ÜBER LÄNGSTE 1-TEILFOLGEN IN 0-1-FOLGEN *)

Helmut Prodinger

1. EINLEITUNG

Das Interesse an der vorliegenden Arbeit ist vorwiegend *methodologischer* Natur; an Hand zweier Beispiele (Kapitel 2 und 3) soll eine Vorgangsweise der asymptotischen Abzähltechniken, die von de Bruijn vorgeschlagen [4] und von Flajolet systematisiert worden ist [1], beschrieben werden. Diese Methode, die auf der *Mellin-Transformation* beruht, ist bei vielen Problemen der Theoretischen Informatik, der Kombinatorik, sowie der Analytischen Zahlentheorie anwendbar; es wird die Hoffnung ausgesprochen, daß die vorliegende Arbeit dazu beiträgt, diese Techniken weiter zu popularisieren. Um dieses Ziel zu erreichen, werden diverse lästige Details fortgelassen, um die generelle Vorgangsweise überschaubarer zu machen. Diese sind in der Diplomarbeit von Ulrich Schmidt (Abzählmethoden der Theoretischen Informatik) ausgearbeitet worden.

Es werden Folgen aus 0 und 1 betrachtet (Kapitel 2; es gibt 2^n Folgen der Länge n), sowie Folgen, wo die Anzahlen von 0 und 1 gleich sind (Kapitel 3; es gibt $\binom{2n}{n}$ derartige Folgen der Länge 2n).

Studiert werden in beiden Fällen die d-längste maximale 1-Folge, d.h. man denkt sich die Längen der maximalen 1-Folgen in absteigender Reihenfolge aufgeschrieben (mit Wiederholungen) und betrachtet das d-te Element.

Beispiel: 0011101011100011011110 hat maximale 1-Folgen der Längen 3,1,3,2,4; also in geordneter Reihenfolge 4,3,3,2,1, sodaß etwa die drittlängste 1-Folge die Länge 3 hat.

Es werden hier in ausgiebiger Weise (gewöhnliche) erzeugende Funktionen verwendet. Sei etwa A eine Familie von Objekten, denen man eine "Größe" $|\ |$ zuordnen kann; dann sei die zugehörige erzeugende Funktion

$$A(z) = \sum_{\sigma \in A} z^{|\sigma|} .$$

In unserem Fall wird man als Größe wohl die Länge wählen.

$[z^n]\, f(z)$ bezeichne den Koeffizienten von z^n in $f(z)$.

---

*) Über diese Resultate ist beim Österreichischen Mathematiker-Kongreß in Graz vorgetragen worden.

2. 0-1-FOLGEN OHNE EINSCHRÄNKUNGEN

Sei $1^{<k}$ die Menge aller Folgen $\varepsilon, 1, 11, \ldots, 1^{k-1}$, wo ε die leere Folge bedeutet. Weiters sei A_k die Familie der Folgen, deren längste 1-Folge eine Länge $<k$ hat. Man erhält sofort

$$A_k = 1^{<k}(01^{<k})^*, \tag{1}$$

wo * der Kleene'sche *-Operator ist ($M^* = \bigcup_{i \geq 0} M^i$).

Die entsprechende erzeugende Funktion $A_k(z)$ ist dann

$$A_k(z) = \frac{1-z^k}{1-z}\left(1 - z\frac{1-z^k}{1-z}\right)^{-1} = \frac{1-z^k}{1-2z+z^{k+1}} . \tag{2}$$

Die entsprechende wahrscheinlichkeitserzeugende Funktion ist dann $A_k(\frac{z}{2})$.

Der Erwartungswert der längsten 1-Folge, wo alle 0-1-Folgen als gleich wahrscheinlich angesehen werden, ist dann (partielle Summation!)

$$\sum_{k \geq 1} \quad 1 - [z^n] \frac{1-(\frac{z}{2})^k}{1 - z + (\frac{z}{2})^{k+1}} . \tag{3}$$

Die asymptotische Entwicklung von (3) ist bereits Knuth [5] gelungen; wir werden im weiteren seine Herleitung skizzieren.

