DIE BESTIMMUNG GEWISSER PARAMETER BEI BINÄREN BÄUMEN MIT
HILFE ANALYTISCHER METHODEN

Helmut Prodinger

Abstract. This paper deals with the average number of nodes with a special property in binary trees with n nodes. Generating functions are set up and considered as analytic functions. A detailed singularity analysis allows to get asymptotic formulas for the considered numbers. The local expansions are derived by use of the Mellin transform. The asymptotic expansion involves periodic terms; the Fourier coefficients are computed in terms of Riemann's ζ-functions etc.

1. EINLEITUNG

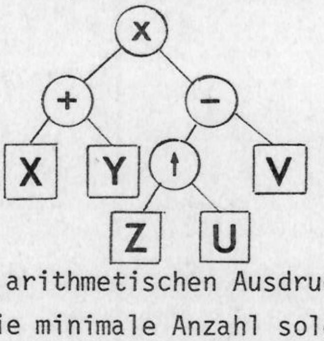
Ein *binärer Baum* ist entweder ein Blatt (□) oder ein (innerer) Knoten (○) mit einem linken und rechten Unterbaum; diese sind selbst binäre Bäume. Die *Familie der binären Bäume* $\mathcal{B}$ erfüllt also die formale Gleichung

$$\mathcal{B} = \square + \begin{array}{c}\circ\\ \diagup\quad\diagdown\\ \mathcal{B}\qquad\mathcal{B}\end{array} \tag{1}$$

(vgl. [3]). Ist b_n die Anzahl der binären Bäume mit n Knoten und $B(z) = \sum_{n\geqslant 0} b_n z^n$, so ergibt sich aus (1)

$$B(z) = 1 + z(B(z))^2 \quad \text{bzw.} \quad B(z) = \frac{1-\sqrt{1-4z}}{2z} \quad \text{und} \quad b_n = \frac{1}{n+1}\binom{2n}{n}. \tag{2}$$

Ein binärer Baum kann verwendet werden, um einen arithmetischen Ausdruck darzustellen. Ein Beispiel illustriert das am besten:



$\longleftrightarrow$ $((X+Y)\times((Z\uparrow U)-V))$

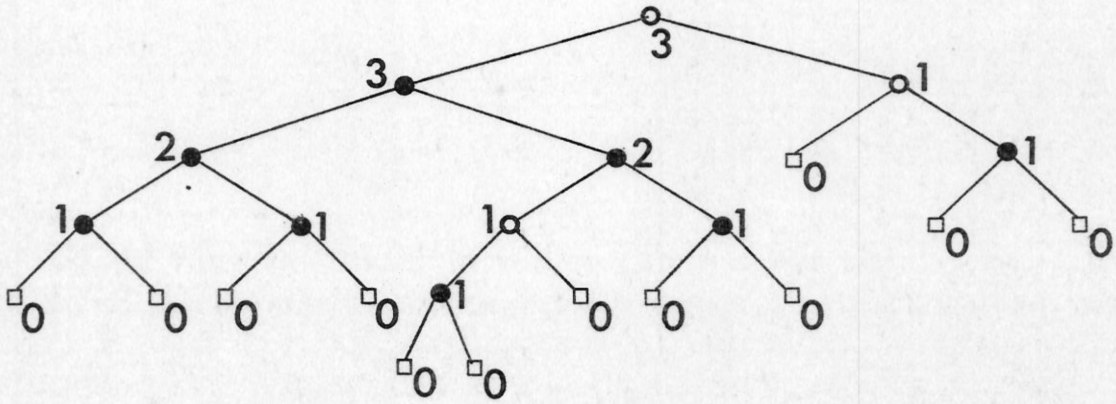
Um einen arithmetischen Ausdruck auszuwerten, benötigt man eine Anzahl von *Hilfsregistern*. Die minimale Anzahl solcher Hilfsregister hängt nur vom Baum t ab und wird mit Reg(t) abgekürzt. Interessanterweise ist dieser Parameter nicht nur in der Informatik von Bedeutung, sondern auch in den Naturwissenschaften. Der Leser sei auf [3,5,6,8,10] verwiesen.

Die Registerfunktion ist in induktiver Weise wie folgt definiert:

$$\mathrm{Reg}(\square) = 0$$

$$\mathrm{Reg}\Big(\overset{\circ}{\bigwedge}_{t_1\ \ t_2}\Big) = \begin{cases} \max\{\mathrm{Reg}(t_1), \mathrm{Reg}(t_2)\} & \text{falls } \mathrm{Reg}(t_1) \neq \mathrm{Reg}(t_2) \\ 1 + \mathrm{Reg}(t_1) & \text{falls } \mathrm{Reg}(t_1) = \mathrm{Reg}(t_2) \end{cases} \tag{3}$$

Diese Definition erlaubt es, von unten nach oben fortschreitend, jedem Knoten bzw. Blatt eine nichtnegative ganze Zahl zuzuordnen, nämlich die Registerfunktion des von dem Knoten induzierten Unterbaumes. Die Registerfunktion kann man dann an der Wurzel ablesen:



(4)

Der Mittelwert D_n der Registerfunktion, wo alle Bäume mit n Knoten als gleich wahrscheinlich angesehen werden, ist in letzter Zeit oft studiert worden: [3,5,6,8,10]. Es gilt:

$$D_n = \log_4 n + D(\log_4 n) + O\left(\frac{\log^{\star} n}{n}\right) ; \tag{5}$$

$\star$ steht für eine positive Zahl, D(x) ist stetig und periodisch mit Periode 1. Schreibt man $D(x) = \sum_{k\in\mathbb{Z}} d_k e^{2k\pi ix}$, so gilt:

$$d_0 = -\frac{1}{2} - \frac{\gamma}{2\log 2} - \frac{1}{\log 2} + \log_2 2\pi ,$$

$$d_k = \frac{1}{\log 2}\zeta(\chi_k)\,\Gamma(\chi_k/2)(\chi_k-1) , \; k \neq 0 , \; \chi_k = \frac{2k\pi i}{\log 2} .$$

Wir wollen uns hier jedoch einer anderen Fragestellung zuwenden: Gewisse Knoten veranlassen die Registerfunktion zu wachsen; in (4) sind diese *kritischen* Knoten hervorgehoben. Es wird im folgenden gezeigt, daß für K_n, die mittlere Anzahl von kritischen Knoten in einem binären Baum mit n Knoten, folgendes gilt:

$$K_n = \frac{n}{3} + \frac{1}{12}\log_2 n + K(\log_4 n) + o(1),$$

wo K eine periodische Funktion mit Periode 1 ist.

