Horton Strahler Zahlen

Manuel Inselmann

3.11.2011

1 Einführung

In diesem Seminarvortrag wird als Hauptresultat gezeigt, dass
$P_n(|S_n - log_4 n| \geq \epsilon) \leq \frac{D}{4^\epsilon}$ für ein $D > 0$ und alle $\epsilon > 0$ gilt, wobei P_n die Gleichverteilung auf der Menge aller Bäume mit n Knoten und $S_n(T)$ die Horton-Strahler Zahl eines Baumes T sei.
Zunächst sei an den Begriff eines Baumes erinnert:

Definition 1.1. $2^{<\mathbb{N}_0} := \bigcup_{n\in\mathbb{N}_0}\{0,1\}^n$ ist der vollständige binäre Baum.
Die Elemte von $2^{<\mathbb{N}_0}$ sind also von der Form
$(i_0, i_1, ..., i_{n-1})$ mit $i_k \in \{0,1\}$. Das einzige Element aus $\{0,1\}^0$ sei mit $\varnothing$ bezeichnet.
Für $t = (i_0, i_1, ..., i_{n-1}) \in 2^{<\mathbb{N}_0}$ und $s = (j_0, j_1, ..., j_{m-1}) \in 2^{<\mathbb{N}_0}$ sei $t^\frown s = (i_0, i_1, ..., i_{n-1}, j_0, j_1, ..., j_{m-1})$, (insbesondere gelte $t^\frown\varnothing = \varnothing^\frown t = t$).
Außerdem soll $s \leq t$ bedeuten, dass es ein $u \in 2^{<\mathbb{N}_0}$ gibt mit $s^\frown u = t$ und $s < t$ soll heißen, dass $s \leq t$ und $s \neq t$ gilt.

Nun können wir definieren, was ein Baum ist:

Definition 1.2. Eine endliche Teilmenge T von $2^{<\mathbb{N}_0}$ heißt Baum, falls für alle $t \in T$ und $s \in 2^{<\mathbb{N}_0}$ mit $s \leq t$ folgt, dass $s \in T$ gilt. Die $t \in T$ heißen Knoten von T.

Definition 1.3. $\Omega = \{T \subset 2^{<\mathbb{N}_0} | T\ ist\ Baum\}$

Definition 1.4. Sei T ein Baum und $t \in 2^{<\mathbb{N}_0}$.
Definiere $T_t := \{s \in 2^{<\mathbb{N}_0} | t^\frown s \in T\}$
und $t^\frown T := \{t^\frown s | s \in T\}$.

Definition 1.5. $r : \Omega \setminus \{\emptyset\} \to \Omega$

$$r(T) = \begin{cases} \emptyset & \text{falls } T = \{\varnothing\} \\ r(T_i) & \text{falls } i \in T \text{ und } (1-i) \notin T\ (i \in \{0,1\}) \\ \{\emptyset\} \cup (0)^\frown r(T_{(0)}) \cup (1)^\frown r(T_{(1)}) & \text{falls } (0) \in T \text{ und } (1) \in T \end{cases}$$

Definition 1.6. Die Abbildung $S : \Omega \to \mathbb{N}_0$ ordnet jedem Baum seine Horton-Strahler-Zahl zu. Sie ist rekursiv definiert durch:
$S(\emptyset) = 0$
$S(T) = S(r(T)) + 1$.
Also ist $S(T)$ das minimale n, für das $r^{\circ n}(T) = \emptyset$ ist.
Mit Induktion nach der Anzahl der Knoten zeigt man leicht, dass man $S(T)$

rekursiv auch wie folgt definieren kann:
$S(\emptyset) = 0$
$S(T) = max\{S(T_{(0)}), S(T_{(1)})\} + 1_{\{S(T_{(0)})=S(T_{(1)})\}}$.

Beispielsweise ist $S(T) = 1$ genau dann, wenn $(T, \leq)$ linear ist. Das andere Extrembeispiel ist $T = \cup_{k<n}\{0,1\}^k$ mit $S(T) = n$. Man beachte, dass in diesem Fall $|T| = 2^n - 1$ gilt.
Allgemein gilt für alle $T \in \Omega$: $S(T) \leq log_2(|T| + 1)$,
denn dies ist klar für $|T| \in \{0, 1\}$ und für $S(T_{(0)}) \neq S(T_{(1)})$ und ansonsten gilt:
$S(T) \leq S(T_{(0)}) + 1 \leq log_2(\frac{|T|-1}{2} + 1) + 1 = log_2(|T| + 1)$
(denn $min\{|T_{(0)}|, |T_{(1)}|\} \leq \frac{|T|-1}{2}$).

