Übungsblatt 1

Aufgabe 1. Es sei (T, v) ein (n, q)-Wurzelbaum mit $q \geq 2$. Zeigen Sie, dass

$$L(T) \geq \lceil \log_q(n) \rceil.$$

Beweis. Wegen $q \geq 2$ gilt offensichtlich $n \geq 2$, d.h. insbesondere, dass die Kardinalität der Kantenmenge mindestens zwei ist.
Für $l = L(T)$ und wegen $q \geq 2$ gilt genau dann $\log_q(n) \leq l$, wenn $n \leq q^l$. Zunächst betrachten wir die Menge aller (n, q)−Wurzelbäume, ohne die anfängliche Einschränkung $q \geq 2$. Wir beweisen letztere Ungleichung per vollständige Induktion nach $l \in \mathbb{N}_0$. Ist $l = 0$, so besteht T nur aus der Wurzel, und es gilt $n = 1$ und $q = 0$. Letzteres impliziert mit der üblichen Konvention $0^0 = 1$ die Ungleichung $n = 1 \leq 1 = q^l$.
Seien nun $l \geq 1$ und $x_1, \ldots, x_t$ die Nachfolger der Wurzel v mit $1 \leq t \leq q$ und bezeichne mit T_i die echten Unterbräume verwurzelt in x_i für $1 \leq i \leq t$. Dies sind die Zusammenhangskomponenten des vollen Teilgraphens der Kantenmenge $E(T) \setminus \{v\}$. Offensichtlich gilt $l_i := L(T_i) < L(T) = l$ für alle $1 \leq i \leq t$ und somit nach Induktionsvoraussetzung $n_i \leq q_i^{l_i}$, wobei (n_i, q_i) der Parameter des Wurzelbaumes (T_i, x_i) für $1 \leq i \leq t$ ist. Wegen $q_i \leq q$, $l_i < l$ und $t \leq q$ erhalten wir

$$n = \sum_{i=1}^{t} n_i \leq \sum_{i=1}^{t} q_i^{l_i} \leq \sum_{i=1}^{t} q^l \leq q^{l+1}.$$

Nun gilt nach anfänglicher Bemerkung für $q \geq 2$ die Behauptung. □

Bemerkung. *Alternativ, kann man auch Satz (2.2), die Kraftsche Ungleichung, in Aufgabe 1 hinzuziehen. Nehme an, dass $n > q^l$ gilt. Seien $l_1, \ldots, l_n$ die Blattlängen, so gilt für den (n, q)-Wurzelbaum*

$$\sum_{i=1}^{n} q^{-l_i} \geq \sum_{i=1}^{n} q^{-\max_{x \in E(T)} l(x)} = \sum_{i=1}^{n} q^{-L(T)} = \frac{n}{q^{L(T)}} = \frac{n}{q^l} > 1,$$

ein Widerspruch zur Kraftschen Ungleichung.
Eine weitere Möglichkeit ist, den Baum T in einen vollständigen Wurzelbaum mit derselben Wurzel einzubetten und für diesen gilt trivialerweise die Ungleichung $n \leq q^l$, sodass sie insbesondere auch für T erfüllt ist.

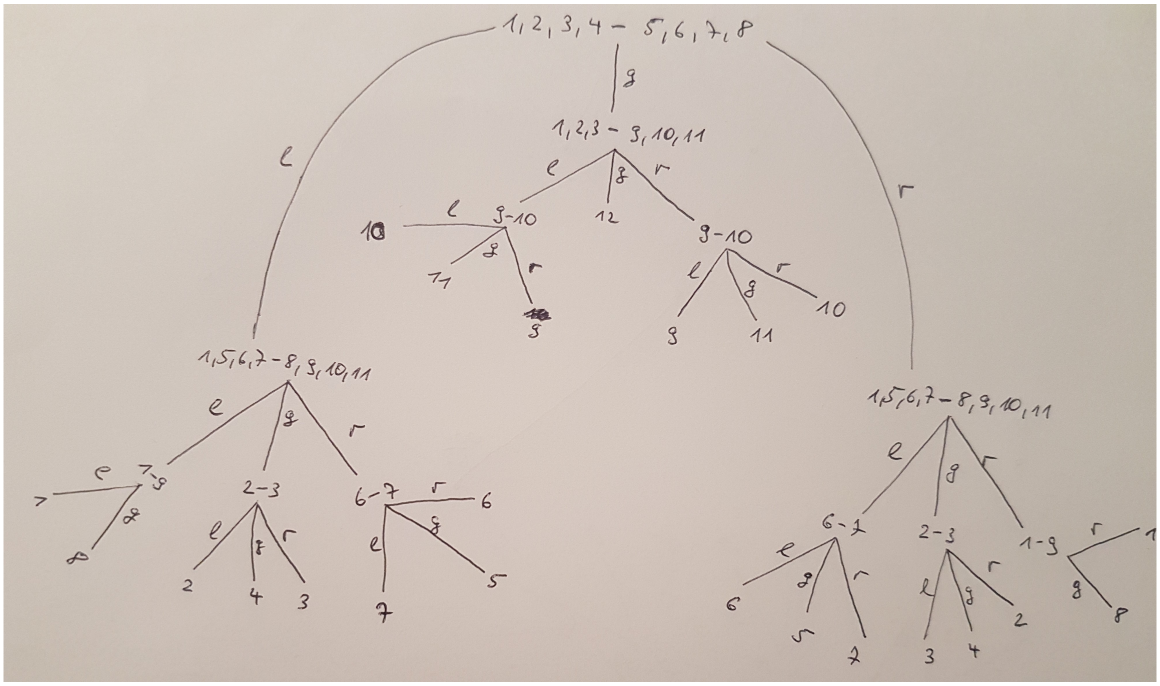
Aufgabe 2. (a) Gegeben seien zwölf Münzen und eine Balkenwaage. Eine der Münzen ist gefälscht und daher entweder leichter oder schwerer als die Übrigen. Wie hoch ist Mindestanzahl von Wägungen, die man im ungünstigsten Fall durchführen muss, um die gefälschte Münze zu finden?

