

Rudolf Grübel

## Zufällige binäre Bäume: Von der average-case Analyse zur Verteilungsasymptotik

**Zusammenfassung** Wir betrachten einige klassische Such- und Sortieralgorithmen bei zufälligem Input. Zählt  $X_n$ , wie viele Operationen eines bestimmten Typs bei einem Datensatz vom Umfang  $n$  benötigt werden, so geht es bei der average-case Analyse um den Erwartungswert  $EX_n$  dieser Zufallsgröße und insbesondere um das Verhalten von  $EX_n$  mit  $n \rightarrow \infty$ . Wir stellen verschiedene Methoden vor, mit denen in vielen solchen Fällen das asymptotische Verhalten der gesamten Verteilung von  $X_n$  bestimmt werden kann. Aussagen über die Verteilungskonvergenz lassen sich beispielsweise verwenden, um die Wahrscheinlichkeit exzessiv langer Laufzeiten zu approximieren.

### 1 Einleitung

An der Schnittstelle der Gebiete Diskrete Mathematik, Stochastik und Theoretische Informatik hat sich in den letzten Jahren viel getan. Eine bewusst kurze und in vieler Hinsicht weit gespannte Liste von Monographien könnte aus den Büchern von Alon und Spencer (2000) zur probabilistischen Methode, Janson, Łuczak and Ruciński (2000) zu zufälligen Graphen sowie Motwani und Raghavan (1995) zu randomisierten Algorithmen bestehen. Bei einem ersten Kontakt mit diesem Bereich fällt auf, welcher durchschlagende Erfolg selbst einfachste Sachverhalte aus der Stochastik haben können: Zählt beispielsweise die Zufallsvariable  $X$  die Anzahl der Objekte eines bestimmten Typs in irgendeinem Zufallsexperiment und stellt sich für  $X$  ein Erwartungswert größer als 0 heraus, so kann  $X$  nicht mit Wahrscheinlichkeit 1 den Wert 0 annehmen — es muss also mindestens ein solches Objekt existieren! Das Buch von Aigner und Ziegler (1998) enthält hierzu einige konkrete und leicht zugängliche Beispiele; der attraktiven Kombination von pfiffig-genialer Idee und geringem technischen Aufwand bei der Umsetzung kann sich wohl kaum

---

R. Grübel  
Institut für Mathematische Stochastik, Universität Hannover, Postfach 60 09,  
30060 Hannover, Deutschland  
e-mail: [rgrubel@stochastik.uni-hannover.de](mailto:rgrubel@stochastik.uni-hannover.de)

ein Mathematiker entziehen. Dies steht in einem interessanten Kontrast zu anderen innermathematischen Anwendungen der Stochastik, beispielsweise dem Beweis des Hypoelliptizitätssatzes von Hörmander in der Theorie der partiellen Differentialgleichungen mit dem Malliavin-Kalkül oder dem Beweis der Indexformel von Atiyah und Singer in der algebraischen Geometrie mit der Brownschen Bewegung, bei denen der Weg von der Problemstellung über das Verständnis der stochastischen Konstruktion bis hin zur schließlichen Anwendung lang und steinig sein kann.

In der vorliegenden Arbeit soll von einem Ideenkreis berichtet werden, der zwischen den angedeuteten Extrema angesiedelt ist, also einen Blick auf nicht-triviale Methoden der Stochastik erlaubt, und dies bei der Anwendung auf Probleme, die durchaus auch interessierten Laien erklärt werden können. Konkret geht es um eine bestimmte Klasse von binären Bäumen, die bei den klassischen Aufgaben des Suchens und Sortierens eine zentrale Rolle spielt. Diese Strukturen ermöglichen unter anderem die probabilistische Analyse der Algorithmen `BINARY SEARCH TREE`, `QUICKSORT` und `FIND`, alle drei seit langem in Theorie und Praxis von großer Bedeutung. Die Laufzeit solcher Algorithmen, im wesentlichen die Zahl der Operationen eines bestimmten Typs, hängt vom Input ab. Bei zufälligem Input (oder bei einer randomisierten Version des jeweiligen Verfahrens) sind diese Bäume zufällig, und als Konsequenz ist die Anzahl der benötigten Operationen bei Inputumfang  $n$  eine Zufallsvariable  $X_n$ . Lange Zeit stand die average-case Analyse, also die Betrachtung des Erwartungswertes  $EX_n$  solcher Zufallsgrößen, im Mittelpunkt; diese verrät bereits mehr über einen Algorithmus als eine worst-case Analyse, bei der es um Obergrenzen für den Wert von  $X_n$  geht. In den letzten etwa 15 Jahren hat man einige Erfolge bei der Analyse der Verteilungen dieser Zufallsgrößen erzielt, wobei ganz unterschiedliche Methoden der klassischen und der modernen Stochastik angewendet wurden. Über die average-case Analyse hinausgehend können mit solchen Resultaten auch brauchbare Approximationen für die Wahrscheinlichkeit exzessiv langer Laufzeiten erhalten werden.

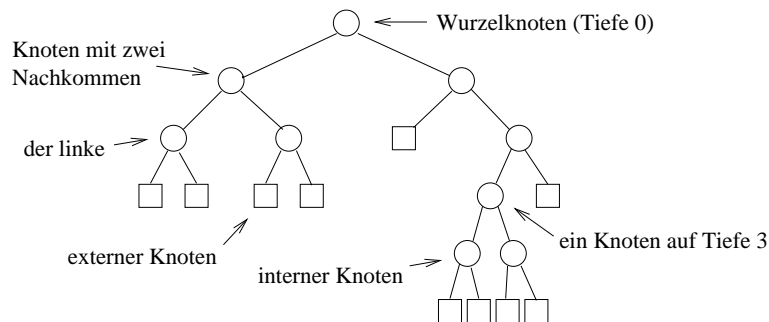


ABBILDUNG 1: Ein binärer Baum

Auf diese Entwicklungen wollen wir im folgenden einen Blick werfen. Unser Ziel ist dabei, ausgehend von Grundkenntnissen der elementaren Stochastik ein Verständnis der verwendeten Methoden zu erreichen. Gelegentlich werden wir ein technisches Detail ignorieren und auch bei manchen Definitionen nicht die größtmögliche Allgemeinheit anstreben.

Obwohl wir uns bei den diskreten Strukturen auf binäre Bäume, bei den algorithmischen Anwendungen auf bestimmte Such- und Sortieralgorithmen und bei den Resultaten auf Verteilungskonvergenz beschränken, müssen wir trotzdem noch aus dem umfangreichen und interessanten Material auswählen. Im Sinne einer flüssigen Darstellung werden wir zudem Literaturhinweise auf einen Schlussabschnitt verschieben.

## 2 Binäre Bäume

Bekanntlich sagt ein Bild mehr als tausend Worte: Abbildung 1 illustriert die im folgenden verwendeten Begriffe. Wir beeilen uns aber, (konstruktive) formale Definitionen aller Elemente dieser Zeichnung nachzuliefern. Ein *Knoten* ist eine endliche Folge  $u = (u_1, \dots, u_k) \in \{0, 1\}^k$  von Nullen und Einsen,  $k$  ist die *Tiefe* des Knotens. Im Falle  $k = 0$  erhalten wir die leere Menge als Repräsentanten des Wurzelknotens. Man kann sich  $u$  als Wegbeschreibung vorstellen:  $u_i = 0$  heißt, dass man auf dem Weg vom Wurzelknoten zum Knoten  $u$  im  $i$ -ten Schritt nach links geht, bei  $u_i = 1$  geht man nach rechts. Den *linken* und *rechten Nachkommen* zu  $u$  erhält man als  $(u_1, \dots, u_k, 0)$  bzw.  $(u_1, \dots, u_k, 1)$ . Im Falle  $k > 0$  sind die Knoten  $(u_1, \dots, u_i)$  mit  $i \in \{0, \dots, k-1\}$  die *Vorfahren* von  $u$ ,  $(u_1, \dots, u_{k-1})$  ist der *direkte Vorfahr* von  $u$ . Ein *binärer Baum* ist nun schlicht eine endliche Teilmenge  $T$  von  $\bigcup_{k=0}^{\infty} \{0, 1\}^k$  mit der Eigenschaft, dass mit jedem  $u \in T \setminus \{\emptyset\}$  auch der direkte Vorfahr von  $u$  ein Element von  $T$  ist. Die Elemente von  $T$  heißen *interne Knoten*. Als *externe Knoten* bezeichnen wir im Falle  $T \neq \emptyset$  die Menge aller  $u \in \bigcup_{k=1}^{\infty} \{0, 1\}^k$ , die nicht in  $T$  liegen, deren direkter Vorfahr aber ein Element von  $T$  ist; der Wurzelknoten ist der einzige externe Knoten des leeren Baumes. Die Menge der externen Knoten kann als *Rand*  $\partial T$  von  $T$  aufgefasst werden. Wenn nicht weiter spezifiziert, sind im folgenden mit Knoten immer interne Knoten gemeint. Wir schreiben generell  $|A|$  für die Anzahl der Elemente einer Menge  $A$ ; bei einem binären Baum  $T$  ist also  $|T|$  die Anzahl der (internen) Knoten von  $T$ .

In der Sprache der Theoretischen Informatik ist ein binärer Baum eine präfix-stabile endliche Menge von Wörtern über dem Alphabet  $\{0, 1\}$ , die Grundmenge  $\bigcup_{k=0}^{\infty} \{0, 1\}^k$  wird gelegentlich zu  $\{0, 1\}^*$  abgekürzt. Es ist natürlich eine für die Praxis sehr wichtige Frage, wie man eine solche fundamentale Datenstruktur beispielsweise über Felder und Zeiger implementiert; diesen Aspekt werden wir aber im folgenden ignorieren. In den Anwendungen enthalten die Knoten in der Regel einen Wert; ein Paar  $(T, \phi)$  mit einem binären Baum  $T$  im obigen Sinn und einer Abbildung  $\phi : T \rightarrow \mathbb{R}$  bezeichnen wir als *bewerteten Baum* (der Wertebereich  $\mathbb{R}$  ist dabei nur eine für unsere Zwecke bequeme Festsetzung).

Die (internen oder externen) Knoten eines binären Baumes lassen sich in kanonischer Weise anordnen, wenn man die zu einem  $u$  gehörende Wegbeschreibung in passend kleiner werdende Schritte nach links bzw. rechts übersetzt. Konkret erhält man eine solche Anordnung beispielsweise durch die Funktion

$$\Psi : \{0, 1\}^* \rightarrow (0, 1), \quad u = (u_1, \dots, u_k) \mapsto \frac{1}{2} + \sum_{j=1}^k \frac{2u_j - 1}{2^{j+1}}. \quad (1)$$

In Abbildung 2 ist diese Zuordnung durch die senkrechten gepunkteten Linien angedeutet. Auch zu diesem mathematischen Formalismus gibt es ein Pendant in der

Theoretischen Informatik (oder Algorithmischen Graphentheorie): Die so erhaltene Anordnung entspricht der Knotenreihenfolge, wenn man den Baum ‘in-order’ durchquert. Diese Durchquerung wird rekursiv definiert durch die Vorschrift ‘linken Teilbaum durchqueren, Wurzel besuchen, rechten Teilbaum durchqueren’.