Der erste Term n/3 scheint bereits in impliziter Form in [11] auf.

2. ERZEUGENDE FUNKTIONEN

Wir bezeichnen mit $\mathcal{R}_p$ bzw. $\mathcal{S}_p$ die Familie der binären Bäume mit Registerfunktion $=p$ bzw. $\geqslant p$, sowie mit $R_p(z)$ und $S_p(z)$ die entsprechenden (gewöhnlichen) erzeugenden Funktionen. Weiters sei $\varepsilon = \sqrt{1-4z}$, sowie $z = u/(1+u)^2$, d.h. $u = (1-\varepsilon)/(1+\varepsilon)$.

<u>LEMMA 1.</u>

$$R_p = \frac{1-u^2}{u} \cdot \frac{u^{2^p}}{1-u^{2^{p+1}}} \tag{6}$$

$$S_p = \frac{1-u^2}{u} \cdot \frac{u^{2^p}}{1-u^{2^p}} \tag{7}$$

$$\frac{R_p}{R_{p-1}^2} = \frac{u}{1-u^2} \cdot \frac{1-u^{2^p}}{1+u^{2^p}} \tag{8}$$

<u>Beweis.</u> Der folgende Beweis wurde gemeinsam mit *P. KIRSCHENHOFER* gefunden und gestattet eine wesentlich schnellere Herleitung von (6)-(8) als dies bisher möglich war. Es wird nur (7) bewiesen; (6) und (8) folgen dann, weil $S_p - S_{p+1} = R_p$. Aus (3) folgt unmittelbar

$$\mathcal{R}_p = \underset{\mathcal{R}_{p-1}\quad\mathcal{R}_{p-1}}{\overset{\circ}{\diagup\;\diagdown}} + \underset{\mathcal{R}_p\quad\mathcal{B}\setminus\mathcal{S}_p}{\overset{\circ}{\diagup\;\diagdown}} + \underset{\mathcal{B}\setminus\mathcal{S}_p\quad\mathcal{R}_p}{\overset{\circ}{\diagup\;\diagdown}}\,, \quad \mathcal{R}_0 = \Box \,, \tag{9}$$

$$\mathcal{S}_p = \underset{\mathcal{S}_{p-1}\quad\mathcal{S}_{p-1}}{\overset{\circ}{\diagup\;\diagdown}} + \underset{\mathcal{S}_p\quad\mathcal{B}\setminus\mathcal{S}_{p-1}}{\overset{\circ}{\diagup\;\diagdown}} + \underset{\mathcal{B}\setminus\mathcal{S}_{p-1}\quad\mathcal{S}_p}{\overset{\circ}{\diagup\;\diagdown}}\,, \qquad \mathcal{S}_0 = \mathcal{B}\,. \tag{10}$$

Aus (10) ergibt sich

$$S_p = zS_{p-1}^2 + 2zS_p(B-S_{p-1}),$$

bzw.

$$S_p = \frac{zS_{p-1}^2}{1-2zB+2zS_{p-1}} .$$

Wir beachten $\varepsilon = 1-2zB$ und nehmen den Kehrwert:

$$\frac{1}{S_p} = \frac{\varepsilon}{z}\,\frac{1}{S_{p-1}^2} + \frac{2}{S_{p-1}} .$$

Wir multiplizieren mit ε/z und addieren 1:

$$\frac{\varepsilon}{z}\,\frac{1}{S_p} + 1 = \left(\frac{\varepsilon}{z}\,\frac{1}{S_{p-1}} + 1\right)^2 = \left(\frac{\varepsilon}{zB} + 1\right)^{2^p} .$$

Man rechnet leicht nach, daß $\varepsilon/z = (1-u^2)/u$ und $1 + \varepsilon/zB = 1/u$ ist, und erhält

$$\frac{1-u^2}{u}\frac{1}{S_p} = -1 + \frac{1}{u^{2^p}} = \frac{1-u^{2^p}}{u^{2^p}} ;$$

das ist (7). □

Im folgenden schreiben wir für den Koeffizienten von z^n in $f(z)$ immer $[z^n]f(z)$ u.ä. Außerdem vereinbaren wir eine Kurzschrift: $\overline{f}(z) = \frac{\partial}{\partial u} f(z,u)\Big|_{u=1}$.

Sei $U_p = R_{p-1}/R_p$; es ist dann $U_1 = (1-2z)/z$. Weiters sei $[z^n u^m] V_p(z,u)$ die Anzahl der binären Bäume mit n Knoten, Registerfunktion p und m kritischen Knoten. Außerdem sei $Q_p = V_{p-1}/V_p$. Die uns interessierende Größe K_n ist dann

$$K_n = \frac{[z^n] \sum_{p\geq 0} \overline{V}_p(z)}{[z^n]\, B(z)} . \tag{11}$$

<u>LEMMA 2.</u> $\overline{V}_0 = 0$, $\overline{V}_1 = z/(1-2z)$; <u>für</u> $p \geq 2$ <u>gilt</u>:

$$\overline{V}_p = R_p - R_p \sum_{j=2}^{p} \frac{R_j 2^j}{R_{j-1}^2} \left[\sum_{k=1}^{j-1} \frac{R_k}{2^k} + 1 - \frac{1}{2z} \right] .$$

<u>Beweis.</u> Von (9) erhält man

$$V_0 = 1 \quad \text{und} \quad V_p = 2z\, V_p \sum_{j<p} V_j + zu\, V_{p-1}^2 , \quad p \geq 1;$$