Vorarbeiten

In diesem Abschnitt wird der die Funktion r, die ja gemäß $S(T) = S(r(T)) + 1$ mit S zusammenhängt, weiter untersucht. Dazu wird die Funktion l eingeführt, die jedem Baum die Anzahl seiner Blätter zuordnet. Für l gilt nämlich
$|r(T)| = l(T) - 1$ und eine genauere Analyse von l ist der letzlich der Schlüssel zum Beweis des Hauptresultats.

Definition 1.7. Sei T ein Baum und $t \in 2^{<\mathbb{N}_0}$. Dann heißt t externer Knoten von T, falls $t \notin T$ und $\forall s < t\ s \in T$ gilt.
Weiter sei $E(T) := \{t \in 2^{<\mathbb{N}_0} | t$ ist externer Knoten von $T\}$
und $e(T) = |E(T)|$.

Bemerkung 1.8. Für alle $T \in \Omega$ gilt: $e(T) = |T| + 1$.

Beweis. $T = \emptyset$ ist offensichtlich.
Ansonsten gilt nach Induktion nach der Anzahl der Knoten:
$e(T) = e(T_{(0)}) + e(T_{(1)}) = |T_{(0)}| + |T_{(1)}| + 2 = |T| + 1$. □

Definition 1.9. Sei T ein Baum und $t \in 2^{<\mathbb{N}_0}$. Dann heißt t Blatt von T genau dann, wenn $t \in T$ und $t^\frown(0), t^\frown(1) \notin T$ gilt.
Sei $L(T) := \{t \in 2^{<\mathbb{N}_0} | t$ ist ein Blatt von $T\}$ und $l(T) = |L(T)|$.

Bemerkung 1.10. Sei $T \in \Omega \setminus \{\emptyset\}$. Dann gilt $e(r(T)) = l(T)$

Beweis. Der Beweis erfolgt wieder über Induktion nach der Anzahl der Knoten.
Der Fall $T = \{\varnothing\}$ ist klar.
Für alle anderen T gilt: $l(T) = l(T_{(0)}) + l(T_{(1)}) = e(r(T))$. □

Bemerkung 1.11. Zusammengesetzt ergibt sich: $|r(T)| = l(T) - 1$.

Definition 1.12. Für $n \in \mathbb{N}_0$ sei $\Omega_n := \{T \in \Omega||T| = n\}$ und P_n die Gleichverteilung auf Ω_n. Ferner sei r_n die Einschränkung von r auf Ω_n.

Lemma 1.13. *Es gilt* $\mathbb{P}(r_n = \cdot||r_n| = k) = P_k(\cdot)$

Beweis. Es ist zu zeigen, dass jedes $T \in \Omega_k$ unter r_n gleich viele Urbilder besitzt.
Definiere dazu $i_k : \Omega_k \to \Omega_{2k+1}$;
$T \mapsto T \cup E(T)$,
sowie $i : \Omega \to \Omega$;
$T \mapsto T \cup E(T)$.
Dann gilt $r \circ i_k = id_{\Omega_k}$. Dies zeigt, dass jeder Baum $T \in \Omega_k$ genau ein r-Urbild in $i(\Omega_k)$ besitzt.
Sei nun $q : \Omega \to \Omega$

$$q(T) = \begin{cases} \{\emptyset\} & \text{falls } T = \emptyset \\ \{\emptyset\} & \text{falls } T = \{\emptyset\} \\ q(T_i) & \text{falls } i \in T \text{ und } (1-i) \notin T \\ \{\emptyset\} \cup 0^\frown q(T_0) \cup 1^\frown q(T_1) & \text{falls } 0 \in T \text{ und } 1 \in T \end{cases}$$

Sei q_n die Einschränkung von q auf Ω_n.
q_n ist für $n > 0$ also die Abbildung, die in einem Baum genau diejenigen Knoten „weglässt“, die genau ein Kind besitzen.
Damit folgt, dass $q(\Omega) = i(\Omega)$ gilt, sowie andererseits $r_n = r \circ q_n$. Wenn nun $Q \in \Omega_n$ ein r-Urbild zu $T \in \Omega_k$ ist, so gilt $q(Q) = i_k(T)$. Es reicht also zu zeigen, dass jeder Baum $i_k(T)$ unter q gleich viele Urbilder in Ω_n besitzt. Dies folgt aber daraus, dass jedes q-Urbild aus $i_k(T)$ hervorgeht, indem über jeden Knoten, der kein Blatt ist, eine lineare Folge von Knoten aufgesetzt wird und der Baum dann wie er vorher war, fortgesettz wird (formaler: $T \mapsto t^\frown s^\frown T_s$, wobei s der linearen Folge entspricht und t dem Knoten, der kein Blatt ist). Um dabei ein Baum mit n Knoten zu erhalten, muss man $n - 2k - 1$ Knoten hinzufügen und hierfür gibt es unabhängig von T immer $\binom{n-1}{2k} 2^{n-2k-1}$ Möglichkeiten. □