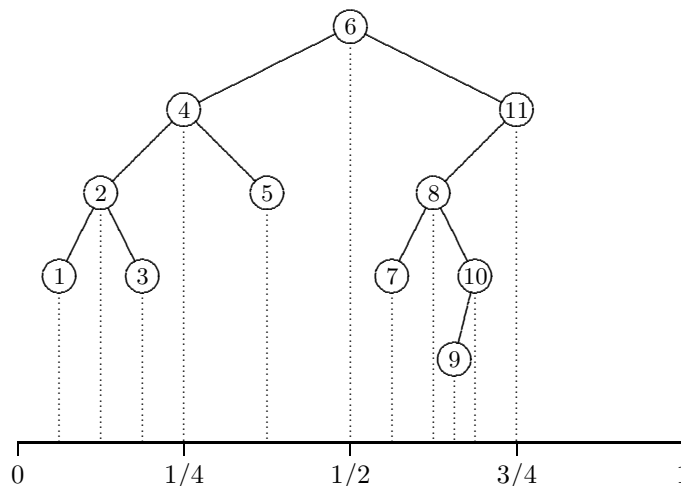


ABBILDUNG 2: Ein bewerteter Baum

In den Anwendungen geht es häufig nicht um den genauen Baum  $T$  oder bewerteten Baum  $(T, \phi)$ , sondern um bestimmte Größen, die als Funktionen von  $T$  oder  $(T, \phi)$  geschrieben werden können. Von besonderer Bedeutung ist die Summe der Tiefen aller (internen) Knoten, die *interne Pfadlänge* von  $T$ . Als Hilfsmittel bei der Analyse dieser Größe wird sich die *externe Pfadlänge*, also die Summe der Tiefen der externen Knoten, als hilfreich erweisen. Von Interesse sind weiter die *Höhe* des Baumes, definiert als das Maximum der Tiefen aller Knoten, oder bei bewerteten Bäumen die Tiefe des Knotens, der einen bestimmten Wert enthält. Weitere solche Größen werden wir in den folgenden Abschnitten kennenlernen.

### 3 Dynamisches Verhalten

Jeder binäre Baum  $T$  lässt sich in  $|T|$  Schritten durch sukzessives Umwandeln eines externen in einen internen Knoten aus dem leeren Baum erhalten. Bereits diese elementare Beobachtung zur Dynamik hat interessante Konsequenzen: Der leere Baum hat 0 interne und einen externen Knoten, beim Umwandeln steigt die Anzahl der internen Knoten um 1, bei den externen verliert man einen und erhält zwei neue; es folgt also  $|\partial T| = |T| + 1$ . Die Differenz der externen und der internen Pfadlänge von  $T$  beträgt  $2|T|$ , denn jede Umwandlung eines externen in einen internen Knoten vergrößert diese Differenz um den Wert 2. Zur Kontrolle: Der Baum in Abbildung 1 hat 9 interne und 10 externe Knoten, die interne Pfadlänge ist 19, die externe 37.

Eine erste wichtige Anwendung bewerteter binärer Bäume ergibt sich im Zusammenhang mit der sequentiellen Speicherung von Daten  $x_1, x_2, \dots$ , wobei wir

einfachheitshalber annehmen, dass diese paarweise verschiedene reelle Zahlen sind. Der *BST-Algorithmus* (das Akronym BST steht für *binary search tree*) konstruiert hierzu eine Folge  $(T_1, \phi_1), (T_2, \phi_2), \dots$  von bewerteten binären Bäumen:  $T_1$  besteht nur aus dem Wurzelknoten mit Inhalt  $x_1$ . Hat man  $T_n$  zu  $x_1, \dots, x_n$  konstruiert, so ergibt sich  $T_{n+1}$  wie in der oben beschriebenen allgemeinen Situation durch Umwandlung eines externen Knotens von  $T_n$  in einen internen: Dieser wird bestimmt, indem man den nächsten einzufügenden Wert  $x_{n+1}$  nacheinander, beginnend mit dem Wurzelknoten, mit dem Inhalt eines Knotens vergleicht, nach links bei Resultat ‘kleiner’, nach rechts bei Resultat ‘größer’ geht, bis schließlich der gewünschte externe Knoten erreicht ist. Diesem wird durch  $\phi_{n+1}$  der Wert  $x_{n+1}$  zugeordnet, auf den Knoten von  $T_n$  stimmt  $\phi_{n+1}$  mit  $\phi_n$  überein. Beispielsweise erhält man bei 6, 11, 8, 4, 10, 2, 3, 1, 7, 5, 9, ... als  $T_{11}$  die in Abbildung 2 wiedergegebene Struktur. Beim Einfügen des letzten Wertes 9 vergleicht man diesen nacheinander mit den Werten 6, 11, 8 und 10, geht nach rechts, links, rechts und wieder links, und landet so schließlich im linken Nachkommen von 10.

Nach Konstruktion spielt für den Baum  $T_n$  selbst nur das Größenverhältnis der ersten  $n$   $x$ -Werte untereinander eine Rolle; insbesondere ändert sich  $T_n$  nicht, wenn man  $x_1, \dots, x_n$  derselben streng monotonen Transformation unterwirft. Eine solche erhält man beispielsweise durch den Übergang zu den *absoluten Rängen*,

$$(x_1, \dots, x_n) \mapsto \pi_n = (\pi_{n,1}, \dots, \pi_{n,n}) \quad \text{mit } \pi_{n,i} := |\{1 \leq j \leq n : x_j \leq x_i\}|.$$

Jeder der möglichen Werte taucht genau einmal als absoluter Rang auf,  $\pi_n$  ist also eine Permutation der Menge  $\{1, \dots, n\}$ . Unter dem *relativen Rang* des Wertes  $x_i$  (in der gesamten Folge, hier muss  $n$  nicht mitgeschleppt werden) versteht man

$$R(x_i) := |\{1 \leq j \leq i : x_j \leq x_i\}|.$$

Die Originaldaten sowie die zugehörigen absoluten und relativen Ränge zu dem Baum in Abbildung 2 sind in der folgenden Tabelle aufgeführt, wobei die  $x$ -Werte auf drei Nachkommastellen verkürzt wurden:

$i$ :	1	2	3	4	5	6	7	8	9	10	11
$x_i$ :	.580	.969	.655	.246	.954	.107	.241	.030	.618	.452	.724
$\pi_{11,i}$ :	6	11	8	4	10	2	3	1	7	5	9
$R(x_i)$ :	1	2	2	1	4	1	2	1	6	5	9

TABELLE 1: Die Daten zu Abbildung 2

Bei der in (1) angegebenen Anordnung der Knoten ergibt sich für Bäume, die mit dem BST-Algorithmus erzeugt werden, eine streng monotone Zuordnung der horizontalen Knotenposition und dem Knoteninhalte. Der Baum hat damit die *Sucheigenschaft*, auf die wir im nächsten Abschnitt zurückkommen werden. Insbesondere kann man die  $n + 1$  externen Knoten, die für das Einsortieren des nächsten Wertes  $x_{n+1}$  zur Verfügung stehen, von links nach rechts anordnen, und sieht so, dass bei dieser Reihenfolge der relative Rang von  $x_{n+1}$  (der natürlich mit dem absoluten Rang in  $(x_1, \dots, x_{n+1})$  übereinstimmt) den umzuwandelnden externen

Knoten festlegt. Denkt man sich in Abbildung 2 den Knoten mit Inhalt 9 gelöscht, so bleiben 11 externe Knoten, und beim Übergang zu  $T_{11}$  wird in der Zählung von links nach rechts der neunte hiervon zu  $T_{10}$  hinzugefügt.

Der BST-Algorithmus liefert also zu  $(x_n)_{n \in \mathbb{N}}$  eine Folge  $(T_n)_{n \in \mathbb{N}}$  von binären Bäumen mit  $|T_n| = n$ , wobei beim Übergang von  $n$  zu  $n + 1$  der durch den relativen Rang von  $x_{n+1}$  bestimmte externe Knoten von  $T_n$  in einen internen Knoten umgewandelt wird. Bei der probabilistischen Analyse geht es um das Verhalten bei zufälligem Input. Konkret nehmen wir an, dass die Daten Realisierungen von unabhängigen und identisch verteilten reellwertigen Zufallsvariablen sind. Um mühsame Fallunterscheidungen beim mehrfachen Auftreten von Werten etc. zu vermeiden, fordern wir zusätzlich, dass die Verteilungsfunktion dieser Zufallsvariablen stetig ist. Da die Bäume, wie bereits erwähnt, bei einer streng monotonen Transformation des Inputs unverändert bleiben, können wir dann ‘o.B.d.A.’ annehmen, dass die den Zufallsvariablen gemeine Verteilung die Gleichverteilung auf dem Einheitsintervall ist; für diese schreiben wir gelegentlich  $\text{unif}(0, 1)$ . Besteht der Input aus einer Folge von unabhängigen,  $\text{unif}(0, 1)$ -verteilten Zufallsvariablen  $(U_k)_{k \in \mathbb{N}}$ , so liefert der BST-Algorithmus eine Folge  $(T_n)_{n \in \mathbb{N}}$  von zufälligen binären Bäumen;  $T_n$  ist der Baum zu  $(U_1, \dots, U_n)$ . Übrigens zeigt bereits der Fall  $n = 3$ , dass  $T_n$  *nicht* gleichverteilt ist auf der Menge der binären Bäume mit  $n$  Knoten.

Der Baum  $T_n$  hängt von  $(U_1, \dots, U_n)$  nur über die nun zufällige Permutation  $\pi_n = (\pi_{n,1}, \dots, \pi_{n,n})$  ab, die man durch die absoluten Ränge erhält. Die Verteilung des Zufallsvektors  $(U_1, \dots, U_n)$  ist invariant unter Permutationen, also ist  $\pi_n$  auf der symmetrischen Gruppe  $S_n$  aller Permutationen von  $\{1, \dots, n\}$  gleichverteilt. Man macht sich leicht klar, dass der Übergang von den absoluten zu den relativen Rängen eine Bijektion zwischen  $S_n$  und der Menge

$$E_n := \{(i_1, \dots, i_n) \in \mathbb{N}^n : 1 \leq i_j \leq j \text{ für } j = 1, \dots, n\}$$

vermittelt, wobei  $E_n$  auch als kartesisches Produkt der Mengen  $\{1, \dots, j\}$ ,  $j = 1, \dots, n$ , geschrieben werden kann. Insbesondere ist  $(R(U_1), \dots, R(U_n))$  auf  $E_n$  gleichverteilt. Dies führt auf den recht bemerkenswerten Sachverhalt, dass unter den angegebenen Bedingungen an  $(U_n)_{n \in \mathbb{N}}$  die relativen Ränge  $(R(U_n))_{n \in \mathbb{N}}$  unabhängig sind, und dass  $R(U_n)$  auf  $\{1, \dots, n\}$  gleichverteilt ist. Für die stochastische Dynamik der Baumfolge  $(T_n)_{n \in \mathbb{N}}$  bedeutet dies, dass man beim Übergang von  $T_n$  zu  $T_{n+1}$  einen der  $n + 1$  externen Knoten von  $T_n$  zufällig und gleichverteilt auswählt und in einen internen Knoten umwandelt; insbesondere ist diese Folge als stochastischer Prozess mit Werten in der Menge der endlichen Teilmengen von  $\{0, 1\}^*$  eine Markovkette.