Wir dividieren durch V_p und subtrahieren die analoge Ungleichung für $p+1$:

$$0 = 2z\, V_p + zu \frac{V_p^2}{V_{p+1}} - zu \frac{V_{p-1}^2}{V_p} ;$$

$$0 = \frac{2}{u} + Q_{p+1} - Q_p^2 , \quad p \geq 1.$$

Wir beachten $Q_p(z,1) = U_p(z)$, differenzieren nach u und ersetzen u durch 1:

$$0 = -2 + \overline{Q}_{p+1} - 2\, \overline{Q}_p\, U_p , \quad p \geq 1.$$

Daraus erhält man

$$\frac{\overline{Q}_p}{2^p U_1 \ldots U_{p-1}} = \frac{\overline{Q}_{p-1}}{2^{p-1} U_1 \ldots U_{p-2}} + \frac{1}{2^{p-1} U_1 \ldots U_{p-1}} , \quad p \geq 2.$$

Es ist $V_1 = zu/(1-2z)$, daher $Q_1 = (1-2z)/zu$ und $\overline{Q}_1 = 2 - 1/z$. Weiters gilt $U_1 \ldots U_p = 1/R_p$. Demnach erhält man für $p \geq 2$:

$$\overline{Q}_p \frac{R_{p-1}}{2^p} = \overline{Q}_{p-1} \frac{R_{p-2}}{2^{p-1}} + \frac{R_{p-1}}{2^{p-1}} = \sum_{j=2}^{p} \frac{R_{j-1}}{2^{j-1}} + \overline{Q}_1 \frac{R_0}{2} ,$$

oder

$$\overline{Q}_p = \frac{2^p}{R_{p-1}} \left[\sum_{j=1}^{p-1} \frac{R_j}{2^j} + 1 - \frac{1}{2z} \right] .$$

Für $p \geqslant 1$ gilt $Q_p V_p = V_{p-1}$, und daher

$$\overline{Q}_p R_p + U_p \overline{V}_p = \overline{V}_{p-1} .$$

Diese Rekursion für $\overline{V}_p$ wird nach demselben Muster wie vorhin gelöst:

$$\overline{V}_p U_1 \dots U_p = \overline{V}_{p-1} U_1 \dots U_{p-1} - \overline{Q}_p R_p U_1 \dots U_{p-1} ,$$

oder für $p \geqslant 1$

$$\frac{\overline{V}_p}{R_p} = \frac{\overline{V}_{p-1}}{R_{p-1}} - \overline{Q}_p \frac{R_p}{R_{p-1}} = \frac{\overline{V}_1}{R_1} - \sum_{j=2}^{p} \overline{Q}_j \frac{R_j}{R_{j-1}} .$$

Hieraus ergibt sich die Behauptung unmittelbar.

<u>LEMMA 3.</u>

$$\eta_s := \sum_{j \geqslant s} 2^j \sum_{k \geqslant 1} (-1)^{k+1} u^{k2^j} = 2^s \frac{u^{2^s}}{1-u^{2^s}} .$$

<u>Beweis.</u> Wir setzen

$$\eta_s := \sum_{n \geqslant 1} \Theta_s(n) u^n \quad \text{mit} \quad \Theta_s(n) = \sum_{j \geqslant s; k \geqslant 1; k2^j = n} (-1)^{k+1} 2^j .$$

Sei nun $n = 2^m(1+2\lambda)$ mit $m \geqslant s$:

$$\Theta_s(n) = \sum_{j=s}^{m} 2^j (-1)^{1+n2^{-j}} = - \sum_{j=s}^{m-1} 2^j + 2^m = 2^s .$$

Also gilt

$$\eta_s = \sum_{m \geqslant s; \lambda \geqslant 0} 2^s u^{2^m(1+2\lambda)} = 2^s \sum_{k \geqslant 1} u^{k2^s} = 2^s \frac{u^{2^s}}{1-u^{2^s}} .$$

<u>SATZ 1.</u> <u>Die arithmetische Funktion</u> ω <u>sei definiert durch</u>

$\omega(n) = i$, <u>falls</u> $n = 2^m(1+2i)$

<u>Dann gilt</u>:

$$\overline{V}(z) = \frac{u(1+u)}{1-u} - 2 \frac{1-u^2}{u} \sum_{k \geqslant 1} \omega(k) u^k .$$

<u>Beweis.</u> Aus Lemma 2 erhalten wir

$$\overline{V} = \overline{V}_0 + \overline{V}_1 + \sum_{p \geqslant 2} \overline{V}_p = R_1 + \sum_{p \geqslant 2} R_p - A_1 - A_2$$

mit

$$A_1 = \sum_{p\geqslant 2} R_p \sum_{j=2}^{p} \frac{R_j 2^j}{R_{j-1}^2} \sum_{k=1}^{j-1} \frac{R_k}{2^k}$$

und

$$A_2 = (1 - \frac{1}{2z}) \sum_{p\geqslant 2} R_p \sum_{j=2}^{p} \frac{R_j 2^j}{R_{j-1}^2} .$$

Es ist $\sum_{p\geqslant 1} R_p = -1 + B = u$, sodaß wir uns A_2 zuwenden können:

$$A_2 = \frac{2u-(1+u)^2}{2u} \sum_{p\geqslant 2} \frac{1-u^2}{u} \frac{u^{2^p}}{1-u^{2^{p+1}}} \sum_{j=2}^{p} \frac{u}{1-u^2} \frac{1-u^{2^j}}{1+u^{2^j}} 2^j$$

$$= -\frac{1+u^2}{2u} \sum_{j\geqslant 2} \frac{1-u^{2^j}}{1+u^{2^j}} 2^j \sum_{p\geqslant j} \frac{u^{2^p}}{1-u^{2^{p+1}}}$$

$$= -\frac{1+u^2}{2u} \sum_{j\geqslant 2} \frac{1-u^{2^j}}{1+u^{2^j}} 2^j \frac{u^{2^j}}{1-u^{2^j}}$$

$$= -\frac{1+u^2}{2u} \sum_{j\geqslant 2} 2^j \sum_{k\geqslant 1} (-1)^{k+1} u^{k2^j} = -\frac{1+u^2}{2u} \eta_2$$

$$= -\frac{1+u^2}{2u} 4 \frac{u^4}{1-u^4} = -\frac{2u^3}{1-u^2}$$

Die Berechnung von A_1 ist ähnlich:

$$A_1 = \sum_{k\geqslant 1} \frac{1-u^2}{u} \frac{u^{2^k}}{1-u^{2^{k+1}}} \frac{1}{2^k} \sum_{j\geqslant k+1} \frac{u}{1-u^2} \frac{1-u^{2^j}}{1+u^{2^j}} 2^j \sum_{p\geqslant j} \frac{1-u^2}{u} \frac{u^{2^p}}{1-u^{2^{p+1}}}$$

$$= \frac{1-u^2}{u} \sum_{k\geqslant 1} \frac{u^{2^k}}{1-u^{2^{k+1}}} \frac{1}{2^k} \sum_{j\geqslant k+1} \frac{u^{2^j}}{1+u^{2^j}} 2^j$$