Nun sei $A_k^{<d>}$ die Familie der Folgen, wo genau d 1-Folgen Länge $\geq k$ haben. Natürlich gilt $A_k^{<0>} = A_k$. Sei $d \geq 1$; man findet dann

$$A_k^{<d>} = (A_k 0+\varepsilon)1^{\geq k}(0A_k 0+0)1^{\geq k} \ldots 1^{\geq k}(0A_k+\varepsilon), \tag{4}$$

wo d-mal $1^{\geq k}$ vorkommt ($1^{\geq k} = \{1^k, 1^{k+1}, \ldots\}$).

Die entsprechende erzeugende Funktion $A_k^{<d>}(z)$ ist daher

$$A_k^{<d>}(z) = \left(\frac{1-z}{1-2z+z^{k+1}}\right)^2 \left(\frac{z^k}{1-z}\right)^d \left(z\,\frac{1-z}{1-2z+z^{k+1}}\right)^{d-1}$$

$$= \frac{z^{kd+d-1}(1-z)}{(1-2z+z^{k+1})^{d+1}} \qquad \text{für } d \geq 1. \tag{5}$$

Die mittlere Länge der (d+1)-längsten Folge ($d \geq 0$) ist dann

$$\sum_{k \geq 1} \text{Prob}\,\{\text{die } (d+1)\text{-längste Folge ist } \geq k\}$$

$$= \sum_{k \geq 1} 1 - \text{Prob}\,\{\text{die } (d+1)\text{-längste Folge ist } < k\}$$

$$= \sum_{k \geq 1} 1 - \text{Prob}\,\{\text{alle 1-Folgen haben Länge } < k\}$$

$$- \text{Prob}\,\{\text{genau eine 1-Folge hat Länge } \geq k\} \tag{6}$$

$$\vdots$$

$$- \text{Prob}\,\{\text{genau } d \text{ 1-Folgen haben Länge} \geq k\}$$

$$= \sum_{k \geq 1} 1 - [z^n]\, A_k(\tfrac{z}{2}) - [z^n]\, A_k^{<1>}(\tfrac{z}{2}) - \ldots - [z^n]\, A_k^{<d>}(\tfrac{z}{2}).$$

Knuth's Beobachtung war nun, daß $1-z+(\frac{z}{2})^{k+1}$ (aus dem Nenner von $A_k(\frac{z}{2})$) eine *dominante* einfache Nullstelle $\rho_k = 1 + \varepsilon_k$ hat, mit

$$\rho_k = 1 + \varepsilon_k = 1 + 2^{-k-1} + O(k2^{-2k}). \tag{7}$$

Durch etwas lästige, aber nicht schwierige Abschätzungen ergibt sich, daß man innerhalb der angestrebten Rechengenauigkeit $A_k(\frac{z}{2})$ durch

$$\frac{1 - (\frac{\rho_k}{2})^k}{-1+(k+1)(\frac{\rho_k}{2})^k} \cdot \frac{1}{z - \rho_k},$$

bzw. sogar durch

$$\frac{1}{1 - z/\rho_k} \tag{8}$$

ersetzen darf; den Koeffizienten von z^n in $A_k(\frac{z}{2})$ kann man dann durch

$$[z^n]\, \frac{1}{1-z/\rho_k} = \rho_k^{-n} \tag{9}$$

ersetzen, oder gemäß (7) durch

$$(1 + 2^{-k-1})^{-n}$$

bzw. sogar durch

$$e^{-n/2^{k+1}}. \tag{10}$$

Knuth's Aufgabe war daher,

$$\sum_{k \geq 2} 1 - e^{-n/2^k} \tag{11}$$

asymptotisch auszuwerten; summiert man bereits ab $k \geq 0$, so muß man vom Endergebnis 2 + exponentiell kleine Bestandteile abziehen.