Die Laufzeit von BST, also die ‘Konstruktionskosten’ für  $T_n$ , wird wesentlich durch die benötigte Anzahl der Vergleiche bestimmt. Nun landet ein Wert genau dann in einem Knoten der Tiefe  $k$ , wenn für seine Einsortierung  $k$  Vergleiche benötigt werden, d.h. die interne Pfadlänge  $X_n$  zu  $T_n$  ist die unter diesem Gesichtspunkt wesentliche Größe. Es sei  $Y_n$  die externe Pfadlänge. Wir wollen nun das asymptotische Verhalten von  $X_n$  bei  $n \rightarrow \infty$  näher untersuchen und erinnern zunächst an einige Grundbegriffe der Stochastik.

Zufallsgrößen werden als messbare Abbildungen auf einem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  mit Werten in einem messbaren Raum  $(E, \mathcal{B})$  betrachtet; bei reellwertigen Zufallsgrößen (Zufallsvariablen)  $X$  ist  $\mathcal{B}$  die  $\sigma$ -Algebra der Borel-Mengen,

bei abzählbarem  $E$  (wie beispielsweise der Menge der endlichen Teilmengen von  $\{0, 1\}^*$ ) ist  $\mathcal{B}$  die Menge aller Teilmengen von  $E$ . Der Raum  $L^2$ , oder ausführlicher  $L^2(\Omega, \mathcal{A}, P)$ , aller Zufallsvariablen mit endlichem zweiten Moment wird mit dem inneren Produkt  $\langle X, Y \rangle := EXY$  zu einem Hilbertraum (man muss, um durch  $\|X\| := \langle X, X \rangle^{1/2}$  eine richtige Norm und nicht nur eine Seminorm zu erhalten, Zufallsvariablen  $X$  und  $Y$  im Falle  $P(X = Y) = 1$  miteinander identifizieren). Für eine Unter- $\sigma$ -Algebra  $\mathcal{F}$  von  $\mathcal{A}$  ist der bedingte Erwartungswert  $X \mapsto E[X|\mathcal{F}]$  die Orthogonalprojektion auf den Teilraum  $L^2(\Omega, \mathcal{F}, P)$  der  $\mathcal{F}$ -messbaren Zufallsvariablen. Im Extremfall  $\{\emptyset, \Omega\}$  erhält man als Projektion auf den Raum der konstanten Zufallsvariablen den bekannten ‘unbedingten’ Erwartungswert, woraus mit der Kompatibilität der Orthogonalprojektionen auf ineinander enthaltene Teilräume  $E(E[X|\mathcal{F}]) = EX$  folgt. In unseren Anwendungen wird  $\mathcal{F}$  von einer diskreten Zufallsgröße  $Y$  erzeugt (die beispielsweise aus dem Vektor  $(T_1, \dots, T_n)$  bestehen kann), eine Zufallsvariable  $X$  ist  $\mathcal{F}$ -messbar, wenn sie als Funktion  $X = \phi(Y)$  von  $Y$  geschrieben werden kann, und der bedingte Erwartungswert hat die Form

$$E[X|\mathcal{F}] = \phi(Y) \quad \text{mit } \phi(y) = \sum_x x P(X = x|Y = y),$$

im Falle  $P(Y = y) = 0$  bleibt  $\phi(y)$  undefiniert.

Beim BST-Algorithmus wird beim Übergang von  $T_n$  zu  $T_{n+1}$  einer der  $n + 1$  externen Knoten von  $T_n$ , mit Tiefen  $k_1(T_n), \dots, k_{n+1}(T_n)$ , zufällig und gleichverteilt ausgewählt und in einen internen Knoten umgewandelt. Dies führt unter Berücksichtigung von  $k_1(T_n) + \dots + k_{n+1}(T_n) = Y_n$  mit  $\mathcal{F}_n = \sigma(T_1, \dots, T_n)$  auf

$$E[Y_{n+1}|\mathcal{F}_n] = Y_n + \frac{1}{n+1} \sum_{i=1}^{n+1} (2(k_i(T_n) + 1) - k_i(T_n)) = \frac{n+2}{n+1} Y_n + 2, \quad (2)$$

wobei die Summe durch eine Zerlegung nach dem Index des zur Umwandlung ausgewählten externen Knotens entsteht. In einer solchen Situation, bei der die Dynamik der bedingten Erwartungswerte durch (deterministische) affine Transformationen ausgedrückt werden kann, freut sich der angewandte Stochastiker: Er hat (fast) ein Martingal gefunden.

Zur Erläuterung dieses in der modernen Stochastik so wichtigen Konzepts verlassen wir kurzzeitig unsere BST-Anwendung und kehren zu der allgemeinen Situation zurück. Eine aufsteigende Folge  $(\mathcal{F}_n)_{n \in \mathbb{N}}$  von Unter- $\sigma$ -Algebren von  $\mathcal{A}$  bezeichnet man als *Filtration*; in unseren Anwendungen ist  $\mathcal{F}_n$  die von den Zufallsvariablen  $U_1, \dots, U_n$  oder den Bäumen  $T_1, \dots, T_n$  erzeugte  $\sigma$ -Algebra und repräsentiert somit die bis zur Zeit  $n$  angesammelte Information. Eine Folge  $(Z_n)_{n \in \mathbb{N}}$  ist zu dieser Filtration *adaptiert*, wenn  $Z_n$   $\mathcal{F}_n$ -messbar ist (also beispielsweise als Funktion von  $T_n$  geschrieben werden kann). Eine solche adaptierte Folge heißt *Martingal*, wenn

$$E[Z_{n+1}|\mathcal{F}_n] = Z_n \quad \text{für alle } n \in \mathbb{N}$$

gilt. Dies ist ein spezieller, aber sehr wichtiger Typ stochastischer Dynamik: Bezüglich des mittleren quadratischen Fehlers  $E(Z_{n+1} - \tilde{Z})^2 = \|Z_{n+1} - \tilde{Z}\|^2$  ist, auf der Basis der Historie bis zur Zeit  $n$ , der aktuelle Wert  $Z_n = \tilde{Z}$  des Prozesses die optimale Vorhersage für den nächsten Wert  $Z_{n+1}$ .

Wir kehren kurz zurück zur BST-Situation. Aus (2) folgt

$$EY_{n+1} = \frac{n+2}{n+1} EY_n + 2$$

und damit unter Verwendung von  $X_n = Y_n - 2n$

$$E[Z_{n+1}|\mathcal{F}_n] = Z_n \quad \text{für } Z_n := \frac{X_n - EX_n}{n+1},$$

d.h.  $(Z_n, \mathcal{F}_n)_{n \in \mathbb{N}}$  ist ein Martingal.

Wir skizzieren den Nachweis, dass  $L^2$ -beschränkte Martingale  $(Z_n, \mathcal{F}_n)_{n \in \mathbb{N}}$  im Hilbertraum  $L^2$  konvergieren: Es seien hierzu  $m, n \in \mathbb{N}$  mit  $m \leq n$ . Die Martingaleigenschaft impliziert  $E[Z_n|\mathcal{F}_m] = Z_m$ . Ganz allgemein hat man für bedingte Erwartungswerte die Regel, dass ein  $\mathcal{F}$ -messbarer Faktor  $X$  'herausgenommen' werden kann, also  $E[XY|\mathcal{F}] = XE[Y|\mathcal{F}]$  gilt. Dies führt auf

$$E(Z_n Z_m) = E(E[Z_n Z_m|\mathcal{F}_m]) = E(Z_m E[Z_n|\mathcal{F}_m]) = EZ_m^2$$

und damit

$$\begin{aligned} \|Z_n - Z_m\|^2 &= E(Z_n - Z_m)^2 \\ &= EZ_n^2 - 2E(Z_n Z_m) + EZ_m^2 \\ &= EZ_n^2 - EZ_m^2 \quad (\geq 0), \end{aligned}$$

bei  $\sup_{n \in \mathbb{N}} EZ_n^2 < \infty$  ist also  $(Z_n)_{n \in \mathbb{N}}$  eine Cauchy-Folge in  $L^2$ .

Man kann zeigen, dass das oben in der BST-Situation gefundene Martingal  $L^2$ -beschränkt ist; siehe hierzu auch den nächsten Abschnitt. Insgesamt erhält man damit den folgenden Satz.

**Satz 1** *Die standardisierte interne Pfadlänge  $Z_n := (X_n - EX_n)/(n+1)$  konvergiert mit  $n \rightarrow \infty$  in  $L^2$  gegen eine Zufallsvariable  $Z$ .*

Natürlich können wir, soweit es den Grenzwertsatz betrifft, auch durch  $n$  anstelle von  $n+1$  dividieren.

## 4 Rekursive Strukturen

Im vorangegangenen Abschnitt haben wir die Dynamik einer Folge  $(T_n)_{n \in \mathbb{N}}$  von Bäumen, also die Frage, wie  $T_{n+1}$  aus  $T_n$  entsteht, in den Vordergrund gestellt. In mancher Hinsicht komplementär hierzu ist die rekursive Sichtweise: Ein binärer Baum ist entweder leer, oder er besteht aus einem Tripel linker Teilbaum, Wurzelknoten, rechter Teilbaum (wobei die Teilbäume natürlich wiederum leer sein können etc.). Diese Sichtweise ist insbesondere dann sinnvoll, wenn zwischen den einzelnen Bäumen keine direkte Verbindung besteht. Dies wiederum ist bei dem Algorithmus der Fall, den wir nun beschreiben und zu binären Bäumen in Beziehung setzen.