Die letzte Summe ist η_{k+1}; daher gilt:

$$A_1 = \frac{1-u^2}{u} \sum_{k\geqslant 1} \frac{u^{2^k}}{1-u^{2^{k+1}}} 2 \frac{u^{2^{k+1}}}{1-u^{2^{k+1}}}$$

$$= -\frac{2u^2}{1-u^2} + 2\frac{1-u^2}{u} \sum_{k\geqslant 0} \frac{u^{3.2^{k-1}}}{(1-u^{2^k})^2}$$

$$= -\frac{2u^2}{1-u^2} + 2\frac{1-u^2}{u} \sum_{k\geqslant 0;\lambda\geqslant 0} u^{3.2^{k-1}} (\lambda+1) u^{\lambda.2^k} .$$

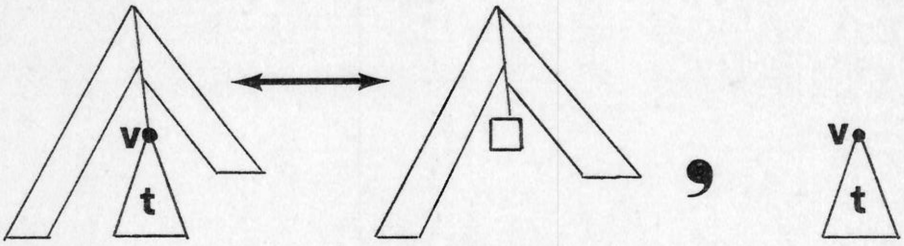
Die Summe ist

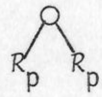
$$\sum_{k\geq 1;\lambda\geq 1} \lambda\, u^{2^k(1+2\lambda)} = \sum_{n\geq 1} \omega(n)\, u^n ,$$

sodaβ der Beweis sich durch Zusammenfassen ergibt.

Es wird noch ein zweiter Beweis angegeben, der auf einer Idee von *P. FLAJOLET* beruht. Er ist kürzer, aber man erzielt keine Formel für $\overline{V}_p$ für festes p.

Man denke sich einen binären Baum mit p kritischen Knoten. Man betrachte einen ausgezeichneten dieser kritischen Knoten v.



Der Auszeichnung von v entspricht eine eindeutige Zerlegung; $t \in \sum_{p\geq 0}$ 

Damit jeder kritische Knoten gezählt wird, muβ man jeden auszeichnen. Mit anderen Worten,

$$\overline{V}(z) = \sum_{n\geq 0} (n+1) \frac{1}{n+1} \binom{2n}{n} z^n \,.\, z \sum_{p\geq 0} R_p^2$$

$$= \frac{1}{\sqrt{1-4z}} \frac{u}{(1+u)^2} \sum_{p\geq 0} \left(\frac{1-u^2}{u}\right)^2 \frac{u^{2^{p+1}}}{\left(1-u^{2^{p+1}}\right)^2}$$

$$= \frac{1-u^2}{u} \sum_{p\geq 1;\lambda\geq 1} \lambda\, u^{\lambda 2^p} = \frac{1-u^2}{u} \sum_{n\geq 1} \psi(n)\, u^n ,$$

mit

$$\psi(n) = \sum_{n=\lambda 2^p;\, p,\lambda\geq 1} \lambda .$$

Sei nun $n = 2^m(1+2i)$, dann gilt

$$\psi(n) = \sum_{p=1}^{m} \frac{n}{2^p} = (1+2i) \sum_{j=0}^{m-1} 2^j = (1+2i)(2^m-1) = n - 1 - 2\,\omega(n),$$

d.h.

$$\overline{V} = \frac{1-u^2}{u} \left[\frac{u}{(1-u)^2} - \frac{u}{1-u} \right] - 2\, \frac{1-u^2}{u} \sum_{n\geq 1} \omega(n)\, u^n ,$$

wie behauptet. □

Man kann auch eine "explizite" Formel für $[z^n]\ \overline{V}(z)$ angeben:

SATZ 2. $[z^n]\ \overline{V}(z)$

$$= \binom{2n}{n-1} - 2 \sum_{k\geqslant 1} \omega(k) \left[\binom{2n}{n+1-k} - 2\binom{2n}{n-k} + \binom{2n}{n-1-k} \right].$$

Beweis. Die Berechnung erfolgt mit der Cauchy'schen Integralformel:

$$[z^n]\ \overline{V}(z) = \frac{1}{2\pi i} \int^{(0_+)} \frac{dz}{z^{n+1}} \overline{V}(z)$$

$$= \frac{1}{2\pi i} \int^{(0_+)} \frac{du\ (1-u)}{(1+u)^3} \frac{(1+u)^{2n+2}}{u^{n+1}} \left[\frac{u(1+u)}{1-u} - 2\,\frac{1-u^2}{u} \sum_{k\geqslant 1} \omega(k)\ u^k \right]$$

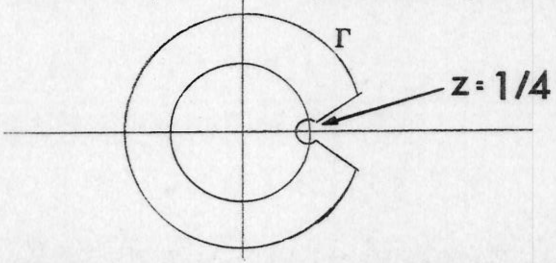
$$= [u^n]\ u(1+u)^{2n} - 2\,[u^n]\ \frac{(1-u)^2\ (1+u)^{2n}}{u} \sum_{k\geqslant 1} \omega(k)\ u^k;$$

hieraus ergibt sich die Behauptung unmittelbar.

