Datenkompression (Rote), SS 2003, 7. Übungsblatt
Abgabe: bis Montag, 16. 6. 2003

28. Verzögerung bei der arithmetischen Kodierung. (10 Punkte)
Wieviele Eingabesymbole muss man bei der arithmetischen Kodierung im schlimmsten Fall lesen, bevor man das erste Bit der kodierten Nachricht ausgeben kann? Wieviele Bits muss man bei der Dekodierung im schlimmsten Fall lesen, bevor man das erste Zeichen der Nachricht ausgeben kann?

    - (5 Punkte) Nehmen Sie an, dass mit exakter Arithmetik gerechnet wird.
    - (5 Punkte) Nehmen Sie an, dass die Intervalllängen auf Vielfache von 1/100 gerundet werden.
    - (0 Punkte) Nehmen Sie an, dass die Intervalllängen auf Vielfache von $1/N$ gerundet werden. (Das heißt mit anderen Worten: Es wird ganzzahlige Arithmetik im Bereich $[0, N]$ verwendet.)

    Beweisen Sie die oberen Schranken, und geben Sie jeweils auch Beispiele an, wo die von Ihnen behaupteten Verzögerungen auftreten.

29. Sequentielle Suche bei der arithmetischen Dekodierung. (11 Punkte)
Bei der Dekodierung muss man das Teilintervall finden, das die Kodezahl enthält. Eine einfache Möglichkeit besteht darin, die Intervalle sequentiell zu durchsuchen. Dabei ist es natürlich vorteilhaft, die Intervalle der Länge nach absteigend zu sortieren. Formal sieht das so aus: Gegeben ist eine Folge von Häufigkeiten $m_1 \geq m_2 \geq \cdots \geq m_n$ mit Summe $M = m_1 + \cdots + m_n$. Für eine zufällige Zahl x im Intervall $[L, U)$ soll der Index i bestimmt werden, sodass
$$L + (U - L)\frac{m_1 + \cdots + m_{i-1}}{M} \leq x < L + (U - L)\frac{m_1 + \cdots + m_i}{M}$$
ist, und zwar durch sequentielles Durchprobieren der Werte $i = 1, 2, \ldots, n$.

    - (0 Punkte) Wie viel Zeit benötigt dieses Verfahren, wenn x im i-ten Intervall liegt?
    - (2 Punkte) Schreiben Sie eine Formel für die erwartete Suchzeit.
    - (2 Punkte) Berechnen Sie die erwartete Suchzeit im Fall, dass alle Symbole gleich häufig sind.
    - (2 Punkte) Zeigen Sie, dass dieser Fall der schlechteste mögliche Fall ist.
    - (3 Punkte) Berechnen Sie die erwartete Suchzeit für den Fall, wo m_i annähernd proportional zu $1/i$ ist. (Diese Verteilung entspricht dem sogenannten Zipf'schen Gesetz.)
    - (2 Punkte) Berechnen Sie die erwartete Suchzeit für den Fall, wo m_i annähernd proportional zu $1/i^2$ ist.

30. Programmieraufgabe (10 Punkte)
Schreiben Sie ein Programm zur Dekodierung der mit Ihrem Programm aus Aufgabe 17 komprimierten Daten, und wenden Sie es auf die Beispiele aus Aufgabe 26 an. (Wenn Sie Ihr Kodierungsprogramm ändern müssen, gibt es 5 Punkte Abzug.)

31. Erweitern nicht-vollständiger Präfix-Kodes (0 Punkte)
Ein Kode heißt *vollständig*, wenn jede Kette $x \in T^*$ über dem Kodealphabet T das Anfangsstück einer kodierten Nachricht ist. Beweisen Sie, dass man jeden Präfix-Kode durch Hinzunahme zusätzlicher Kodewörter zu einem vollständigen Präfix-Kode ergänzen kann.