Eine vorgegebene Liste  $x = (x_1, \dots, x_n)$  von reellen Zahlen ist in aufsteigende Reihenfolge zu bringen; wir nehmen wieder Einfachheit halber an, dass keine mehrfachen Werte auftreten. Der Algorithmus QUICKSORT verfährt wie folgt: Zunächst konstruiert man durch sukzessiven Vergleich von  $x_1$  mit  $x_i$ ,  $i = 2, \dots, n$ , die



Teillisten  $x^- = (x_1^-, \dots, x_k^-)$  und  $x^+ = (x_1^+, \dots, x_j^+)$  der kleineren bzw. größeren Werte. In der sortierten Liste gehört  $x_1$  zwischen diese beiden Listen. Man macht nun rekursiv mit den beiden Teillisten weiter, bis man bei Listen der Länge 0 oder 1 angekommen ist. Schreibt man das erste Element der jeweiligen Liste in den Wurzelknoten und ordnet  $x^-$  und  $x^+$  dem linken bzw. rechten Teilbaum zu, so ergibt sich ein binärer Baum im Sinne von Abschnitt 2; konkret führt  $x = (6, 11, 8, 4, 10, 2, 3, 1, 7, 5, 9)$  auf den Baum in Abbildung 2. Eine randomisierte Version dieses Verfahrens erhält man, wenn anstelle des jeweils ersten Elementes der Liste ein Element mit zufälligem, auf den möglichen Positionen  $\{1, \dots, n\}$  (bei einer Liste der Länge  $n$ ) gleichverteilten Index gewählt wird. Die Anzahl  $X_n$  der benötigten Vergleiche bei unabhängigem,  $\text{unif}(0, 1)$ -verteilten Input ist ein gutes Maß für die benötigte Laufzeit und somit ein zentrales Objekt bei der probabilistischen Analyse dieses Verfahrens.

Kaum ein Algorithmus wurde so intensiv untersucht wie QUICKSORT. Auf einer Beliebtheitskala der Anwendungsbeispiele in einer Stochastikvorlesung für Informatiker dürfte die zugehörige average-case Analyse, also die Betrachtung von  $EX_n$ , Platz 1 erhalten: Der absolute Rang von  $U_1$  unter  $U_1, \dots, U_n$  ist gleichverteilt auf der Menge  $\{1, \dots, n\}$ , man hat also mit Wahrscheinlichkeit  $1/n$  die Länge  $k - 1$  für  $x^-$  und  $n - k$  für  $x^+$ . Da  $U_2, \dots, U_n$  bei der Rechts-Links-Entscheidung mit  $U_1$  verglichen werden müssen, führt dies auf

$$EX_n = n - 1 + \frac{1}{n} \sum_{k=1}^n (EX_{k-1} + EX_{n-k}).$$

Diese einfache Differenzgleichung für die Zahlenfolge  $(EX_n)_{n \in \mathbb{N}_0}$  lässt sich leicht lösen. Man erhält unter Berücksichtigung der Randbedingung  $EX_0 = 0$  die explizite Formel  $EX_n = 2(n+1)H_n - 4n$  mit  $H_n := \sum_{k=1}^n 1/k$ ; man nennt  $H_n$  die  $n$ -te harmonische Zahl. Eine Betrachtung der Varianz in diesem Stil führt nach einer erheblich mühsameren Rechnung auf

$$\text{var}(X_n) = 7n^2 - 4(n+1)H_n^{(2)} - 2(n+1)H_n + 13n, \quad (3)$$

wobei nun  $H_n^{(2)} := \sum_{k=1}^n 1/k^2$  die harmonischen Zahlen zweiter Art bezeichnen. Aus diesen Resultaten erhält man mit der Ungleichung von Chebyshev und der bekannten Asymptotik der harmonischen Zahlen bereits, dass  $X_n/(2n \log n)$  mit  $n \rightarrow \infty$  in Wahrscheinlichkeit gegen 1 geht. Da  $X_n$  in der Baum-Interpretation, wie die Notation nahelegt, zu der in Abschnitt 3 behandelten Pfadlänge wird, liefern diese ‘Momentbetrachtungen’ auch den Nachweis, dass das dort als Hilfsmittel verwendete Martingal  $(Z_n, \mathcal{F}_n)_{n \in \mathbb{N}}$  in der Tat  $L^2$ -beschränkt ist.

Satz 1 impliziert nun, dass  $(X_n - 2n \log n)/(n+1)$  mit  $n \rightarrow \infty$  in Verteilung konvergiert, gibt aber, was die Grenzverteilung betrifft, wenig an Information her. Wir wollen nun zeigen, dass die rekursive Struktur auf eine Gleichung für diese Grenzverteilung führt, aus der solche Information erhalten werden kann; es stehen also nun Verteilungen im Vordergrund und nicht die Zufallsgrößen selbst. Wir schreiben ganz allgemein  $\mathcal{L}(Z)$  für die Verteilung einer Zufallsgröße  $Z$ ; der Buchstabe ‘ $\mathcal{L}$ ’ steht dabei für das englische Wort ‘law’, eine der im internationalen Schrifttum üblichen Bezeichnungen für ‘Verteilung(sgesetz)’.

Zu einer Permutation  $\pi = (\pi_1 \dots, \pi_n)$  seien  $\pi^-$  und  $\pi^+$  die Teilpermutationen, die man erhält, wenn man alle Werte  $\geq \pi_1$  bzw.  $\leq \pi_1$  streicht und im letzteren Fall noch  $\pi_1$  von den übriggebliebenen Einträgen subtrahiert. Konkret ergeben sich beispielsweise bei  $\pi = (6, 11, 8, 4, 10, 2, 3, 1, 7, 5, 9)$  die Teilpermutationen  $\pi^- = (4, 2, 3, 1, 5)$  und  $\pi^+ = (5, 2, 4, 1, 3)$ . Ist  $\pi$  zufällig und gleichverteilt über  $S_n$ , so zeigt eine einfache Rechnung, dass  $\pi^-$  und  $\pi^+$  unter  $\pi_1$  bedingt unabhängig sind und dass  $\pi_-$  unter  $\pi_1 = m$  gleichverteilt ist auf  $S_{m-1}$ ,  $\pi^+$  auf  $S_{n-m}$ . Bereits verwendet haben wir, dass  $\pi_1$  auf der Menge  $\{1, \dots, n\}$  gleichverteilt ist. Da  $T_n$  als Funktion von  $\pi_n$  betrachtet werden kann, ergibt sich so das folgende induktive Verfahren zur Konstruktion der Verteilungen  $\mathcal{L}(T_n)$ ,  $n \in \mathbb{N}_0$ :  $\mathcal{L}(T_0)$  ist auf den leeren Baum konzentriert. Hat man  $\mathcal{L}(T_0), \dots, \mathcal{L}(T_{n-1})$  konstruiert, so erhält man  $\mathcal{L}(T_n)$  als Verteilung des Baumes mit linkem Teilbaum  $T^-$  und rechtem Teilbaum  $T^+$ , indem man zunächst eine auf  $\{1, \dots, n\}$  gleichverteilte Zufallsvariable  $M$  und dann bei  $M = m$  die Teilbäume gemäß  $\mathcal{L}(T_{m-1})$  bzw.  $\mathcal{L}(T_{n-m})$  wählt. Im Rahmen der in Abschnitt 2 gegebenen formalen Definition bezeichnen wir dabei als *linken Teilbaum* eines binären Baumes  $T$  die Menge aller Knoten  $u = (u_1, \dots, u_k)$  mit  $(0, u_1, \dots, u_k) \in T$ , beim *rechten Teilbaum* geht es stattdessen um  $(1, u_1, \dots, u_k)$ ; entsprechend kann man für die Prozedur des Zusammenfügens eine formale Definition geben.

Die rekursive Struktur überträgt sich natürlich auf Funktionen von  $T_n$ , wie beispielsweise die interne Pfadlänge. Für diese erhalten wir

$$X_n \stackrel{= \text{distr}}{=} n - 1 + X_{I_n - 1} + X'_{n - I_n}.$$

Hierbei sind die drei Zufallsgrößen  $I_n, (X_0, \dots, X_{n-1})$  und  $(X'_0, \dots, X'_{n-1})$  stochastisch unabhängig,  $I_n$  ist auf  $\{1, \dots, n\}$  gleichverteilt und  $X_i, X'_i$  haben dieselbe Verteilung (nämlich die der Pfadlänge von  $T_i$ ), und ‘ $\stackrel{= \text{distr}}{=}$ ’ steht für Gleichheit in Verteilung. Wie in Abschnitt 3 führen wir auch hier eine affine Transformation aus und erhalten

$$\begin{aligned} \frac{X_n - \phi(n)}{n + 1} &\stackrel{= \text{distr}}{=} \frac{n - 1 + \phi(I_n - 1) + \phi(n - I_n) - \phi(n)}{n + 1} \\ &+ \frac{X_{I_n - 1} - \phi(I_n - 1)}{I_n} \cdot \frac{I_n}{n + 1} \\ &+ \frac{X_{n - I_n} - \phi(n - I_n)}{n - I_n + 1} \cdot \frac{n - I_n + 1}{n + 1}. \end{aligned} \quad (4)$$

Wir verwenden dies mit  $\phi(n) := EX_n = 2(n + 1)H_n - 4n$  und lassen  $n$  gegen  $\infty$  gehen: Auf der linken Seite erhalten wir (die Verteilung von)  $Z$  aus Satz 1. Mit  $\lim_{n \rightarrow \infty} \phi(n)/(2n \log n) = 1$  ergibt sich auf der rechten Seite von (4)

$$\frac{n - 1 + \phi(I_n - 1) + \phi(n - I_n) - \phi(n)}{n + 1} \rightarrow_{\text{distr}} 1 + 2U \log(U) + 2(1 - U) \log(1 - U)$$

und natürlich auch

$$\frac{I_n}{n + 1} \rightarrow_{\text{distr}} U, \quad \frac{n - I_n + 1}{n + 1} \rightarrow_{\text{distr}} 1 - U$$

mit  $\mathcal{L}(U) = \text{unif}(0, 1)$ ; hierbei steht ' $\rightarrow_{\text{distr}}$ ' für Konvergenz in Verteilung. Da  $I_n$  auf  $\{1, \dots, n\}$  gleichverteilt und von  $(X_1, \dots, X_n)$  unabhängig ist, folgt weiter

$$E\left(\frac{X_{I_n-1} - \phi(I_n - 1)}{I_n} - Z\right)^2 = \frac{1}{n} \sum_{j=0}^{n-1} E\left(\frac{X_j - EX_j}{j+1} - Z\right)^2.$$

Satz 1 zusammen mit der bekannten Aussage zu Cesàro-Summen liefert hierfür den Grenzwert 0. Analog verfährt man bei  $(X_{n-I_n} - \phi(n - I_n))/(n - I_n + 1)$  und erhält so insgesamt für die rechte Seite von (4) den Grenzwert  $C(U) + UZ + (1 - U)Z'$ , wobei  $U$ ,  $Z$  und  $Z'$  unabhängig sind, sowohl  $Z$  als auch  $Z'$  dieselbe Verteilung haben wie die Zufallsvariable  $Z$  auf der linken Seite und schließlich  $C$  definiert wird durch

$$C(u) := 1 + 2u \log(u) + 2(1 - u) \log(1 - u) \quad \text{für } 0 < u < 1. \quad (5)$$