Nun wollen wir eine entsprechende Betrachtung anstellen, um

$$[z^n] \sum_{k \geq 1} A_k^{<t>}(\tfrac{z}{2}) \tag{12}$$

für festes $t \geq 1$ asymptotisch zu bestimmen (vgl. (6)). Die Nullstellen im Nenner von

$$A_k^{<t>}(\tfrac{z}{2}) = \frac{(1 - \frac{z}{2})\, (\frac{z}{2})^{kt+t-1}}{(1 - z + (\frac{z}{2})^{k+1})^{t+1}}$$

sind ja wie bei Knuth; nur die Vielfachheit ist eine andere. Wir ersetzen daher

$A_k^{<t>}(\frac{z}{2})$ durch

$$\frac{(1-\frac{\rho_k}{2})(\frac{\rho_k}{2})^{kt+t-1}}{(-1+(\frac{\rho_k}{2})^{k+1})^{t+1}} \cdot \frac{1}{(z-\rho_k)^{t+1}}$$

bzw. durch

$$\frac{1}{2^{kt+t}} \frac{1}{(1-z/\rho_k)^{t+1}} . \tag{13}$$

$[z^n]\ A_k^{<t>}(\frac{z}{2})$ ersetzen wir durch den Koeffizienten von z^n in (13), also in weiterer Folge durch

$$\frac{1}{2^{kt+t}} \frac{n^t}{t!} \rho_k{}^{-n}$$

und durch

$$\frac{n^t}{2^{(k+1)t}\, t!} e^{-n/2^{k+1}} . \tag{14}$$

Unsere Aufgabe ist es also,

$$\sum_k \frac{1}{2^{kt}} e^{-n/2^k} \tag{15}$$

asymptotisch auszuwerten; man kann bei k=0 zu summieren beginnen.

Die Methode, um (11) und auch (15) zu behandeln, ist die *Mellin-Transformation*, für die wir auf die Übersichtsarbeit von Flajolet, Régnier und Sedgewick [1] verweisen. Sei

$$f(x) = \sum_{k\geq 0} 1 - e^{-x/2^k} ; \tag{16}$$

die Mellin-Transformierte $f^*(s)$ ist dann

$$f^*(s) = \int_0^\infty x^{s-1} f(x)\, dx$$

$$= - \sum_{k\geq 0} 2^{ks}\ \Gamma(s) = - \frac{1}{1-2^s} \Gamma(s) \tag{17}$$

Dies gilt für $-1 < \mathrm{Re}(s) < 0$ ("fundamentaler Streifen"). Die Umkehrformel für die Mellin-Transformation liefert dann

$$f(x) = \frac{1}{2\pi i} \int_{-\frac{1}{2}-i\infty}^{-\frac{1}{2}+i\infty} f^*(s)\, x^{-s}\, ds. \tag{18}$$

Man verschiebt die Integrationsgerade nach *rechts*, wobei man die *Residuen* in Betracht zieht:

$$f(x) \sim - \sum \mathrm{Res}(f^*(s) x^{-s}); \tag{19}$$

die Summe wird dabei erstreckt über alle Pole mit Realteil ≥ 0. Man hat einfache Pole an den Stellen

$$\chi_k = \frac{2k\pi i}{\log 2}, \quad k \in \mathbb{Z},\ k \neq 0, \tag{20}$$

sowie einen doppelten Pol bei $s=0$. Die Bestimmung der Residuen ist sehr einfach, sodaß man erhält:

$$\begin{aligned} f(x) &\sim \log_2 x + \frac{\gamma}{\log 2} + \frac{1}{2} \\ &\quad - \frac{1}{\log 2} \sum_{k\neq 0} \Gamma\left(\frac{2k\pi i}{\log 2}\right) x^{-2k\pi i/\log 2}. \end{aligned} \tag{21}$$

