Universität Konstanz
Fachbereich Informatik & Informationswissenschaft
Prof. Dr. Sven Kosub / Michael Aichem, Julian Müller, Dagmar Sorg, Michael Strecke, Nadja Willenborg

Mathematik: Diskrete Strukturen
SS 2015

4. Übungsblatt
(Musterlösung)

Ausgabe: 08.05.2015 **Abgabe:** 15.05.2015, bis spätestens 12:00 per Mail an den Tutor

Vertiefung: **10 Punkte**

(a) Drücken Sie die Anzahl der surjektiven Funktionen $f : \{0,1\}^n \to \{0,1\}^2$ mit Hilfe der Stirling-Zahlen zweiter Art aus.

(b) Wieso gilt $n^m = \sum_{k=0}^{m} S_{m,k} \cdot n^{\underline{k}}$ für $n, m \in \mathbb{N}_+$?

(c) Von 18 Studierenden in einer Spezialvorlesung studieren 7 Mathematik, 9 Physik und 10 Informatik. Davon studieren 3 Mathematik und Physik, 3 Mathematik und Informatik sowie 5 Physik und Informatik. Ein Student studiert sogar all drei Fächer. Wie viele Studierende studieren keines der drei Fächer?

(d) Für drei Mengen A, B und C gelten folgende Eigenschaften: $\|A\| = 63$, $\|B\| = 91$, $\|C\| = 44$, $\|A \cap B\| = 25$, $\|A \cap C\| = 23$, $\|C \cap B\| = 21$. Außerdem gelte $\|A \cup B \cup C\| = 139$. Wie groß ist $\|A \cap B \cap C\|$?

(e) Für zwei Mengen A und B gelte: $\|A\| = 100$, $\|B\| = 60$ und die Anzahl der Elemente von $A \cup B$, die zu genau einer der beiden Mengen gehören, ist genau doppelt so groß, wie die Anzahl der Elemente, die in beiden Mengen liegen. Wie viele Elemente liegen in beiden Mengen?

(f) Wie viele Zahlen im Bereich $1, 2, \ldots, 200$ sind durch keine der Zahlen $3, 7, 11, 27$ teilbar?

(g) Wie viele Zahlen im Bereich $1, \ldots, 10^9$ sind weder von der Form x^3 noch x^7 noch x^{13} für ein geeignetes $x \in \mathbb{N}$?

(h) Auf einer großen Informatikerkonferenz finden immer parallel 5 Vorträge statt. Die dafür vorgesehenen 5 Konferenzräume bieten $150, 200, 200, 300$ und 500 Teilnehmern einen Sitzplatz. Für die Konferenz haben sich 1523 Teilnehmer registriert. Wie viele registrierte Teilnehmer müssen in einem der Konferenzräume auf jeden Fall stehen, wenn alle registrierten Teilnehmer zeitgleich genau einen Vortrag besuchen?

Beachtung: Vortragende besuchen keine Vorträge anderer und zählen auch nicht als registrierte Teilnehmer.

(i) Menschen haben bis zu 150.000 Kopfhaare. Mindestens wie viele Chinesen haben die exakt gleiche Anzahl von Kopfhaaren (zu einem bestimmten Zeitpunkt), wenn Sie von einer chinesischen Bevölkerung von 1,33 Millarden ausgehen?

(j) Wie oft müssen Sie mindestens würfeln, damit eine Zahl n-mal vorkommt?

Lösung:

(a) Eine surjektive Funktion $f : \{0,1\}^n \to \{0,1\}^2$ zerlegt die Menge der Argumente in eine 4-Partition. Da die Nummer der Komponente innerhalb der Partition eine Rolle spielt, ergeben sich somit

$$4! \cdot S_{2^n,4} = 24 \cdot S_{2^n,4}$$

surjektive Funktionen.

(b) Die Zahl n^m ist einerseits die Anzahl der Funktionen $f : \{1,\ldots,m\} \to \{1,\ldots,n\}$. Andererseits hat jede solche Funktion genau $k \in \{1,\ldots,n\}$ verschiedene Funktionswerte. Eine Funktion mit k verschiedenen Funktionswerten zerlegt die Menge $\{1,\ldots,m\}$ in eine k-Partition. Jede Komponente einer k-Partition hat eine andere Nummer aus $\{1,\ldots,n\}$. Somit ergeben sich $S_{m,k} \cdot n^{\underline{k}}$ Funktionen mit genau k verschiedenen Funktionswerten. Summation über k ergibt die Gleichung

$$n^m = \sum_{k=1}^{n} S_{m,k} \cdot n^{\underline{k}}.$$

(c) Bezeichnen M, P sowie I die Mengen der Mathematik-, Physik- und Informatikstudenten, so ergibt sich für die Anzahl der Studenten, die mindestens eines dieser Fächer studieren:

$$\begin{aligned} \|M \cup P \cup I\| &= \|M\| + \|P\| + \|I\| - \|M \cap P\| - \|M \cap I\| - \|P \cap I\| + \|M \cap P \cap I\| \\ &= 7 + 9 + 10 - 3 - 3 - 5 + 1 = 16 \end{aligned}$$