3. ASYMPTOTIK .

Um das asymptotische Verhalten von K_n zu bestimmen, kann man wie in [5,8] vorgehen: Der erste Schritt ist, die Binomialkoeffizienten in Satz 2 zu approximieren; die auftretenden Summen $\sum_{k\geqslant 1} \omega(k)\ k^b \exp(-k^2/n)$ können dann mithilfe der Mellin-Transformation ausgewertet werden, oder aber durch partielle Summation und Information über $\sum_{k<n} \omega(k)$ (vgl. [14]). Das ist interessant, aber wir schlagen hier einen schnelleren Weg ein, vgl. [4,6,12].

Es wird der Koeffizient von z^n in $\overline{V}(z)$ mittels der Cauchyschen Integralformel ausgedrückt, und zwar unter Verwendung eins Integrationsweges wie unten angedeutet:



Hiezu braucht man eine analytische Fortsetzung von $\overline{V}(z)$ außerhalb des Konvergenzkreises. Das ist jedoch nicht schwierig, weil der Konvergenzradius von $\overline{V}(z)$ in der u-Ebene 1 ist, und der dem obigen Integrationsweg entsprechende Integrationsweg in der u-Ebene innerhalb des Einheitskreises liegt. Falls wir nun eine lokale Entwicklung um $z = 1/4$ haben, kann sie "übersetzt" werden in eine asymptotische Entwicklung von K_n. (Vgl. [4, 6,12].) Die lokale Entwicklung wird mit der Mellin-Transformation gewonnen. Man braucht eigentlich nur

$$e^{-\tau} = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \Gamma(s)\, \tau^{-s}\, ds, \qquad \operatorname{Re} \tau > 0,\ c > 0.$$

<u>LEMMA 4. Die erzeugende Dirichletreihe von $\omega(k)$ ist gegeben durch:</u>

$$\sum_{k\geq 1} \omega(k)\, k^{-s} = \frac{2^{s-1} - 1}{2^s - 1}\, \zeta(s-1) - \frac{1}{2}\zeta(s).$$

<u>Beweis.</u> Man beachte zunächst, daß

$$\sum_{\lambda\geq 0} (1+2\lambda)^{-s} = \zeta(s) - 2^{-s}\zeta(s) = (1-2^{-s})\,\zeta(s).$$

Daher gilt

$$\sum_{k\geq 1} \omega(k)\, k^{-s} = \sum_{k=2^i(1+2\lambda)} \lambda\, 2^{-is}(1+2\lambda)^{-s}$$

$$= \sum_{i\geq 0} 2^{-is} \sum_{\lambda\geq 0} \left[\frac{1}{2}(1+2\lambda) - \frac{1}{2}\right](1+2\lambda)^{-s}$$

$$= \frac{1}{1-2^{-s}}\left[\frac{1}{2}(1-2^{-s+1})\,\zeta(s-1) - \frac{1}{2}(1-2^{-s})\,\zeta(s)\right]. \qquad \square$$

<u>LEMMA 5. Die lokale Entwicklung von $\overline{V}(z)$ für $z \to 1/4$ ist gegeben durch</u>

$$\overline{V}(z) \sim \frac{1}{3\varepsilon} + \left[\frac{11}{18} + \frac{4\,\zeta'(-1)}{\log 2} + \frac{\gamma}{3\log 2}\right]\varepsilon + \frac{1}{3\log 2}\,\varepsilon\log\varepsilon$$

$$+ \frac{4}{\log 2} \sum_{k\neq 0} \Gamma(\chi_k)\,\zeta(\chi_k - 1)\,\varepsilon^{1-\chi_k}.$$

<u>mit</u> $\chi_k = \dfrac{2k\pi i}{\log 2}$.

<u>Beweis.</u> Wir setzen $u = e^{-t}$. Dann gilt $t = -\log u = -\log\frac{1-\varepsilon}{1+\varepsilon} = 2\varepsilon + \frac{2\varepsilon^3}{3} + O(\varepsilon^5)$.
Weiters gilt

$$\frac{u(1+u)}{1-u} = \frac{(1 - t + \frac{t^2}{2} + \ldots)(2 - t + \frac{t^2}{2} + \ldots)}{t - \frac{t^2}{2} + \frac{t^3}{6} + \ldots} = \frac{2}{t} - 2 + \frac{7}{6}t + \ldots$$

$$= \frac{1}{\varepsilon} - \frac{\varepsilon}{3} + \ldots - 2 + \frac{7}{3}\varepsilon + \ldots = \frac{1}{\varepsilon} - 2 + 2\varepsilon + \ldots .$$

In ähnlicher Weise erhält man

$$\frac{1-u^2}{u} = 2t + \frac{1}{3}t^3 + \ldots = 4\varepsilon + 4\varepsilon^3 + \ldots .$$

Nun können wir uns dem komplizierteren Ausdruck zuwenden:

$$\sum_{k\geq 1} \omega(k)\, e^{-kt} = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \Gamma(s) \sum_{k\geq 1} \omega(k)\, k^{-s}\, t^{-s}\, ds$$

$$= \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \Gamma(s)\, t^{-s} \left[\frac{2^{s-1}-1}{2^s-1} \zeta(s-1) - \frac{1}{2}\zeta(s) \right] ds.$$

Man kann den Integrationsweg beliebig weit nach links verschieben, wenn man nur die Residuen in Betracht zieht. Für $s = 2$, $s = 1$ liegen einfache Pole vor; sie stammen von den ζ-Funktionen. Weiters gibt es Pole bei $s = 2k\pi i/\log 2$, $k \in \mathbb{Z}$, und schließlich bei $s = -k$, $k \in \mathbb{N}_0$; sie stammen von der Γ-Funktion.