(b) Wie viele Wägungen benötigt man, wenn 13 Münzen gegeben sind?

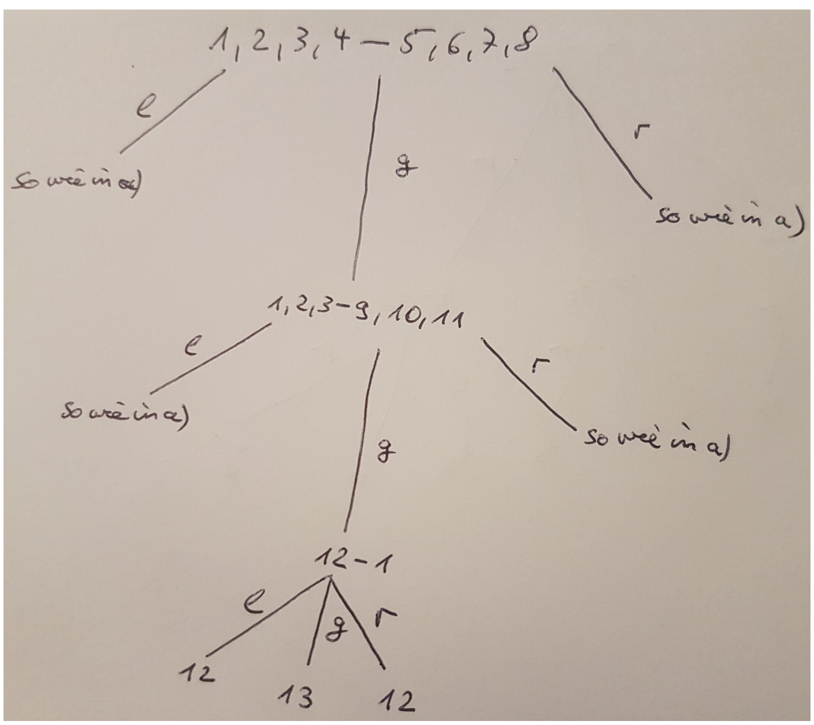
Beweis. (a) Zunächst überlegen wir uns, dass wir mindestens drei Wägungen benötigen, um die gefälschte Münze zu finden. Sei hierfür T ein $(m, 3)$-Wurzelbaum, der ein Verfahren zur Findung der gefälschten Münze mit der minimalen Anzahl von Wägungen beschreibt. So gilt $m \geq 12$ und wegen Aufgabe 1 ist

$$L(T) \geq \lceil \log_3(m) \rceil \geq \lceil \log_3(12) \rceil = 3$$

die Mindestanzahl der Wägungen. Nun geben wir ein Verfahren an, das die gefälschte Münze mit Hilfe von drei Wägungen findet, wobei l, g, r die jeweiligen Zustände der Waage sind.



(b) Seien 13 Münzen gegeben. So folgt aus denselben Gründen wie zuvor, dass man mindestens drei Wägungen benötigt, um die gefälschte Münze zu finden. Im Folgenden wird ein Verfahren beschrieben, welches mit Hilfe von maximal drei Wägungen die gesuchte Münze findet.



□

Aufgabe 3. Geben Sie eine Bijektion zwischen der Menge der Präfix-Codes C über dem Alphabet $\{0,1\}$ mit $|C| = n$ und der Menge der mit $\{0,1\}$-gewichteten $(n,2)$-Wurzelbäume an.