Dem aufmerksamen Leser wird nicht entgangen sein, dass in der obigen Argumentation etwas sehr freizügig zwischen Zufallsvariablen und ihren Verteilungen hin und her gesprungen wurde. Wir geben nun das formale Gerüst, mit dessen Hilfe dann auch der entscheidende Aspekt dieses Zugangs herausgearbeitet werden kann. Wir fassen hierzu Wahrscheinlichkeitsmaße als Punkte in einem metrischen Raum auf und betrachten konkret die Menge  $\mathcal{M}_2$  aller Wahrscheinlichkeitsmaße  $\mu$  auf den Borelschen Teilmengen von  $\mathbb{R}$ , zu denen das zweite Moment existiert, also  $\int x^2 \mu(dx) < \infty$  gilt. Weiter sei  $\mathcal{M}_{2,0}$  die Teilmenge der  $\mu \in \mathcal{M}_2$  mit  $\int x \mu(dx) = 0$ , für die also der zugehörige Erwartungswert verschwindet. Für  $\mu, \nu$  in  $\mathcal{M}_2$  wird die *Wasserstein-Distanz* definiert durch

$$d_W(\mu, \nu) := \inf\{\|X - Y\| : \mathcal{L}(X) = \mu, \mathcal{L}(Y) = \nu\}.$$

Der metrische Raum  $(\mathcal{M}_2, d_W)$  ist vollständig,  $\mathcal{M}_{2,0}$  ist hierin abgeschlossen. Darüber hinaus ist bekannt, dass Konvergenz bzgl.  $d_W$  von  $(\mu_n)_{n \in \mathbb{N}}$  gegen  $\mu$  äquivalent ist zur schwachen Konvergenz zusammen mit der Konvergenz der zweiten Momente. Mit der schwachen Konvergenz ist hierbei  $\int f d\mu_n \rightarrow \int f d\mu$  für alle stetigen und beschränkten Funktionen gemeint; dies entspricht der Konvergenz in Verteilung von  $X_n$  gegen  $X$  im Falle  $\mu_n = \mathcal{L}(X_n)$ ,  $\mu = \mathcal{L}(X)$ . Die  $d_W$ -Konvergenz einer Folge von Wahrscheinlichkeitsmaßen kann also beispielsweise durch Angabe einer Folge von Zufallsvariablen nachgewiesen werden, die in  $L^2$  (und damit in Verteilung) konvergiert, und für die auch die zweiten Momente konvergieren. Genau dies führt nun in Verbindung mit dem obigen Argument auf eine Gleichung für die Grenzverteilung  $\mu$  der standardisierten Pfadlänge:

$$Z \stackrel{\text{distr}}{=} C(U) + UZ + (1 - U)Z' \quad (6)$$

mit  $U, Z, Z'$  unabhängig,  $\mathcal{L}(U) = \text{unif}(0, 1)$ ,  $\mathcal{L}(Z) = \mathcal{L}(Z') = \mu$  und  $C$  wie in (5).

Wir definieren nun einen Operator  $S$  auf  $\mathcal{M}_2$  durch

$$S(\mu) := \mathcal{L}(C(U) + UZ + (1 - U)Z'),$$

wobei die Zufallsvariablen  $U, Z, Z'$  unabhängig sind und  $\mathcal{L}(U) = \text{unif}(0, 1)$ ,  $\mathcal{L}(Z) = \mathcal{L}(Z') = \mu$  gilt. Für die Zufallsgröße  $C(U)$  erhalten wir

$$\begin{aligned} EC(U) &= \int_0^1 (1 + 2u \log(u) + 2(1-u) \log(1-u)) du \\ &= 0, \end{aligned} \tag{7}$$

$$\begin{aligned} EC(U)^2 &= \int_0^1 (1 + 2u \log(u) + 2(1-u) \log(1-u)) du \\ &= \frac{7}{3} - \frac{2\pi^2}{9} < \infty. \end{aligned} \tag{8}$$

Aus (8) folgt, dass das zweite Moment zu  $S(\mu)$  endlich ist, und (7) zeigt, dass  $\mathcal{M}_{2,0}$  unter  $S$  nicht verlassen wird. Die Verbindung zu unseren binären Bäumen wird durch die Tatsache hergestellt, dass die interessierende Grenzverteilung ein Fixpunkt von  $S$  als Selbstabbildung des Raumes  $\mathcal{M}_{2,0}$  ist.

Ist der Fixpunkt eindeutig? Sicher nicht in  $\mathcal{M}_2$ , da bei einer festen deterministischen Verschiebung der Variablen  $Z$  und  $Z'$  in (6) wegen  $U + (1-U) = 1$  die Verteilungsgleichheit erhalten bleibt. Hat man aber  $\mu, \nu \in \mathcal{M}_{2,0}$ , und ist  $(X, Y)$  ein zweidimensionaler Zufallsvektor mit  $\mathcal{L}(X) = \mu$ ,  $\mathcal{L}(Y) = \nu$  und  $\|X - Y\| \leq d_W(\mathcal{L}(X), \mathcal{L}(Y)) + \epsilon$ , so erhält man unter Zuhilfenahme eines von  $(X, Y)$  unabhängigen Zufallsvektors  $(X', Y')$  mit derselben Verteilung wie  $(X, Y)$  sowie einer von beiden Zufallsvektoren unabhängigen,  $\text{unif}(0, 1)$ -verteilten Zufallsvariablen  $U$

$$\begin{aligned} d_W^2(S(\mu), S(\nu)) &\leq E((C(U) + UX + (1-U)X') - (C(U) + UY + (1-U)Y'))^2 \\ &= E(U^2(X - Y)^2) + E((1-U)^2(X' - Y')^2) \\ &= (EU^2 + E(1-U)^2)E(X - Y)^2 \\ &\leq \frac{2}{3} (d_W(\mu, \nu) + \epsilon)^2. \end{aligned}$$

Hierbei haben wir im ersten Schritt ausgenutzt, dass  $C(U) + UX + (1-U)X'$  die Verteilung  $S(\mu)$  und  $C(U) + UY + (1-U)Y'$  die Verteilung  $S(\nu)$  hat, im zweiten, dass aufgrund der vorausgesetzten Unabhängigkeiten und wegen  $E(X - Y) = 0$  die gemischten Terme verschwinden. Da  $\epsilon > 0$  beliebig war, erhalten wir hieraus die folgende Abschätzung.

**Satz 2** Für alle  $\mu, \nu \in \mathcal{M}_{2,0}$  gilt  $d_W(S(\mu), S(\nu)) \leq \sqrt{2/3} d_W(\mu, \nu)$ .

Mit dem bekannten Fixpunktsatz von Banach folgt nun, dass die Verteilungsgleichheit (6) die Verteilung  $\mathcal{L}(Z)$  in  $\mathcal{M}_{2,0}$  charakterisiert, aber beispielsweise auch, dass der  $d_W$ -Abstand der durch Iteration von  $S$  erhaltenen Wahrscheinlichkeitsmaße  $\mu, S(\mu), S(S(\mu)), \dots$  zu  $\mathcal{L}(Z)$  bei beliebigem Startmaß  $\mu \in \mathcal{M}_{2,0}$  exponentiell schnell gegen 0 geht. Die Fixpunktgleichung kann auch zum Erhalt von Informationen über die Grenzverteilung verwendet werden. Beispielsweise folgt aus (6) mit der Unabhängigkeit von  $U, Z$  und  $Z'$

$$EZ^2 = EC(U)^2 + EU^2EZ^2 + E(1-U)^2Z^2,$$

da die gemischten Terme wegen  $EZ = EZ' = 0$  verschwinden. Dies ergibt  $EZ^2 = EC(U)^2 + 2EZ^2/3$  und somit wegen (8)

$$\text{var}(Z) = EZ^2 = 3EC(U)^2 = 7 - \frac{2\pi^2}{3},$$

was man natürlich auch als Grenzwert in (3) bei  $n \rightarrow \infty$  erhält. Eine einfache qualitative Schlussfolgerung aus (6) ist die, dass  $\mathcal{L}(Z)$  absolutstetig ist,  $Z$  also eine Wahrscheinlichkeitsdichte hat. In der Tat liefert (6) eine Integralgleichung für eine solche Dichte, die wiederum als Basis für numerische Approximationen dienen kann.

## 5 Poisson-Approximation

Eines der klassischen Resultate der Stochastik ist das *Gesetz der seltenen Ereignisse*, demgemäß die Anzahl  $S_n$  der ‘Erfolge’ bei kleiner Erfolgswahrscheinlichkeit  $p$  und großer Anzahl  $n$  von unabhängigen Versuchswiederholungen näherungsweise Poisson-verteilt ist mit Parameter  $np$ . Über reine Grenzwertsätze und Konvergenzgeschwindigkeitsaussagen hinaus lassen sich bei diesem Gesetz sogar einfache, explizite Oberschranken für den Abstand der Verteilungen erhalten. Wir schreiben im folgenden  $\text{Ber}(p)$  für die Bernoulli-Verteilung mit Parameter  $p$ ,  $\mathcal{L}(X) = \text{Ber}(p)$  bedeutet dann  $P(X = 1) = p = 1 - P(X = 0)$ , und  $\text{Po}(\lambda)$  für die Poisson-Verteilung mit Parameter  $\lambda$ . Der *Totalvariationsabstand* zweier Wahrscheinlichkeitsmaße  $\mu$  und  $\nu$  auf  $\mathbb{R}$  wird definiert durch

$$d_{\text{TV}}(\mu, \nu) := \sup_{B \in \mathcal{B}} |\mu(B) - \nu(B)|. \quad (9)$$

Sind  $\mu$  und  $\nu$  auf eine abzählbare Menge  $A$  konzentriert, so gilt

$$d_{\text{TV}}(\mu, \nu) = \frac{1}{2} \sum_{x \in A} |\mu(\{x\}) - \nu(\{x\})|.$$

Eine für unsere Anwendungen interessante Oberschranke ist, mit  $S_n = X_1 + \dots + X_n$ ,  $X_1, \dots, X_n$  unabhängig und  $\mathcal{L}(X_i) = \text{Ber}(p_i)$  für  $i = 1, \dots, n$ ,

$$d_{\text{TV}}(\mathcal{L}(S_n), \text{Po}(ES_n)) \leq \frac{\sum_{i=1}^n p_i^2}{\sum_{i=1}^n p_i}. \quad (10)$$

Wir wollen dies nun in der BST-Situation aus Abschnitt 3 anwenden, behalten aber die in Abschnitt 4 beschriebene Verbindung zu QUICKSORT im Auge. Es sei also  $(T_n)_{n \in \mathbb{N}}$  die Folge der mit dem BST-Algorithmus erzeugten zufälligen binären Bäume zu einer Folge  $(U_n)_{n \in \mathbb{N}}$  von unabhängigen,  $\text{unif}(0, 1)$  verteilten Zufallsvariablen. Jede 0-1-Folge  $\theta = (\theta_k)_{k \in \mathbb{N}} \in \{0, 1\}^{\mathbb{N}}$  definiert einen Pfad durch die  $T_n$ 's, es sei

$$Z_n(\theta) := \max\{k \in \mathbb{N}_0 : (\theta_1, \dots, \theta_k) \in T_n\}$$

die Länge dieses Pfades zur ‘Zeit’  $n$ . Wie in Abschnitt 2 erklärt, wird beim Übergang von  $T_k$  zu  $T_{k+1}$  einer der  $k + 1$  externen Knoten von  $T_k$  zufällig und gleichverteilt