Das Residuum bei $s = 2$ ist

$$\frac{1}{3}\, t^{-2},$$

das Residuum bei $s = 1$ ist

$$-\frac{1}{2}\, t^{-1},$$

das Residuum bei $s = 0$ ist

$$-\frac{\log t}{24 \log 2} + \frac{3}{16} - \frac{\zeta'(-1)}{2 \log 2} - \frac{\gamma}{24 \log 2}$$

das Residuum bei $s = \chi_k$ für $k \neq 0$ ist

$$-\frac{1}{2 \log 2}\, \Gamma(\chi_k)\, \zeta(\chi_k - 1)\, t^{-\chi_k}.$$

Indem man zusammenfaßt, erhält man

$$\overline{V}(z) \sim \frac{1}{3\varepsilon} + \left[\frac{11}{18} + \frac{4\,\zeta'(-1)}{\log 2} + \frac{\gamma}{3 \log 2} \right] \varepsilon + \frac{1}{3 \log 2}\, \varepsilon \log \varepsilon$$

$$+ \frac{4}{\log 2} \sum_{k \neq 0} \Gamma(\chi_k)\, \zeta(\chi_k - 1)\, \varepsilon^{1-\chi_k}. \quad \square$$

<u>SATZ 3.</u>

$$K_n = \frac{n}{3} + \frac{1}{12} \log_2 n + K(\log_4 n) + o(1),$$

<u>wo</u> $K(x)$ <u>eine periodische Funktion mit Periode</u> 1 <u>ist; schreibt man</u> $K(x) = \sum_{k \in \mathbb{Z}} c_k e^{2k\pi i x}$, <u>so gilt</u>

$$c_0 = -\frac{3}{16} - \frac{\gamma}{12 \log 2} - \frac{2\,\zeta'(-1)}{\log 2} - \frac{1}{6 \log 2},$$

$$c_k = \frac{1}{\log 2}\, \zeta(\chi_k - 1)\, \Gamma(\chi_k/2)\, (\chi_k - 1), \quad k \neq 0, \quad \chi_k = \frac{2k\pi i}{\log 2}.$$

<u>Beweis.</u> Wir beziehen uns auf (11):

$$\frac{[z^n]\, \frac{1}{\varepsilon}}{[z^n]\, B} = n + 1.$$

Für $n \geqslant 1$ gilt:

$$\frac{[z^n]\ \varepsilon}{[z^n]\ B} = -\frac{n+1}{2n-1} = -\frac{1}{2} + \dots\ .$$

Weiters gilt [7]

$$[z^n]\ \log(1-z)\cdot(1-z)^{\alpha} = -\frac{n^{-\alpha-1}\log n}{\Gamma(-\alpha)} + \frac{n^{-\alpha-1}}{(\Gamma(-\alpha))^2}\,\Gamma'(-\alpha) + O\left(\frac{\log^{\star} n}{n^{\alpha+2}}\right),$$

und daher, unter Beachtung von $\Gamma(-\frac{1}{2}) = -2\sqrt{\pi}$, $\Gamma'(-\frac{1}{2}) = 2\sqrt{\pi}\,(\gamma + 2\log 2 - 2)$ (vgl. [15]):

$$[z^n]\ \varepsilon\log\varepsilon = \frac{1}{2}\,4^n\left[\frac{n^{-3/2}\log n}{2\sqrt{\pi}} + \frac{n^{-3/2}\,2\sqrt{\pi}\,(\gamma + 2\log 2 - 2)}{4\pi} + \dots\right]$$

$$= \frac{1}{4}\,\frac{4^n\,n^{-3/2}}{\sqrt{\pi}}\left[\log n + (\gamma + 2\log 2 - 2) + \dots\right].$$

Weiters ist [1]

$$[z^n]\ \varepsilon^{1-\chi_k} = 4^n\,\frac{n^{-\frac{1-\chi_k}{2}-1}}{\Gamma(\frac{\chi_k-1}{2})} + \dots$$

und

$$[z^n]\ B = \frac{1}{n+1}\binom{2n}{n} = \frac{4^n}{\sqrt{\pi}\,n^{3/2}} + \dots,$$

sodaß wir erhalten

$$K_n = \frac{1}{3}(n+1) + \left[\frac{11}{8} + \frac{4\,\zeta'(-1)}{\log 2} + \frac{\gamma}{3\log 2}\right].(-\tfrac{1}{2}) + \dots$$

$$+ \frac{1}{3\log 2}\,\frac{1}{4}\left[\log n + (\gamma + 2\log 2 - 2)\right] + \dots$$

$$+ \frac{4}{\log 2}\,\Gamma(\chi_k)\,\zeta(\chi_k-1)\sum_{k\neq 0}\frac{n^{\chi_k/2}}{\Gamma(\frac{\chi_k-1}{2})}\sqrt{\pi}\ .$$

Die endgültige Formel ergibt sich unter Beachtung von [15]:

$$\frac{\Gamma(\chi_k)\sqrt{\pi}}{\Gamma(\frac{\chi_k-1}{2})} = 2^{\chi_k-1}\,\Gamma(\chi_k/2)\,\Gamma(\tfrac{\chi_k+1}{2})\,/\,\Gamma(\tfrac{\chi_k-1}{2}) = \frac{1}{4}\,\Gamma(\chi_k/2)\,(\chi_k-1)$$

und $n^{\chi_k/2} = e^{2k\pi i\,\log_4 n}$. $\square$

4. DIE HÖHEREN MOMENTE

SATZ 4. Das s-te Moment $K_n^{(s)}$ der Anzahl der kritischen Knoten, wobei alle binären Bäume mit n Knoten als gleich wahrscheinlich angesehen werden, erfüllt folgende Beziehung ($s \geq 1$):

$$K_n^{(s)} \sim \frac{1}{3} n^s .$$

Beweis. Es wird hier darauf verzichtet, eine genauere Formel herzuleiten, obwohl dies nach dem Muster des vorigen Kapitels möglich wäre. Wir verwenden die Idee von P. Flajolet und geben nur eine Beweisskizze:

$$K_n^{(s)} = \frac{[z^n] \left(\sum\limits_{m \geq 0} (m+1)^{s-1} \binom{2m}{m} z^m \right) z \sum\limits_{p \geq 0} (R_p(z))^2}{\frac{1}{n+1} \binom{2n}{n}} .$$

Der Zähler ist asymptotisch äquivalent mit

$$[z^n] \sum_{m \geq 0} m^{s-1} \binom{2m}{m} z^m \; \frac{1}{4} \sum_{p \geq 0} (R_p(z))^2 ,$$

$z = 1/4$ bedeutet $u = 1$; daher kann der zweite Faktor wie folgt ausgewertet werden:

$$\frac{1}{4} \sum_{p \geq 0} \lim_{u \to 1} \left(\frac{1-u^2}{u} \, \frac{u^{2^p}}{1-u^{2^{p+1}}} \right)^2 = \frac{1}{4} \sum_{p \geq 0} \left(\frac{1}{2^p} \right)^2 = \frac{1}{3} .$$

Also gilt

$$K_n^{(s)} \sim \frac{n^{s-1} \binom{2n}{n}}{n^{-1} \binom{2n}{n}} \, \frac{1}{3} = \frac{1}{3} n^s .$$

5. ÜBER DIE LINKSSEITIGE HÖHE BINÄRER BÄUME

Wir wenden uns nun einer anderen, jedoch sehr ähnlichen Fragestellung zu (vgl. [8]). Die Vorgangsweise ist ähnlich wie vorher; daher können wir uns kürzer fassen.