Wegen der Beziehungen $|r(T)| = l(T) - 1$, $\{S(T) > 0\} = \{|r^{\circ k}(T)| > 0\}$ und $S(T) = S(r(T)) + 1$, macht es Sinn, sich nun zunächst mit der Funktion l zu beschäftigen, bzw. mit ihrer Einschränkung auf Ω_n, die mit l_n bezeichnet sei.

Definition 1.14. $Q_{nk} := \{T \in \Omega_n | l(T) = k\}$;
$q_{nk} := |Q_{nk}|$
$Q(w, z) = \sum\limits_{n,k \geq 0} q_{nk} w^n z^k$.

Bemerkung 1.15. Es gilt für $n > 1$ und $k \in \mathbb{N}_0$:

$$q_{nk} = \sum_{n_1+n_2=n-1, k_1+k_2=k} q_{n_1k_1} q_{n_2k_2}$$

Beweis. Definiere $+ : \Omega \times \Omega \to \Omega$ mit $(T, S) \mapsto \{\varnothing\} \cup (0)^\frown T \cup (1)^\frown S$
Dann gilt für $n > 1$:

$$Q_{nk} = \bigcup_{n_1+n_2=n-1, k_1+k_2=k} (Q_{n_1k_1} + Q_{n_2k_2})$$

Da diese Vereinigung disjunkt ist, folgt die Behauptung. □

Bemerkung 1.16. Es gilt in einer offenen Umgebung von $(0,0)$:
$Q(w,z) = \frac{1-\sqrt{1-4w(wz-w+1)}}{2w}$.

Beweis. Es gilt: $q_{00} = 1$, $q_{10} = 0$, $q_{1n} = 0$ für $n > 1$ und $q_{0n} = 0$ für $n > 0$.
Hieraus und aus obiger Bemerkung erhält man:

$$
\begin{aligned}
&(2wQ(w,z)-1)^2 \\
&= 4 \sum_{n,k\geq 0} \sum_{n_1+n_2=n, k_1+k_2=k} q_{n_1k_1} q_{n_2k_2} w^{n+2} z^k - 4 \sum_{n,k\geq 0} q_{nk} w^{n+1} z^k + 1 \\
&= 1 + 4w^2 - 4w - 4w^2 z = 1 - 4w(wz-w+1)
\end{aligned}
$$

$Q'(w,z) := \frac{1-\sqrt{1-4w(wz-w+1)}}{2w}$ erfüllt diese Gleichung anstelle von $Q(w,z)$ ebenso. Ferner ist Q' in einer Umgebung von $(0,0)$ in der Form

$$Q'(w,z) = \sum_{n,k\geq 0} q'_{nk} w^n z^k$$

darstellbar, wobei man aus

$$Q'(w,z) = \frac{1 - \sum\limits_{k\geq 0} \binom{\frac{1}{2}}{k} (-1)^k (4w(wz-w+1))^k}{2w}$$

abliest, dass $q'_{nk} = q_{nk}$ für alle $n \leq 1$ und $k \in \mathbb{N}_0$ gilt.
Da die q'_{nk} für $n > 1$ aber auch obige Rekursionsgleichung erfüllen und diese durch diese Informationen schon vollständig bestimmt sind, folgt $q'_{nk} = q_{nk}$ für alle $n,k \in \mathbb{N}_0$. $\square$

Bemerkung 1.17. Es gilt für alle $n \in \mathbb{N}_0$:
$E_n(l_n) = \frac{n(n+1)}{2(2n-1)}$, sowie
$Var_n(l_n) = \frac{n(n+1)(n-1)(n-2)}{2(2n-3)(2n-1)^2}$.

