.

[41] Günter Rote. Binary Trees having a given Number of Nodes with 0, 1, and 2 Children. `http://opt.math.tu-graz.ac.at`, 1997.

[42] Robert Sedgewick. *Quicksort*. Garland, New York, 1978.

[43] Robert Sedgewick. *Algorithmen*. Addison-Wesley, Bonn, 1991.

[44] Robert Sedgewick und Philippe Flajolet. *An Introduction to the Analysis of Algorithms*. Addison-Wesley, Reading, Massachusetts, 1996.

[45] Renzo Sprugnoli. Counting Labels in Binary Trees. *BIT*, 30:62–69, 1990.

[46] William Tutte. The number of planted plane trees with a given partition. *American Mathematical Monthly*, 71:272–277, 1964.

[47] A. Walker und D. Wood. Locally balanced binary trees. *The Computer Journal*, 19(4):322–325, 1976.

[48] Horst Wenzel und Peter Meinhold. *Gewöhnliche Differentialgleichungen*. Teubner, Stuttgart, 1994.

[49] Herbert S. Wilf. *Generatingfunctionology*. Academic Press, San Diego, 1990.

[50] D. A. Zave. A Series Expansion Involving the Harmonic Numbers. *Information Processing Letters*, 5:75–77, 1976.

Curriculum vitae

Name:	Alois Panholzer
Adresse:	Rudolf-Köppl-Gasse 5/3/3 1220 Wien
Geburtsdaten:	8. März 1971 in Linz
Familienstand:	ledig
Ausbildung:	
1977-1981	Volksschule, 4113 St. Martin i. Mkr.
1981-1985	Hauptschule, 4113 St. Martin i. Mkr.
1985-1991	Höhere Technische Lehranstalt Wels für Chemische Betriebstechnik 11. 6. 1991 Matura mit Auszeichnung
1991-1996	Technische Universität Wien - Studienrichtung Technische Mathematik 9. 1. 1996 Diplom mit Auszeichnung
Seit 1996	Doktoratsstudium der Technischen Wissenschaften
Wissenschaftliche Arbeiten:	A. Panholzer und H. Prodinger. *Descendants and Ascendants in Binary Trees*, zur Veröffentlichung vorgelegt, 1997.
	A. Panholzer und H. Prodinger. *A Generating Functions Approach for the Analysis of Grand Averages for Multiple Quickselect*, zur Veröffentlichung vorgelegt, 1997.