(unabhängig von der Historie  $T_1, \dots, T_k$ ) ausgewählt und in einen internen Knoten umgewandelt. Dies zeigt, dass sich  $Z_n(\theta)$  als Summe von unabhängigen,  $\text{Ber}(1/i)$ -verteilten Zufallsvariablen schreiben lässt, also ist (10) anwendbar und man erhält

$$d_{\text{TV}}(\mathcal{L}(Z_n(\theta)), \text{Po}(H_n)) \leq \frac{H_n^{(2)}}{H_n}. \quad (11)$$

Die Oberschranke kann wiederum durch  $\pi^2/(6 \log n)$  nach oben abgeschätzt werden, geht also insbesondere mit  $n \rightarrow \infty$  gegen 0. Beachtet man, dass sich die  $d_{\text{TV}}$ -Distanz der Verteilungen von  $X$  und  $Y$  unter affinen Transformationen der Zufallsvariablen nicht ändert, dass wegen (9) diese Distanz eine Oberschranke für den Abstand der Verteilungsfunktionen zu  $X$  und  $Y$  liefert und somit Konvergenz in Totalvariation die Verteilungskonvergenz impliziert, und dass schließlich  $(X_\lambda - \lambda)/\sqrt{\lambda}$  mit  $\mathcal{L}(X_\lambda) = \text{Po}(\lambda)$  bei  $\lambda \rightarrow \infty$  in Verteilung gegen die Standardnormalverteilung konvergiert, so folgt aus (11), dass  $Z_n(\theta)$  nach Standardisierung asymptotisch normalverteilt ist.

Konkret liefert dieses Resultat bei  $\theta = (0, 0, \dots)$  die Verteilungsasymptotik zur Einfügetiefe des Minimums. Größen dieser Art sind insbesondere dann interessant, wenn in einer Such- oder Sortieranwendung nicht der gesamte Baum konstruiert wird (oder werden muss). Das Standardbeispiel ist der Algorithmus FIND, der bei Input  $x = (x_1, \dots, x_n) \in \mathbb{R}^n$  und  $l \in \{1, \dots, n\}$  das Element  $x_{n:l}$  mit Rang  $l$  findet, also den Wert  $x_i$ , für den  $|\{1 \leq j \leq n : x_j \leq x_i\}| = l$  gilt; bei  $l = 1$  ist dies das Minimum. Wie bei QUICKSORT werden im ersten Schritt durch Vergleich mit  $x_1$  die Teillisten  $x^-$  und  $x^+$  der kleineren bzw. größeren Elemente gebildet. Ist  $m$  die Länge von  $x^-$ , so macht man im Falle  $m > l$  mit  $x^-$  anstelle von  $x$  weiter,  $l$  bleibt unverändert. Bei  $m = l$  ist  $x_1$  die gesuchte Ordnungsstatistik. Bei  $m < l$  macht man mit  $x^+$  weiter, muss dabei aber natürlich  $l$  durch  $l - m$  ersetzen. Wie bei QUICKSORT gibt es auch hier eine randomisierte Version, bei der der Index des Vergleichselements zufällig ausgewählt wird.

Wie viele Rekursionen benötigt man, um mit FIND die  $l$ -te Ordnungsstatistik zu finden oder, äquivalent, welche Tiefe  $Y_{n,l}$  hat der Knoten mit Inhalt  $l$  in dem mit dem BST-Algorithmus konstruierten Baum zu einer zufälligen und gleichverteilten Permutation von  $\{1, \dots, n\}$ ? Ein klassisches Resultat ist

$$EY_{n,l} = H_l + H_{n+1-l} - 2. \quad (12)$$

Die folgende Beobachtung erweist sich als Schlüssel für die Analyse der Verteilung von  $Y_{n,l}$ : Auf dem Pfad vom Wurzelknoten zu dem Knoten mit Inhalt  $l$  geht man in einem Knoten mit Inhalt  $k$  nach links, wenn  $k$  ein absteigender Rekord in den größeren und nach rechts, wenn  $k$  ein aufsteigender Rekord unter den kleineren Werten ist. Beispielsweise geht man auf dem Weg vom Wurzelknoten zu dem Knoten mit Inhalt 9 in Abbildung 2 bei 6 und 8 nach rechts, bei 11 und 10 nach links. Die zugehörigen Daten in Tabelle 1 zeigen, dass diese Werte unter den Werten kleiner als 9 die aufsteigenden ( $\pi_{11,1}$  und  $\pi_{11,3}$ ) bzw. absteigenden Rekorde sind ( $\pi_{11,2}$  und  $\pi_{11,5}$ ); bei nicht zuletzt einsortierten Werten  $l \neq 9$  ist nur der Bereich links von  $l$  relevant. (Natürlich wird bei einer Anwendung von FIND der Datensatz  $x$  selbst und nicht die zufällige Permutation betrachtet. Bei letzterer hat die  $l$ -te Ordnungsstatistik  $x_{n:l}$  stets den Wert  $l$ , mit ‘Inhalt 9’ ist also in Wirklichkeit ‘Inhalt 0.724’ gemeint.)

Wir formalisieren diese Beobachtung: Es seien  $N := \pi^{-1}(l)$  die Position des gesuchten Wertes  $l$  in der zufälligen Permutation  $\pi$ ,

$$\begin{aligned} S_- &:= \{1 \leq i < N : \pi(i) < l\} = \{i_1, \dots, i_G\}, \\ S_+ &:= \{1 \leq i < N : \pi(i) > l\} = \{j_1, \dots, j_{N-1-G}\} \end{aligned}$$

mit  $i_1 < \dots < i_G$  und  $j_1 < \dots < j_{N-1-G}$  die Indizes zu den relevanten Teilpermutationen und

$$\begin{aligned} \pi_- &:= (\pi(i_1), \dots, \pi(i_G)), \quad \pi_+ := (\pi(j_1), \dots, \pi(j_{N-1-G})), \\ R_- &:= \sum_{r=1}^G \prod_{k=1}^{r-1} 1_{\{\pi(i_k) < \pi(i_r)\}}, \quad R_+ := \sum_{r=1}^{N-1-G} \prod_{k=1}^{r-1} 1_{\{\pi(i_k) > \pi(i_r)\}} \end{aligned}$$

( $1_A$  bezeichnet die Indikatorfunktion zur Menge  $A$ , also  $1_A(x) = 1$  bei  $x \in A$  und  $1_A(x) = 0$  sonst). Mit diesen Definitionen gilt  $Y_{n,l} = R_- + R_+$  und man erhält, dass  $\mathcal{L}(Y_{n,l})$  aus bestimmten handelsüblichen Verteilungen zusammengesetzt ist:  $N$  ist gleichverteilt auf  $\{1, \dots, n\}$ . Unter  $N = m$  ist  $(\pi_1, \dots, \pi_{m-1})$  eine zufällige und gleichverteilte Permutation von  $\{\pi_i : 1 \leq i < m\}$ . Der Wert  $\pi_i$  kann als Resultat des  $i$ -ten Zugs ohne Zurücklegen aus einer Urne mit  $n - 1$  Kugeln angesehen werden, wobei  $l - 1$  Kugeln ‘weiß’, also kleiner als  $l$  sind, d.h.  $G = |S_-|$  hat eine hypergeometrische Verteilung mit Parametern  $n - 1$ ,  $l - 1$  und  $m - 1$ . Schließlich hat unter  $G = k$  die Anzahl der aufsteigenden Rekorde (das sind ja gerade die Elemente mit maximalem relativen Rang, siehe Abschnitt 3) dieselbe Verteilung wie die Summe  $\xi_1 + \dots + \xi_k$  von unabhängigen Zufallsvariablen  $\xi_i$  mit  $\mathcal{L}(\xi_i) = \text{Ber}(1/i)$ ,  $1 \leq i \leq k$ . Bei diesem letzten Schritt kann nun (10) angewendet werden; analog behandelt man  $R_+$ . Eine technische Schwierigkeit ergibt sich noch aus der Tatsache, dass  $R_-$  und  $R_+$  nicht unabhängig sind, aber schließlich erhält man eine Approximation der interessierenden Verteilung  $\mathcal{L}(Y_{n,l})$  durch eine Poissonverteilung, in Form einer einfachen und expliziten Oberschranke für den Approximationsfehler in der Totalvariationsdistanz.

### Satz 3

$$\sup_{l \in \{1, \dots, n\}} d_{\text{TV}}(\mathcal{L}(Y_{n,l}), \text{Po}(EY_{n,l})) \leq \frac{28 + \pi^2}{\log n} \quad \text{für alle } n \geq 2.$$

Wie bei der Einfügetiefe des Minimums kann man auch hier als Korollar die asymptotische Normalität erhalten, beispielsweise bei  $l = l_n = \lfloor n/2 \rfloor$  für den (linken) Median.

## 6 Begleitende Konstruktionen

Wie bereits seine Bezeichnung andeutet, nimmt der *Zentrale Grenzwertsatz* eine besondere Position in der Stochastik ein: Sind  $X_1, X_2, \dots$  unabhängige und identisch verteilte Zufallsvariablen mit endlichem zweiten Moment und positiver Varianz, so konvergieren die standardisierten Partialsummen  $(S_n - ES_n)/\sqrt{n}$ ,  $S_n := \sum_{i=1}^n X_i$ , mit  $n \rightarrow \infty$  in Verteilung gegen eine Normalverteilung. Zu diesem Satz gibt es neben zahlreichen wichtigen Verallgemeinerungen auch eine Vielfalt ganz unterschiedlicher Beweise. (Übrigens kann man auch hier, beispielsweise über eine Aufspaltung

der Partialsummen nach geraden und ungeraden Summationsindizes, eine rekursive Struktur finden und auf diesem Wege eine Fixpunktgleichung für die Grenzverteilung herleiten, die diese charakterisiert.) Eines dieser Beweisverfahren beruht auf einer Art ‘begleitenden Konstruktion’: Bei normalverteilten  $X$ -Variablen ist die Aussage des Zentralen Grenzwertsatzes trivial (die Verteilung der standardisierten Partialsummen hängt nicht von  $n$  ab). Die Verbindung zwischen der Ausgangssituation und der Situation mit normalverteilten Summanden kann dann beispielsweise dadurch hergestellt werden, dass in den Partialsummen nacheinander die  $X$ -Variablen durch ein normalverteiltes Pendant ersetzt werden.

Es sei nun  $X_{n,l}$  die Anzahl der von  $\text{FIND}(\pi_n, l)$  bei einer zufälligen und gleichverteilten Permutation  $\pi_n$  der Menge  $\{1, \dots, n\}$  und einem  $l \in \{1, \dots, n\}$  benötigten Vergleiche; diese Anzahl kann wieder als Anhaltspunkt für die benötigte Laufzeit dienen. Wir wollen nun unter Verwendung einer geeigneten begleitenden Konstruktion das asymptotische Verhalten der Verteilung von  $X_{n,l}$  untersuchen.