Beweis. Die Behauptungen ergeben sich aus der Beziehung

$$Q(w,z) = \frac{1-\sqrt{1-4w(wz-w+1)}}{2w}$$

So ist

$$E_n(l_n) = \frac{\sum\limits_{k\geq 0} kq_{nk}}{\sum\limits_{k\geq 0} q_{nk}}$$

und $\sum\limits_{k\geq 0} q_{nk}$ ist der n-te Koeffizient in der Potenzreihenentwicklung von $Q(w,1) = \frac{1-\sqrt{1-4w}}{2w}$ und $\sum\limits_{k\geq 0} kq_{nk}$ ist der n-te Koeffizient in der Potenzreihenentwicklung

von $\frac{\partial Q}{\partial z}(w,1) = \frac{w}{\sqrt{1-4w}}$.
Also

$$E_n(l_n) = \frac{\binom{-\frac{1}{2}}{n-1}(-4)^{n-1}}{2\binom{\frac{1}{2}}{n+1}(-4)^n} = \frac{n(n+1)}{2(2n-1)}$$

Analog ist

$$Var_n(l_n) = \frac{\sum\limits_{k\geq 0} k(k-1)q_{nk}}{\sum\limits_{k\geq 0} q_{nk}} + E_n(l_n)(1-E_n(l_n))$$

und $\sum\limits_{k\geq 0} k(k-1)q_{nk}$ ist der n-te Koeffizient der Potenzreihenentwicklung von

$$\frac{\partial^2 Q}{\partial^2 z}(w,1) = 2w^3(1-4w)^{-\frac{3}{2}}$$

also $\sum\limits_{k\geq 0} k(k-1)q_{nk} = 2(-4)^{n-3}\binom{-\frac{3}{2}}{n-3}$.

Setzt man dies in $Var_n(l_n) = \frac{\sum\limits_{k\geq 0} k(k-1)q_{nk}}{\sum\limits_{k\geq 0} q_{nk}} + E_n(l_n)(1-E_n(l_n))$ ein und formt um, erhält man die Behauptung. $\square$

Hauptresultat

Satz 1.18. *Es gilt für alle $\epsilon \geq 0$ und $n \in \mathbb{N}$:*

$$P_n(S_n > \lceil log_4 n + \epsilon \rceil) \leq \frac{1}{4^\epsilon}$$

Beweis. Setze $T_k := r^{\circ k}(T)$.
Dann gilt wegen Lemma 1.13 und $\frac{n(n+1)}{2(2n-1)} - 1 \leq \frac{n}{4}$:

$$E|T_{k+1}| = E(l(T_k)-1)1_{\{|T_k|>0\}} = E(E(l_j-1)||T_k|=j) = EE_j(l_j-1) = E(\frac{j(j+1)}{2(2j-1)} - 1)1_{\{j>0\}}$$
$$\leq E\frac{|T_k|}{4}$$

Induktiv folgt: $E|T_k| \leq \frac{n}{4^k}$.
Wegen $P_n(|T_k| \geq 1) \leq E|T_k|$ und $\{S_n > k\} = \{|T_k| > 0\}$ folgt für $k := \lceil log_4 n + \epsilon \rceil$):

$$P_n(S_n > \lceil log_4 n + \epsilon \rceil) \leq \frac{n}{4^{\lceil log_4 n + \epsilon \rceil)}} \leq \frac{1}{4^\epsilon}$$

$\square$

Satz 1.19. *Es gibt ein $C > 0$, so dass für alle $\epsilon \geq 0$*

$$P_n(S_n \leq \lfloor log_4 n - \epsilon \rfloor) \leq \frac{C}{4^\epsilon}$$

gilt.

Beweis. Mit $Var(X|Y) = E(X^2|Y) - (E(X|Y))^2$ hat man:

$$Var|T_{k+1}| = Var(E(|T_{k+1}|||T_k|)) + E(Var(|T_{k+1}|||T_k|))$$

Wegen Bemerkung 1.17 und Lemma 1.13 gilt:

$$E(Var(|T_{k+1}|||T_k|)) = E(E(|T_{k+1}|^2||T_k| = j) - E(|T_{k+1}|||T_k| = j)^2)$$
$$= EVar(l_j - 1)1_{\{j>0\}} = E\frac{|T_k|(|T_k|+1)(|T_k|-1)(|T_k|-2)}{2(2|T_k|-3)(2|T_k|-1)^2} \leq \frac{1}{8}E|T_k|$$