Üblicherweise führt man ein:

$$P(u) = \frac{1}{\log 2} \sum_{k\neq 0} \Gamma\left(\frac{2k\pi i}{\log 2}\right) e^{-2k\pi i.u}$$

und hat dann

$$f(x) \sim \log_2 x + \frac{\gamma}{\log 2} + \frac{1}{2} - P(\log_2 x). \tag{22}$$

Nun folgt die entsprechende Diskussion für

$$f(x) = \sum_{k\geq 0} \frac{1}{2^{kt}} e^{-x/2^k}. \tag{23}$$

Die Mellin-Transformierte $f^*(s)$ ist dann

$$f^*(s) = \sum_{k\geq 0} \frac{1}{2^{kt}} 2^{ks} \Gamma(s) = \frac{1}{1-2^{s-t}} \Gamma(s). \tag{24}$$

Der fundamentale Streifen ist $0 < \mathrm{Re}(s) < t$. Man hat also wieder

$$f(x) = \frac{1}{2\pi i} \int_{\frac{1}{2} - i\infty}^{\frac{1}{2} + i\infty} f^*(s)\, x^{-s}\, ds, \tag{25}$$

und durch Verschiebung des Integrationsweges nach rechts:

$$\begin{aligned} f(x) &\sim - \sum_{k\in\mathbb{Z}} \mathrm{Res}(f^*(s_k) x^{-s_k};\ s_k = t + \frac{2k\pi i}{\log 2}) \\ &\sim \frac{1}{\log 2} \sum_{k\in\mathbb{Z}} \Gamma\left(t + \frac{2k\pi i}{\log 2}\right) x^{-t-2k\pi i/\log 2}. \end{aligned} \tag{26}$$

Unter Verwendung von (6), (14), (22) und (26) können wir daher folgenden Satz aussprechen, der einen Satz von Knuth [5] als Spezialfall enthält:

SATZ 1: *Seien alle 0-1-Folgen der Länge n als gleich wahrscheinlich angenommen. Für den Erwartungswert $E_n^{<d>}$ der $(d+1)$-längsten 1-Folge erhält man als asymptotisches Äquivalent*

$$E_n^{<d>} \sim \log_2 n + \frac{\gamma}{\log 2} - \frac{3}{2} - \frac{H_d}{\log 2} -$$

$$- \left[P(\log_2 n) + P^{<1>}(\log_2 n) + \ldots + P^{<d>}(\log_2 n) \right] ;$$

die periodischen Funktionen $P^{<t>}(u)$ mit Periode 1 (und kleiner Amplitude) haben die folgende Fourierentwicklung:

$$P^{<t>}(u) = \frac{1}{\log 2} \frac{1}{t!} \sum_{k \neq 0} \Gamma\left(t + \frac{2k\pi i}{\log 2}\right) e^{-2k\pi i . u},$$

wobei $P(u)$ als $P^{<0>}(u)$ aufzufassen ist; H_d ist die d-te harmonische Zahl

$$H_d = 1 + \frac{1}{2} + \ldots + \frac{1}{d},$$

die von den jeweiligen Summanden für $k=0$ herrührt. Für $d=0$ muß man (wie üblich) die leeren Summen als 0 auffassen. □

3. 0-1-FOLGEN MIT EINSCHRÄNKUNGEN

Es sollen hier alle 0-1-Folgen der Länge 2n mit jeweils n Nullen bzw. Einsen betrachtet werden. Die Notation ist ähnlich wie im Kapitel 2, allerdings verwenden wir B statt A.