Als Funktional des binären Baumes  $T_n$  zu  $\pi_n$  lässt sich  $X_{n,l}$  als Summe der Knotenzahlen der jeweiligen rechten oder linken Teilbäume zu der Ahnenreihe des  $l$  enthaltenden Knotens schreiben. Im Falle  $l = 1$ , also beim Auffinden des Minimums, macht man in den Rekursionsschritten stets mit dem linken Teilbaum und dem zweiten Argument  $l = 1$  weiter. Als Konsequenz hat die Folge der Knotenzahlen eine recht einfache stochastische Struktur: Die Knotenzahl  $M$  des linken Teilbaumes des Wurzelknotens ist gleichverteilt auf  $\{0, \dots, n-1\}$ . Aus der Diskussion der rekursiven Struktur ist bekannt, dass der linke Teilbaum unter  $M = m$  dieselbe Verteilung hat wie  $T_m$ . Damit erhalten wir entlang des durch die Wegbeschreibung  $\theta = (0, 0, \dots)$  gegebenen Pfades durch  $T_n$  die Knotenzahlen  $Y_1, Y_2, \dots$ , wobei  $Y_1$  auf  $\{0, \dots, n-1\}$  gleichverteilt ist und für alle  $k \in \mathbb{N}$  die Zufallsvariable  $Y_{k+1}$  unter  $Y_k = m$  bei  $m > 0$  auf  $\{0, \dots, m-1\}$  gleichverteilt ist. Die Gesamtzahl  $X_{n,1}$  ergibt sich als Summe aller dieser Größen, von denen höchstens  $n-1$  von 0 verschieden sein können; die  $n-1$  Vergleiche, die auf den ersten linken Teilbaum führen, müssen noch hinzugerechnet werden. (Man beachte, dass  $n$  nur hier und über die erste Variable  $Y_1$  eingeht.)

Ist  $U$  eine  $\text{unif}(0, 1)$ -verteilte Zufallsvariable, so ist  $\lfloor kU \rfloor$  auf  $\{0, \dots, k-1\}$  gleichverteilt. Als begleitende Konstruktion bietet sich daher die Folge  $(\prod_{i=1}^k U_i)_{k \in \mathbb{N}}$  der Partialprodukte zu einer Folge  $(U_k)_{k \in \mathbb{N}}$  von unabhängigen,  $\text{unif}(0, 1)$ -verteilten Zufallsvariablen an. Konkret hat mit solchen  $U$ -Variablen die Folge  $(Y_k)_{k \in \mathbb{N}}$  dieselbe Verteilung wie die durch  $\tilde{Y}_1 := \lfloor nU_1 \rfloor$ ,  $Y_{k+1} := \lfloor U_{k+1} \tilde{Y}_k \rfloor$  für alle  $k \in \mathbb{N}$  induktiv definierte Folge  $(\tilde{Y}_k)_{k \in \mathbb{N}}$ . Insbesondere gilt

$$\frac{1}{n} X_{n,1} \stackrel{\text{distr}}{=} \frac{1}{n}(n-1) + \frac{1}{n} \lfloor nU_1 \rfloor + \frac{1}{n} \lfloor U_2 \lfloor nU_1 \rfloor \rfloor + \frac{1}{n} \lfloor U_3 \lfloor U_2 \lfloor nU_1 \rfloor \rfloor \rfloor + \dots$$

Mit  $n \rightarrow \infty$  konvergiert die rechte Seite fast sicher gegen

$$1 + U_1 + U_2 \cdot U_1 + U_3 \cdot U_2 \cdot U_1 + \dots$$

Wir erhalten also, wenn wir im Interesse einer kompakten Notation das leere Produkt als 1 interpretieren,

$$\frac{1}{n} X_{n,1} \xrightarrow{\text{distr}} \sum_{k=0}^{\infty} \prod_{i=1}^k U_i \quad \text{mit } n \rightarrow \infty. \quad (13)$$



Mit einer analogen Argumentation (oder einer einfachen Symmetrieüberlegung) erhält man diesen Grenzwert auch für das Maximum, also bei  $l = n$ .

Wie bereits bei der im letzten Abschnitt behandelten Rekursionstiefe besteht allerdings auch hier die eigentliche Schwierigkeit des Problems darin, dass es neben  $n$  noch einen zweiten Parameter  $l$  gibt, der sich in den Rekursionsschritten ändern kann (wenn nicht gerade das Minimum oder Maximum gesucht wird, die ja sowieso beide sehr einfach mit einem einzigen Durchgang durch die Daten gefunden werden können). Beispielsweise erhält man in unserem bewährten Beispiel  $\pi = (6, 11, 8, 4, 10, 2, 3, 1, 7, 5, 9)$  bei  $l = 9$  für FIND in den Rekursionsschritten nacheinander die Argumente

$$\begin{array}{ll} (6, 11, 8, 4, 10, 2, 3, 1, 7, 5, 9), & 9, \\ (11, 8, 10, 7, 9), & 3, \\ (8, 10, 7, 9), & 4, \\ (10, 9), & 1, \\ (9), & 1. \end{array}$$

Eine Möglichkeit für die Behandlung eines allgemeinen zweiten Arguments  $l$  ergibt sich aus der Beobachtung, dass die bei den Rekursionsschritten aufeinander folgenden Paare  $(n_i, l_i)$ , wobei jeweils  $n_i$  die Länge des Inputs und  $l_i$  den Rang vor dem  $i$ -ten solchen Schritt bezeichnen,  $i = 1, 2, \dots$ , eine Markovkette bilden. In jedem Schritt tritt nun einer von drei Fällen ein: Man erhält das gewünschte Element (diesen Fall formalisieren wir durch den Eintritt in einen künstlichen Zustand  $(0, 0)$ ), man macht mit der Liste der kleineren Werte weiter, oder man macht mit der Liste der größeren Werte weiter. (In einer praktischen Implementierung dieses Verfahrens würde man natürlich bereits bei einer Inputliste der Länge 1 stoppen; der Unterschied ist aber für die asymptotische Analyse bedeutungslos.) Bei der Markovkette wird also, ausgehend von einem Zustand  $(n, l)$ , der nächste Zustand zufällig und gleichverteilt aus den  $n$  Möglichkeiten  $(0, 0)$  (erster Fall),  $(l+1, l), (l+2, l), \dots, (n-1, l)$  (zweiter Fall) und  $(n-1, l-1), (n-2, l-2), \dots, (n-l+1, 1)$  (dritter Fall) ausgewählt;  $(0, 0)$  ist ein absorbierender Zustand. Im obigen Beispiel startet die Markovkette in  $(11, 9)$  und besucht nacheinander die Zustände  $(5, 3)$ ,  $(4, 4)$ ,  $(2, 1)$  und  $(1, 1)$ , bevor sie schließlich, in unserer formalen Konstruktion, einen letzten Schritt ausführt und im Zustand  $(0, 0)$  ankommt (und bleibt).

Es sei nun  $(Y_k)_{k \in \mathbb{N}_0}$  eine Markovkette mit Zustandsraum  $E = \{(y_1, y_2) \in \mathbb{R}^2 : 0 \leq y_2 \leq y_1\}$  und Übergangsmechanismus  $Y_{k+1} = \phi(Y_k, U_{k+1})$ , wobei  $(U_k)_{k \in \mathbb{N}}$  wie im Falle  $l = 1$  eine Folge von unabhängigen,  $\text{unif}(0, 1)$ -verteilten Zufallsvariablen bezeichnet und  $\phi : (0, 1) \times E \rightarrow E$  definiert wird durch

$$\phi(u, (y_1, y_2)) := \begin{cases} (y_1 - uy_1, y_2 - uy_1), & \text{falls } uy_1 < y_2, \\ (uy_1, y_2), & \text{falls } uy_1 \geq y_2. \end{cases}$$

Mit dieser begleitenden Konstruktion erhalten wir das gewünschte Resultat zur Verteilungasymptotik von  $X_{n,l}$ . Es sei  $Y_{k,1}$  die erste Komponente von  $Y_k$ ,  $k \in \mathbb{N}_0$ .

**Satz 4** *Ist  $(l_n)_{n \in \mathbb{N}}$  eine Folge natürlicher Zahlen mit*

$$\lim_{n \rightarrow \infty} \frac{l_n}{n} = t \quad \text{und} \quad 1 \leq l_n \leq n \quad \text{für alle } n \in \mathbb{N},$$

so gilt

$$\frac{1}{n}X_{n,l_n} \xrightarrow{\text{distr}} \sum_{k=0}^{\infty} Y_{k,1}, \quad (14)$$

wobei  $(Y_k)_{k \in \mathbb{N}_0}$  eine Markovkette ist mit dem oben angegebenen Übergangsmechanismus und Start in  $(1, t)$ , also  $Y_0 = (1, t)$ .

Es sei  $F(t, \cdot)$  die Verteilungsfunktion zu dem Grenzwert in (14). Die Markoveigenschaft von  $(Y_k)_{k \in \mathbb{N}_0}$  führt auf die Integralgleichung

$$F(t, x) = \int_0^t F\left(\frac{t-u}{1-u}, \frac{(x-1)^+}{1-u}\right) du + \int_t^1 F\left(\frac{t}{u}, \frac{(x-1)^+}{u}\right) du,$$

wobei wir  $x^+$  für das Maximum von  $x$  und 0 geschrieben haben. Mit numerischen Mitteln lassen sich hieraus  $\alpha$ -Quantile der Grenzverteilung erhalten; beispielsweise ergeben sich bei  $\alpha = 0.99$  und  $t = 0, 1/4, 1/2$  die Näherungswerte 4.11, 5.81 bzw. 6.16. Für den Erwartungswert des Grenzwertes, also  $m_1(t) := \int x F(t, dx)$ , erhält man eine Integralgleichung, die sich explizit lösen lässt:

$$m_1(t) = 2 - 2t \log t - 2(1-t) \log(1-t). \quad (15)$$

Wie bei der Varianz der internen Pfadlänge in Abschnitt 4 kann man auch hier den Erwartungswert zum Grenzwert alternativ über die klassische explizite Formel

$$EX_{n,l} = 2(n+3 + (n+1)H_n - (l+2)H_l - (n+3-l)H_{n+1-l}) \quad (16)$$

bestimmen. Grenzwertaussagen dienen oft als Ersatz für die eigentlich interessierende Aussage im endlichen Fall, aber gelegentlich werden durch den Übergang zum Grenzwert die wesentlichen Aspekte (selbst des endlichen Falls) deutlicher. In diesem Sinn könnte man die (formal schwächere) Aussage (15) durchaus für informativer halten als (16).