Beweis. In der ursprünglichen Aufgabenstellung wurde nach einer Bijektion zwischen der Menge der Präfix-Codes $\mathcal{C}$ und der $(n,2)$-Wurzelbäume gefragt. Wir widmen uns zunächst diesem Problem.

Sei $\mathcal{A} = \{0,1\}$ das Alphabet und $\mathcal{C}$ die Menge der Präfix-Codes über $\mathcal{A}$ mit n Wörtern. Da $\mathcal{C}$ offensichtlich nicht endlich ist und eine Teilmenge der abzählbar unendlichen Menge $\mathcal{A}^*$ ist, ist $\mathcal{C}$ auch abzählbar unendlich. Analog gilt, dass die Menge der $(n,2)$-Wurzelbäume, wir nennen diese $\Gamma(n,2)$, abzählbar unendlich ist. Somit sind $\mathcal{C}$ und $\Gamma(n,2)$ gleichmächtig.

Gehen wir nun zu der modifizierten Aufgabenstellung über. Wir definieren im Folgenden eine Abbildung $\Gamma(n,2)_G \longrightarrow \mathcal{C}$, wobei $\Gamma(n,2)_G$ die Menge der gewichteten $(n,2)$-Bäume ist. Ein gewichteter Baum ist das Paar (T,g) bestehend aus dem Baum T und einer Abbildung $g : K(T) \longrightarrow \mathcal{A}$.
Sei hierfür $(T,g) \in \Gamma(n,2)_G$ ein gewichteter Wurzelbaum. Jeder maximale gewichtete Pfad, d.h. der Pfad von der Wurzel zu einem Blatt, definiert ein eindeutiges Wort über dem Alphabet $\mathcal{A}$, wie folgt:

$$\bullet \xrightarrow{\varepsilon_1} \bullet \xrightarrow{\varepsilon_2} \bullet \quad \dots \quad \bullet \xrightarrow{\varepsilon_{l_i}} \bullet \quad \rightsquigarrow \quad (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{l_i}),$$

wobei $\varepsilon_j \in \{0,1\}$ für alle $1 \le j \le l_i$, $1 \le i \le n$. Sei $C = \{c_1, \dots, c_n\}$ die Menge der zu (T,g) gehörigen Wörter, die wie oben beschrieben konstruiert wurden. So gilt wegen der Anzahl n der Blätter, dass keines der c_i ein Teilwort der c_j für $1 \le i \ne j \le n$ ist. Die Menge C ist also ein Präfix-Code. Somit ist die Zuordnung wohldefiniert.
Die Umkehrabbildung definiert man analog. Sei $C = \{c_1, \dots, c_n\} \in \mathcal{C}$. So konstruiere auf offensichtlicher Weise einen Baum mit n-Blättern und die dazugehörigen gewichteten n-Pfade mit Hilfe der Wörter aus C. Diese Abbildung ist wohldefiniert, da C ein Präfix-Code ist.
Die obigen Abbildungen sind nach Konstruktion invers zueinander. □

Aufgabe 4. (a) Gibt es einen Präfix-Code über $\{0,1\}$ mit sechs Codewörtern der Länge $1,3,3,3,3,3$? Existiert ein Präfix-Code mit Wörtern der Länge $2,3,3,3,3,3$?

(b) Betrachte die Menge $\{n \in \mathbb{N}_0 \mid n \le 99\}$. Die übliche dezimale Repräsentation ist kein Präfix-Code. Bestimmen Sie einen optimalen binären Quellencode C unter der Annahme, dass alle Zahlen gleichwahrscheinlich sind. Wie groß ist $\overline{L}(C)$?

Nach Aufgabe 3 können wir jedem Präfix-Code mit n Wörtern einem gewichteten $(n,2)$-Wurzelbaum T zuordnen. Seien $l_1, \dots, l_n$ die Blattlängen von T, so gilt nach Satz (2.2) (Kraftsche Ungleichung)