Wir betrachten wieder die Zerlegung (1):

$$1^{<k}(01^{<k})^{*}. \tag{27}$$

Wir markieren 1 durch z, sowie 0 durch w, sodaß man aus (27) zu (28) gelangt:

$$F(z,w) = \frac{1 - z^k}{1 - z - w + wz^k} . \tag{28}$$

Die erzeugende Funktion B_k der Folgen mit maximalen 1-Folgen der Länge k ist dann die "Diagonale":

$$B_k(z) = \sum_{n \geq 0} z^n . [z^n w^n] F(z,w). \tag{29}$$

Man bekommt diese Diagonale durch Kontur-Integration. Wir verweisen hiezu beispielsweise auf Greene, Knuth [7], Hautus, Klarner [3], sowie Post [6]:

$$B_k(z) = \frac{1}{2\pi i} \oint F\left(t, \frac{z}{t}\right) \frac{dt}{t}$$

$$= \frac{1}{2\pi i} \oint \frac{1 - t^k}{t - t^2 - z + zt^k} dt. \tag{30}$$

Wir betrachten den Nenner des Integranden. Die Gleichung

$$\Phi(t,z) = t - t^2 - z + zt^k = 0 \tag{31}$$

hat Lösungen

$$t_1(z),\ t_2(z),\ \ldots\ ,\ t_k(z). \tag{32}$$

Das simultane System

$$\begin{aligned} \Phi(z,t) &= 0 \\ \frac{\partial\Phi}{\partial t}(z,t) &= 0 \end{aligned} \tag{33}$$

gibt uns die singuläre Stelle (ρ, τ):

$$\begin{aligned} \tau - \tau^2 - \rho + \rho\tau^k &= 0 \\ 1 - 2\tau + k\rho\tau^{k-1} &= 0 \end{aligned} \tag{34}$$

oder

$$1 - 2\tau + k\,\frac{(1-\tau)\tau^k}{1-\tau^k} = 0 \tag{35}$$

"Bootstrapping" ergibt

$$\tau \sim \frac{1}{2} \quad \text{bzw.} \quad \tau \sim \frac{1}{2} + \frac{k}{2^{k+2}}. \tag{36}$$

Weiters gewinnt man

$$\rho = \frac{\tau(1-\tau)}{1-\tau^k} \sim \frac{1}{4}\left(1 + \frac{1}{2^k}\right). \tag{37}$$

Die Lösungen t_1, t_2 haben (geeignete Numerierung vorausgesetzt!) bei $z = \rho$ eine algebraische Singularität; die übrigen sind bei $z = \rho$ regulär. Eine ähnliche Situation tritt bei einem anderen Problem auf (Flajolet, Prodinger [2]). Man hat für $z \to \rho$:

$$t_{1,2}(z) \sim A \mp B\sqrt{1 - \frac{z}{\rho}} + \ldots \tag{38}$$

Entwickelt man

$$z(t) = t\,\frac{1-t}{1-t^k}$$

in eine Taylorreihe um $t = \tau$, erhält man nach einiger Rechnung

$$z \sim \rho - (t - \tau)^2,$$

oder

$$t \sim \tau \mp \sqrt{\rho}\,\sqrt{1 - \frac{z}{\rho}}.$$

Demnach ist

$$A = \tau \sim \frac{1}{2} \quad \text{und} \quad B = \sqrt{\rho} \sim \frac{1}{2}. \tag{39}$$

Das Integral (30) ist nach dem Residuenkalkül

$$\operatorname*{Res}_{t=t_1(z)} \frac{1 - t^k}{\Phi(t,z)}$$

$$= \frac{1 - (t_1(z))^k}{\lim\limits_{t \to t_1(z)} \dfrac{t - t^2 - z + zt^k}{t - t_1(z)}} \tag{40}$$