Nun gilt für

$$X = \frac{1}{4}1_{\{|T_k \in \{1,2,3\}\}} + 1_{\{|T_k|>3\}}$$

dass $EX \leq 1$ und andererseits:

$$0 \leq \frac{|T_k|}{4} - X \leq E(|T_{k+1}|||T_k|) \leq \frac{|T_k|}{4}$$

Hieraus folgt:

$$Var(E(|T_{k+1}|||T_k|)) = E(E(|T_{k+1}|||T_k|))^2 - (E(E(|T_{k+1}|||T_k|))^2 \leq$$
$$E(\frac{|T_k|}{4})^2 - (E(\frac{|T_k|}{4} - X))^2 = Var\frac{|T_k|}{4} + 2E\frac{|T_k|}{4}EX - (EX)^2 \leq Var\frac{|T_k|}{4} + 2E\frac{|T_k|}{4}$$

Insgesamt ergibt sich:

$$Var|T_{k+1}| \leq cE|T_k| + \frac{1}{16}Var|T_k| \leq \frac{cn}{4^k} + \frac{1}{16}Var|T_k|$$

(mit $c = \frac{5}{8}$).
Induktiv ergibt sich

$$Var|T_{k+1}| \leq \frac{cn}{4^k}\frac{1-4^{-k-2}}{1-\frac{1}{4}} \leq \frac{4cn}{3 \cdot 4^k}$$

Wie im Beweis des vorherigen Satzes zeigt man mithilfe von $\frac{n(n+1)}{2(2n-1)} \geq \frac{n}{4}$, dass

$$E|T_{k+1}| \geq E\frac{|T_k|}{4} - 1$$

gilt, also

$$E|T_{k+1}| \geq \frac{n}{4^{k+1}} - \frac{4}{3}$$

Nun kommen wir zur Abschätzung:

$$\begin{aligned} P_n(S_n \leq k) &= P_n(|T_k| = 0) = P_n(|T_k| - E|T_k| = -E|T_k|) \\ \leq P_n(||T_k| - E|T_k|| \geq E|T_k|) &\leq \frac{Var|T_k|}{(E|T_k|)^2} \leq \frac{4cn}{3 \cdot 4^k} \frac{1}{(\frac{n}{4^k} - \frac{4}{3})^2} \end{aligned}$$

Für $k = \lfloor log_4 n - x \rfloor$ erhält man für $x \geq 1$ (woraus : $4^x - \frac{4}{3} > 0$ folgt):

$$P_n(S_n \leq \lfloor log_4 n - x \rfloor) \leq \frac{4c}{3 \cdot 4^{-x-1}} \frac{1}{(4^x - \frac{4}{3})^2} \leq 12c \frac{1}{4^x}$$

(da $4^x - \frac{4}{3} \geq \frac{2}{3} 4^x$ gilt für $x \geq 1$).
Also gilt

$$P_n(S_n \leq \lfloor log_4 n - x \rfloor) \leq \frac{C}{4^x}$$

für $C = \frac{15}{2}$, falls $x \geq 1$, aber die Behauptung gilt trivialerweise auch für $x < 1$, für jedes $C \geq 4$. □

Die beiden vorangegangenen Sätze liefern unmittelbar:

Satz 1.20. *Es gibt ein $D > 0$, so dass für alle $x > 0$ gilt:*
$P_n(|S_n - log_4 n| \geq x) \leq \frac{D}{4^x}$.

Korollar 1.21. *Für alle $s > 0$ gilt:*

$$E_n |S_n - log_4 n|^s = O(1)$$

.

Beweis. Es gilt:

$$\begin{aligned} E_n |S_n - log_4 n|^s = \int_0^\infty s t^{s-1} P_n(|S_n - log_4 n| > t) \leq D \int_0^\infty s t^{s-1} 4^{-t} = D(ln4)^s \int_0^\infty s t^{s-1} e^{-t} \\ = Ds(ln4)^s \Gamma(s) < \infty \end{aligned}$$

□

Korollar 1.22. *Für alle $\lambda < log4$ gilt:*

$$E e^{\lambda |S_n - log_4 n|} < \infty$$

Beweis. Es gilt:

$$\begin{aligned} E e^{\lambda |S_n - log_4 n|} &\leq \sum_{k=0}^\infty e^{\lambda(k+1)} P_n(k \leq |S_n - log_4 n| < k+1) \\ &\leq D e^\lambda \sum_{k=0}^\infty (\frac{e^\lambda}{4})^k = \frac{D e^\lambda}{1 - \frac{e^\lambda}{4}} \end{aligned}$$

für $\lambda < log4$. □

Literatur

[1] Luc Devroye, Paul Kruszewski : *A note on the Horton-Strahler number for random trees*,