Die linksseitige Höhe h ist für binäre Bäume wie folgt definiert (vgl [2,9]):

$$h(\square) = 0$$

$$h\left(\begin{smallmatrix} & \circ & \\ / & & \backslash \\ t_1 & & t_2 \end{smallmatrix} \right) = \max \{ 1 + h(t_1) , h(t_2) \} .$$

Folgende Analoga zu den kritischen Knoten können betrachtet werden:

Ein Knoten heißt *links-abhängig*, falls die (linksseitige) Höhe seines linken Unterbaumes um 1 kleiner ist als die des Knotens.

Ein Knoten heißt *rechts-abhängig*, falls sein rechter Unterbaum dieselbe Höhe hat. Treffen beide Bedingungen zu, heißt der Knoten *links-rechts-abhängig*.

Es werden nun die entsprechenden Mittelwerte l_n, r_n, m_n der Anzahl der abhängigen Knoten betrachtet.

SATZ 5.

$$l_n = \frac{3}{4} n + \frac{5}{8} + O(\frac{1}{n}) ,$$

$$r_n = (2 - \frac{\pi^2}{6}) n + (\frac{1}{12} - \frac{5}{36} \pi^2) + O(\frac{1}{n}) ,$$

$$m_n = (\frac{7}{4} - \frac{\pi^2}{6}) n + (\frac{17}{24} - \frac{5}{36} \pi^2) + O(\frac{1}{n}) .$$

Beweis. Sei $C_h(z)$ die erzeugende Funktion der Anzahl der Bäume mit Höhe $=h$. Es gilt [2,9]:

$$C_h(z) = (1+u) \left[\frac{1 - u^{h+1}}{1 - u^{h+2}} - \frac{1 - u^h}{1 - u^{h+1}} \right] .$$

Sei nun $L_h(z,y)$ bzw. $R_h(z,y)$ die erzeugende Funktion der Bäume mit Höhe h, wobei der Koeffizient von $z^n y^m$ sich auf die Bäume mit n Knoten und m links- bzw. rechts-abhängigen Knoten bezieht. Weiters sei

$$\overline{L}(z) = \sum_{h\geq 0} \overline{L}_h(z) \quad \text{bzw.} \quad \overline{R}(z) = \sum_{h\geq 0} \overline{R}_h(z) .$$

Indem man die Idee von P. Flajolet benützt, erhält man

$$\overline{L}(z) = \frac{1}{\sqrt{1-4z}} z \sum_{h\geq 0} C_h \sum_{i\leq h+1} C_i ,$$

$$\overline{R}(z) = \frac{1}{\sqrt{1-4z}} z \sum_{h\geq 1} C_h \sum_{i\leq h-1} C_i .$$

Es sei d(k) die Anzahl der Teiler von k und $\sigma(k)$ die Summe der Teiler von k. Weiters sei für den Augenblick

$$A = \sum_{k\geq 1} d(k) u^k \quad \text{und} \quad B = \sum_{k\geq 1} \sigma(k) u^k .$$

$$\overline{L}(z) = \frac{u}{1-u^2} \sum_{h\geq 0} C_h C_0 + \frac{u}{1-u^2} \sum_{h\geq 0} C_h \sum_{1\leq i\leq h+1} C_i$$

$$= \frac{u}{1-u} + \frac{u}{1-u^2} \sum_{i\geq 0} C_{i+1} \sum_{h\geq i} C_h$$

$$= \frac{u}{1-u} + u(1+u) \sum_{i\geq 0} \left[\frac{1-u^{i+2}}{1-u^{i+3}} - \frac{1-u^{i+1}}{1-u^{i+2}} \right] \frac{u^i}{1-u^{i+1}}$$

$$= \frac{u}{1-u} + u(1+u) \sum_{i\geq 0} \left[\frac{u^{i+3}-u^{i+2}}{1-u^{i+3}} - \frac{u^{i+2}-u^{i+1}}{1-u^{i+2}} \right] \frac{u^i}{1-u^{i+1}}$$

$$= \frac{u}{1-u} + (1-u^2) \sum_{i\geq 1} \left[\frac{u^i}{1-u^{i+1}} - \frac{u^{i+1}}{1-u^{i+2}} \right] \frac{u^i}{1-u^i}$$

$$= \frac{u}{1-u} + (1+u) \sum_{i\geq 0} u^i \left[\frac{1}{1-u^i} - \frac{1}{1-u^{i+1}} \right] - \sum_{i\geq 1} u^{i+1} \left[\frac{1}{1-u^i} - \frac{1}{1-u^{i+2}} \right]$$

$$= \frac{u}{1-u} + (1+u)\, A - \frac{1+u}{u} \left[A - \frac{u}{1-u} \right] - uA + \frac{1}{u} \left[A - \frac{u^2}{1-u^2} - \frac{u}{1-u} \right]$$

$$= \frac{u(1+2u)}{u^2} .$$

Wie in Satz 2 finden wir daher

$$[z^n]\, \overline{L}(z) = [u^n]\, u(1+2u)(1+u)^{2n-2} = \binom{2n-2}{n-1} + 2 \binom{2n-2}{n-2} ,$$

und daher

$$l_n = \frac{3n-2}{4n-2} (n+1) .$$

In ähnlicher Weise geht man nun bei $\overline{R}(z)$ vor:

Man beachtet zunächst, daß

$$\sum_{k\geq 1} \sigma(k)\, u^k = \sum_{i,j\geq 1} j\, u^{ij} = \sum_{i\geq 1} \frac{u^i}{(1-u^i)^2} .$$