Von der Funktion (40) wollen wir nur den ersten Term der lokalen Entwicklung um $z = \rho$. Der Limes in (40) ist

$$\lim_{t \to t_1(z)} z(t-t_2(z)) \ldots (t-t_k(z)). \tag{41}$$

Nun ist

$$t_1(z) - t_2(z) \sim - \sqrt{1-z/\rho} \quad ;$$

der restliche Faktor ist

$$\lim_{t \to t_1(z)} z(t-t_3(z)) \ldots (t-t_k(z))$$

$$= \lim_{t \to t_1(z)} \frac{t - t^2 - z + zt^k}{(t-t_1(z))(t-t_2(z))}$$

$$= \lim_{t \to \tau} \frac{t - t^2 - \rho + \rho t^k}{(t - \tau)^2}$$

hier wurden (38) und (39) verwendet. Nun wird de l'Hospitals Regel zweimal benützt:

$$= \lim_{t \to \tau} \frac{1 - 2t + k\rho t^{k-1}}{2(t - \tau)}$$

$$= \lim_{t \to \tau} \frac{- 2 + k(k-1)\rho t^{k-2}}{2}$$

$$= - 1 + \binom{k}{2} \rho \tau^{k-2} \sim -1 . \tag{42}$$

Demnach können wir (40) durch

$$\frac{1 - \tau^k}{\sqrt{1-z/\rho}}$$

bzw. durch

$$\frac{1}{\sqrt{1-z/\rho}} \tag{43}$$

ersetzen. Der Koeffizient von z^n in (43) ist

$$\sim \frac{1}{\sqrt{\pi n}} \quad \rho^{-n}$$

$$\sim \frac{1}{\sqrt{\pi n}} 4^n \left(1 + \frac{1}{2^k}\right)^{-n}$$

$$\sim \frac{1}{\sqrt{\pi n}} 4^n e^{-n/2^k} \tag{44}$$

Demnach hat man für den Erwartungswert der längsten 1-Folge:

$$\sum_{k \geq 1} 1 - \binom{2n}{n}^{-1} [z^n] B_k(z)$$

$$\sim \sum_{k \geq 1} 1 - e^{-n/2^k};$$

diese Größe ist bereits behandelt worden.

Wir haben somit

SATZ 2: *Seien alle 0-1-Folgen mit n Nullen und n Einsen als gleich wahrscheinlich angesehen. Der Erwartungswert der längsten 1-Folge hat als asymptotisches Äquivalent*

$$\log_2 n + \frac{\gamma}{\log 2} - \frac{1}{2} - P(\log_2 n)$$

mit der periodischen Funktion aus Satz 1.

Bemerkung: Würde man der Zahl n nicht die Folgen der Länge 2n, sondern die der Länge n zuordnen, hätte man in der asymptotischen Formel 1 abzuziehen, sodaß dann die Übereinstimmung von den Ergebnissen in den Kapiteln 2 und 3 (in den ersten beiden Termen der asymptotischen Entwicklung) gegeben ist.

LITERATUR

[1] P.FLAJOLET, M.REGNIER, R.SEDGEWICK, Some Uses of the Mellin Integral Transform in the Analysis of Algorithms, in: Proceedings of NATO Advanced Research Workshop on Combinatorial Algorithms on Words, Springer Verlag (1985); Erweiterte Version in Vorbereitung.

[2] P.FLAJOLET, H.PRODINGER, On the height of simply generated families of lattice paths, manuscript in preparation.

[3] M.L.HAUTUS,D.A.KLARNER, The Diagonal of a Double Power Series, Duke Math.J.38 (1971), 229-235.

[4] D.E.KNUTH, The Art of Computer Programming, Vol.3, Reading, Mass.: Addison-Wesley, 1973.

[5] D.E. KNUTH, The Average Time for Carry Propagation, Indagationes Mathematicae 40 (1978), 238-242.

[6] K.A.POST, Binary Sequences with Restricted Repetitions, T.H.Report 74-WSK-02, Eindhoven, 1974.

[7] D.H.GREENE, D.E.KNUTH, Mathematics for the Analysis of Algorithms, Birkhäuser, 1982.

Helmut Prodinger
Inst.f.Algebra u.Diskrete Mathematik
der Technischen Universität Wien
Wiedner Hauptstr.8-10
1040 WIEN