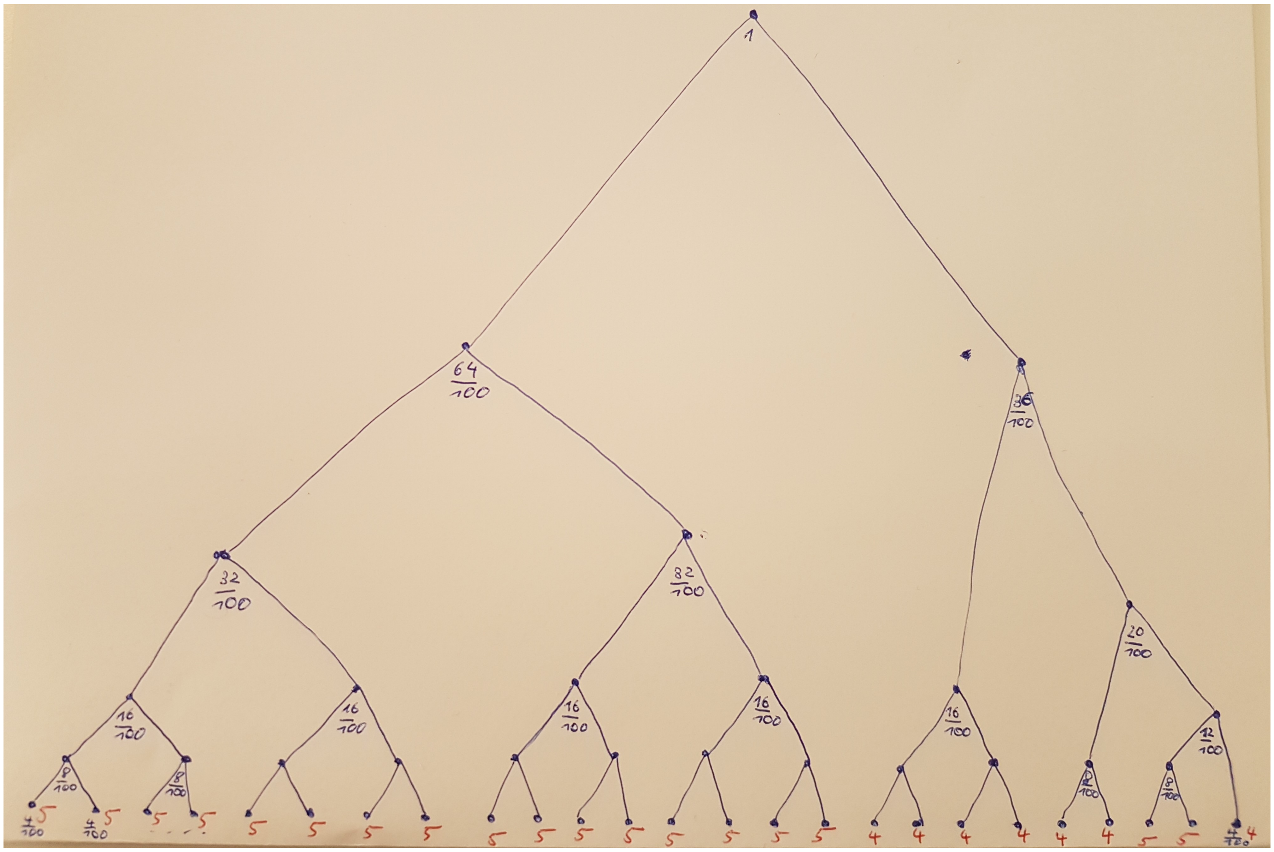
$$\sum_{i=1}^{n} 2^{-l_i} \le 1.$$

Angenommen es existiert ein Präfix-Code C mit Wortlängen $1,3,3,3,3,3$. Ordne diesem den Graphen T zu (und vergiss das Gewicht). Die Kraftsche Ungleichung liefert

$$2^{-1} + \sum_{i=1}^{5} 2^{-3} = \frac{1}{2} + \frac{5}{8} > 1,$$

einen Widerspruch. Also existiert kein solcher Präfix-Code.

Wegen $2^{-2} + \sum_{i=1}^{5} 2^{-3} = \frac{1}{4} + \frac{5}{8} \le 1$ existiert nach Satz (2.2) ein $(6,2)$-Code mit den geforderten Längen $2,3,3,3,3,3$. Ein Beispiel eines solchen Codes ist $C = \{00, 010, 011, 100, 101, 111\}$.
Nach Satz (2.6) aus der Vorlesung liefert eine Huffman Kodierung einen optimalen binären Quellencode. Betrache zunächst die Menge $\{n \in \mathbb{N}_0 \mid n \le 24\}$ und die Gleichverteilung P auf dem dazugehörigem Mengensystem definiert durch $P(i) = \frac{4}{100}$ für $0 \le i \le 24$. Eine Huffman Kodierung liefert den folgenden Wurzelbaum.



Erweitert man nun den obigen Baum symmetrisch, so dass er 100 Blätter besitzt, so erhält man einen von der Huffman Kodierung induzierten Wurzelbaum, der zur ursprünglichen Menge (Quelle) und darauf definierten Gleichverteilung gehört. Dieser enthält $18 \cdot 4 = 72$ Pfade der Länge 7 und $7 \cdot 4 = 28$ Pfade der Länge 6. Der gesuchte Wert beträgt also $\overline{L} = \frac{72 \cdot 7 + 28 \cdot 6}{100} = 6,72$.