Daher gilt

$$\sum_{i\geq 1} \frac{u^{2i}}{(1-u^i)^2} = \sum_{i\geq 1} \frac{u^i\,(1-(1-u^i))}{(1-u^i)^2} = B - A.$$

$$\overline{R}(z) = \frac{u}{1-u^2} \sum_{i\geq 0} C_i \sum_{h\geq i+1} C_h$$

$$= u(1+u) \sum_{i\geq 0} \left[\frac{1-u^{i+1}}{1-u^{i+2}} - \frac{1-u^i}{1-u^{i+1}} \right] \frac{u^{i+1}}{1-u^{i+2}}$$

$$= u(1-u^2) \sum_{i\geq 0} \left[\frac{u^i}{1-u^{i+1}} - \frac{u^{i+1}}{1-u^{i+2}} \right] \frac{u^{i+1}}{1-u^{i+2}}$$

$$= u(1+u) \sum_{i\geq 0} u^i \left[\frac{1}{1-u^{i+1}} - \frac{1}{1-u^{i+2}} \right] - \frac{1-u^2}{u} \sum_{i\geq 2} \frac{u^{2i}}{(1-u^i)^2}$$

$$= (1+u)\, A - \frac{1+u}{u} \left[A - \frac{u}{1-u} \right] - \frac{1-u^2}{u} \left[B - A - \frac{u^2}{(1-u)^2} \right]$$

$$= \frac{1+u}{u}\, A \left[u - 1 + 1 - u \right] + \frac{1+u}{1-u} + \frac{u(1+u)}{1-u} - \frac{1-u^2}{u}\, B$$

$$= - \frac{1-u^2}{u}\, B + \frac{(1+u)^2}{1-u} . \qquad \square$$

Wieder kann man eine "explizite" Formel für die Koeffizienten finden:

$$[z^n]\ \overline{R}(z) = \binom{2n+1}{n} - \sum_{k\geq 1} \sigma(k) \left[\binom{2n}{n+1-k} - 2\binom{2n}{n-k} + \binom{2n}{n-1-k} \right].$$

Das asymptotische Verhalten wird wieder mittels der lokalen Entwicklung von $\overline{R}(z)$ für $z \to 1/4$ gefunden. Man benützt

$$\sum_{k\geq 1} \sigma(k)\, k^{-s} = \zeta(s-1)\,\zeta(s)$$

und

$$\sum_{k\geq 1} \sigma(k)\, e^{-tk} = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \Gamma(s)\, t^{-s}\, \zeta(s-1)\,\zeta(s)\, ds$$

$$\sim \frac{\pi^2}{24\varepsilon^2} - \frac{1}{4\varepsilon} + \left(\frac{1}{24} - \frac{\pi^2}{36}\right) + \ldots .$$

Insgesamt ergibt sich

$$\overline{R}(z) \sim \frac{1}{\varepsilon}\left(2 - \frac{\pi^2}{6}\right) - 1 + \varepsilon\left(\frac{11}{6} - \frac{\pi^2}{18}\right) + \ldots$$

Hieraus ergibt sich r_n nach dem Muster von Kapitel 3.

Um m_n zu finden, beachtet man, daß

$$l_n + r_n - m_n = n.$$

DANKSAGUNG. Die Endfassung dieser Arbeit wurde erstellt, während der Autor das *Laboratoire de Recherche en Informatique* der *Universität Paris XI (Orsay)* besuchte. Für die gewährte Gastfreundschaft sei an dieser Stelle herzlichst gedankt.

Weiters sei angemerkt, daß Diskussionen mit *Philippe Flajolet* für die Abfassung dieser Arbeit sehr hilfreich waren.

LITERATUR

[1] E.A.BENDER, Asymptotic methods in enumeration, SIAM Reviews 16 (1974), 485-515.

[2] N.G. de BRUIJN, D.E.KNUTH, S.O.RICE, The average height of planted plane trees, in: Graph Theory and Computing (R.C.Read, Ed.), 15-22, Academic Press, New York-London, 1972.

[3] P.FLAJOLET, Analyse d'algorithmes de manipulation d'arbres et de fichiers, Cahiers du BURO, 34-35 (1981), 1-209.

[4] P.FLAJOLET, A.ODLYZKO, The average height of binary trees and other simple trees, J. Comput. Syst. Sci. 25 (1982), 142-158.

[5] P.FLAJOLET, J.-C.RAOULT, J.VUILLEMIN, The number of registers required for evaluating arithmetical expressions, Theoretical Computer Science 9 (1979), 99-125.

[6] P.FLAJOLET, H.PRODINGER, Register allocation for unary-binary trees, SIAM J. on Computing, im Druck.

[7] R.JUNGEN, Sur les séries de Taylor n'ayant que des singularités algébro-logarithmiques sur leur cercle de convergence, Commentarii Math. Helvetici 3 (1931), 266-306.

[8] R.KEMP, The average number of registers to evaluate a binary tree optimally, Acta Informatica 11 (1979), 363-372.

[9] D.E.KNUTH, The art of computer programming, Vol. 1, Addison Wesley 1968.

[10] A.MEIR, J.W.MOON, J.R.POUNDER, On the order of random channel networks, SIAM J. Alg. Discr. Meth. 1 (1980), 25-33.

[11] J.W.MOON, On Horton's Law for random channel networks, Annals of Discrete Mathematics 8 (1980), 117-121.

[12] A.ODLYZKO, Periodic oscillations of coefficients of power series that satisfy functional equations, Advances in Mathematics 44 (1982), 180-205.

[13] H.PRODINGER, The influence of the nodes on the leftsided height of a binary tree, submitted.

[14] H.PRODINGER, R.F.TICHY, Über ein zahlentheorétisches Problem aus der Informatik, Sitzungsberichte der Österreichischen Akademie der Wissenschaften, im Druck.

[15] E.T.WHITTAKER, G.N.WATSON, A course of modern analysis, Cambridge University Press, 1927.

Helmut Prodinger,
Institut für Algebra und Diskrete Mathematik,
Abteilung für Theoretische Informatik,
Technische Universität Wien,
Gußhausstraße 27-29,
A-1040 WIEN
ÖSTERREICH