Bereits im einführenden Abschnitt haben wir erwähnt, dass mit Aussagen zur Verteilungasymptotik die Wahrscheinlichkeit langer Laufzeiten (genauer: großer Werte von  $X_{n,l}$ ) näherungsweise bestimmt werden kann. Eine Obergrenze für solche Wahrscheinlichkeiten ergibt sich natürlich bereits mit dem Erwartungswert, denn für alle nichtnegativen Zufallsvariablen  $Z$  und alle  $z > 0$  gilt  $P(Z > z) \leq EZ/z$ . Bei  $t = 1/2$  erhält man so unter Verwendung von (15) als Obergrenze für das 0.99-Quantil der Grenzverteilung den Wert 338.63, der die über Satz 4 erhaltene Näherung 6.16 doch erheblich übersteigt.

## 7 Bemerkungen zur Literatur

Zur Analyse von Algorithmen ist nach wie vor das mehrbändige Werk ‘The Art of Computer Programming’ von D. E. Knuth die erste Referenz. Für uns ist vor allem Band III, Knuth (1973, 1998), interessant. Ein sehr schönes neueres Buch zu diesem Themenkreis ist Sedgewick und Flajolet (1996). Speziell mit Bäumen im Zusammenhang mit den klassischen Aufgaben des Suchens und Sortierens befassen sich die beiden Monographien von Mahmoud (1992, 2000). Ein viel benutzter

Text für ‘Undergraduates’ ist das Buch von Cormen, Leiserson und Rivest (1997); dort findet man beispielsweise in Section 13.4 die oben in Abschnitt 5 verwendete Darstellung von  $Y_{n,l}$  über auf- und absteigende Rekorde.

Bei den hier betrachteten diskreten Strukturen erhält man Zufälligkeit einerseits über eine Annahme an die Form des Inputs für den Algorithmus (das Datenmodell), andererseits über eine etwaige Randomisierung, die in den Algorithmus eingebaut wird. Beide können durchaus, wie bei QUICKSORT, auf ein und dasselbe mathematische Problem führen. Eine interessante Arbeit, bei der im Zusammenhang mit binären Suchbäumen beide Aspekte eine Rolle spielen, ist Seidel und Aragon (1996).

Wir haben uns auf Aussagen zur Verteilungskonvergenz konzentriert, aber es gibt natürlich auch eine Reihe von interessanten Aussagen zur fast sicheren Konvergenz und zur Konvergenz in Wahrscheinlichkeit für zufällige binäre Bäume. Besonders schön ist das Resultat von Devroye (1986) über das asymptotische Verhalten der Höhe der in der vorliegenden Arbeit betrachteten Bäume. Methodisch ergibt sich hierdurch ein sehr fruchtbarer Zusammenhang zu Verzweigungsprozessen, der in einem Übersichtsartikel von Devroye (1998) ausführlich dargestellt wird.

Bekanntlich lässt sich Verteilungskonvergenz oft recht elegant mit charakteristischen Funktionen beweisen. Für manche der obigen stochastischen Methoden gibt es Alternativen, die kombinatorische Überlegungen mit klassischen Hilfsmitteln der Funktionentheorie kombinieren; in vielen Fällen führt beispielsweise die auf Flajolet zurückgehende Singularitätenanalyse sogar auf asymptotische Entwicklungen. Das bereits erwähnte Buch von Sedgewick und Flajolet (1996) bietet einen exzellenten Einstieg in diese Methodik; Flajolet und Sedgewick (2005) wird sehnsüchtig erwartet.

Grundlage für Abschnitt 3 ist Régnier (1989). Wir haben einen Martingalkonvergenzsatz benutzt, der einen sehr kurzen Beweis hat und auf die  $L^2$ -Konvergenz führt, die Methode liefert aber auch die fast sichere Konvergenz. Generell sind Martingale als Hilfsmittel nützlich, weil sie nicht nur unter einfachen Bedingungen konvergieren, sondern auch eine große Zahl nützlicher Ungleichungen erfüllen und sich unter Stopzeiten vorteilhaft verhalten. Ein sehr schönes einführendes Buch zur Wahrscheinlichkeitstheorie, das explizit auf Martingale setzt, ist Williams (1991). Zu Abschnitt 4 ist Rösler (1991) die zentrale Referenz, dort wird mit dem Kontraktionsargument sogar die Konvergenz bewiesen. Die mittlerweile so benannte ‘Kontraktionsmethode’ ist seither bei vielen anderen Algorithmen erfolgreich angewendet worden; siehe beispielsweise Rösler und Rüschemdorf (1998) und Neininger und Rüschemdorf (2004). Einer der Vorteile dieser Methode ist, dass sie über Grenzwertresultate hinaus auch Konvergenzgeschwindigkeitsaussagen und sogar Abschätzungen für die Distanz zwischen der eigentlich interessierenden Verteilung und der Grenzverteilung liefern kann; siehe Neininger und Rüschemdorf (2002) im Zusammenhang mit QUICKSORT. Die Formel (3) für die Varianz ist Theorem 3.12 in Sedgewick und Flajolet (1996). Abstandsbegriffe für Wahrscheinlichkeitsmaße sind oben an verschiedenen Stellen aufgetaucht, das Buch von Rachev (1991) ist eine allgemeine Referenz zu solchen ‘probability metrics’. Die in Abschnitt 5 skizzierte Anwendung der Verbindung zur Theorie der Rekorde ist in Grüberl und Stefanoski (2005) ausgeführt, die genannte Verbindung selbst geht auf Devroye (1988) zurück. Dobrow und Smythe (1996) verwenden die Poisson-Approximation, um asymptotische Normalität für die Tiefe eines zufällig ausgewählten Knotens zu

erhalten. Das average-case Resultat (12) stammt aus Arora und Dent (1985). Das Buch von Barbour et al. (1992) ist ein Standardwerk zur Poisson-Approximation. Die Resultate zu FIND in Abschnitt 6 sind in Grübel (1998) ausgeführt; eine begleitende Konstruktion für die Asymptotik der stochastischen Prozesse  $(X_{n, \lceil nt \rceil})_{0 < t \leq 1}$  erscheint bereits in Grübel und Rösler (1996). Die Dichte der Grenzverteilung in (13) taucht (teilweise unter dem Namen Dickman-Funktion) in ganz unterschiedlichen Zusammenhängen auf, von der Zahlentheorie bis zur Finanzmathematik. Die verschiedenen in dieser Arbeit vorgestellten Methoden lassen sich durchaus kombinieren; so werden beispielsweise in Neininger und Rüschemdorf (2004) begleitende Konstruktionen in Verbindung mit der Kontraktionsmethode benutzt.

Zum Abschluss sei noch einmal betont, dass hier nur ein kleines Segment eines sehr aktiven Gebiets betrachtet wurde (und dabei die Interessen des Autors überproportional vertreten sind). Zumindest erwähnt werden sollte noch, dass *gleichverteilte* zufällige binäre Bäume ihre eigene Theorie haben, mit faszinierenden Verbindungen zur Brownschen Bewegung und Anwendungen bis hin zu bestimmten nichtlinearen partiellen Differentialgleichungen — aber dies ist eine andere Geschichte.

**Danksagung.** Ich danke beiden Gutachtern für die außerordentlich sorgfältige Durchsicht der Arbeit und für ihre sehr hilfreichen Verbesserungsvorschläge.

## Literatur

- Aigner, M. und Ziegler, G. (1998) *Proofs from THE BOOK*. Springer, Berlin.
- Alon, N. und Spencer, J.H. (2000) *The Probabilistic Method*. Second edition. Wiley, New York.
- Arora, S. and Dent, W. (1985) Randomized binary search technique. *Communications of the ACM* **12**, 77–80.
- Barbour, A.D., Holst, L. und Janson, S. (1992) *Poisson Approximation*. Clarendon Press, Oxford.
- Cormen, Th.H., Leiserson, Ch.E. and Rivest, R.L. (1997) *Introduction to Algorithms*. MIT Press, Cambridge (USA).
- Devroye, L. (1986) A note on the height of binary search trees. *Journal of the ACM* **33**, 489–498.
- Devroye, L. (1988) Applications of the theory of records in the study of random trees. *Acta Informatica* **26**, 123–130.
- Devroye, L. (1998) Branching processes and their applications in the analysis of tree structures and tree algorithms. In: *Probabilistic Methods for Algorithmic Discrete Mathematics*. Habib, M., McDiarmid, C., Ramirez-Alfonsin, J. and Reed, B. (ed.), Springer, Berlin.
- Dobrow, R.P. and Smythe, R.T. (1996) Poisson approximations for functionals of random trees. *Random Structures and Algorithms* **9**, 79–92.
- Flajolet, Ph. und Sedgewick, R. (2005) *Analytic Combinatorics*. Buch in Vorbereitung; Vorabversion erhältlich unter:  
<http://algo.inria.fr/flajolet/Publications/book050805.pdf>

- Grübel, R. (1998) Hoare's selection algorithm: a Markov chain approach. *Journal of Applied Probability* **35** (1998), 36–45.
- Grübel, R. und Rösler, U. (1996) Asymptotic distribution theory for Hoare's selection algorithm. *Adv. in Applied Prob.* **28**, 252–269.
- Grübel, R. und Stefanoski, N. (2005) Mixed Poisson approximation of node depth distributions in random binary search trees. *Ann. Appl. Probab.* **15**, 279–297.
- Janson, S., Łuczak T. und Ruciński, A. (2000) *Random Graphs*. Wiley, New York.
- Knuth, D.E. (1973, 1998) *The Art of Computer Programming* Vol. 3, Sorting and Searching. Second Edition. Addison-Wesley, Reading.
- Mahmoud, H.M. (1992) *Evolution of Random Search Trees*. Wiley, New York.
- Mahmoud, H.M. (2000) *Sorting. A Distribution Theory*. Wiley, New York.
- Motwani, R. und Raghavan, P. (1995) *Randomized Algorithms*. Cambridge University Press, Cambridge.
- Neininger, R. und Rüschemdorf, L. (2002) Rates of convergence for Quicksort. Analysis of algorithms. *J. Algorithms* **44**, 51–62.
- Neininger, R. und Rüschemdorf, L. (2004) A general limit theorem for recursive algorithms and combinatorial structures. *Ann. Appl. Probab.* **14**, 378–418.
- Rachev, S.T. (1991) *Probability Metrics and the Stability of Stochastic Models*. Wiley, New York.
- Régnier, M. (1989) A limiting distribution for quicksort. *RAIRO Inform. Théor. Appl.* **23**, 335–343.
- Rösler, U. (1991) A limit theorem for "Quicksort". *RAIRO Inform. Théor. Appl.* **25**, 85–100.
- Rösler, U. und Rüschemdorf, L. (1998) The contraction method for recursive algorithms. Average-case analysis of algorithms (Princeton, NJ, 1998). *Algorithmica* **29**, 3–33.
- Sedgewick, R. und Flajolet, Ph. (1996) *An Introduction to the Analysis of Algorithms*. Addison-Wesley, Reading.
- Seidel, R. und Aragon, C.R. (1996) Randomized search trees. *Algorithmica* **16**, 464–497.
- Williams, D. (1991) *Probability with Martingales*. Cambridge University Press, Cambridge.