Somit studieren $18 - 16 = 2$ Studenten keines dieser Fächer.

(d) Umstellung gemäß dem Inklusions-Exklusions-Prinzips ergibt

$$\begin{aligned} \|A \cap B \cap C\| &= \|A \cup B \cup C\| - \|A\| - \|B\| - \|C\| + \|A \cap B\| + \|A \cap C\| + \|B \cap C\| \\ &= 139 - 63 - 91 - 44 + 25 + 23 + 21 \\ &= 10 \end{aligned}$$

(e) Es gilt $A \cup B = (A \Delta B) \cup (A \cap B)$ mit $(A \Delta B) \cap (A \cup B) = \emptyset$. Wir erhalten

$$\|A \cap B\| = \|A \cup B\| - \|A \Delta B\| = \|A\| + \|B\| - \|A \cap B\| - 2 \cdot \|A \cap B\| = \|A\| + \|B\| - 3 \cdot \|A \cap B\|$$

Es folgt

$$\|A \cap B\| = \frac{\|A\| + \|B\|}{4} = \frac{160}{4} = 40$$

(f) Es sei A_i die Menge der Vielfachen von i im Bereich $1,\ldots,200$. Da die 3 die 27 teilt, gibt es somit

$$\begin{aligned} \|A_3 \cup A_7 \cup A_11\| &= \|A_3\| + \|A_7\| + \|A_{11}\| - \|A_3 \cap A_7\| - \|A_3 \cap A_{11}\| - \|A_7 \cap A_{11}\| \\ &= \left\lfloor \frac{200}{3} \right\rfloor + \left\lfloor \frac{200}{7} \right\rfloor + \left\lfloor \frac{200}{11} \right\rfloor - \left\lfloor \frac{200}{3 \cdot 7} \right\rfloor - \left\lfloor \frac{200}{3 \cdot 11} \right\rfloor - \left\lfloor \frac{200}{7 \cdot 11} \right\rfloor + \left\lfloor \frac{200}{3 \cdot 7 \cdot 11} \right\rfloor \\ &= 66 + 28 + 18 - 9 - 6 - 2 + 0 \\ &= 95 \end{aligned}$$

Zahlen, die durch $3, 7, 11$ oder 27 teilbar sind. Es gibt also $200 - 95 = 105$ Zahlen im Bereich $1, 2, \ldots, 200$, die durch keine der Zahlen $3, 7, 11, 27$ teilbar sind.

(g) In der Menge $\{1, \ldots, 10^9\}$ gibt es

- 1000 Zahlen der Form x^3 (wegen $(10^3)^3 = 10^9$)
- 19 Zahlen der Form x^7 (wegen $19^7 \leq (10^{9/7})^7 = 10^9$ und $20^7 > 10^9$) und
- 4 Zahlen der Form x^{13} (wegen $4^{11} \leq (10^{9/13})^{13} = 10^9$ und $5^{13} > 10^9$).
- 2 Zahlen der Form $(x^3)^7 = x^{21}$ (wegen $2^{21} \leq (10^{9/21})^{21} = 10^9$ und $3^{21} > 10^9$).
- 1 Zahl der Form $(x^3)^{13} = x^{39}$ (wegen $1^{39} \leq (10^{9/39})^{39} = 10^9$ und $2^{39} > 10^9$).
- 1 Zahl der Form $(x^7)^{13} = x^{91}$ (wegen $1^{91} \leq (10^{9/91})^{91} = 10^9$ und $2^{91} > 10^9$).
- 1 Zahl der Form $((x^3)^7)^{13} = x^{273}$ (wegen $1^{273} \leq (10^{9/273})^{273} = 10^9$ und $2^{273} > 10^9$).

Es gibt nun also

$$10^9 - 1000 - 19 - 4 + 2 + 1 + 1 - 1 = 999\ 998\ 980$$

Zahlen in $\{1, \ldots, 10^9\}$, die nicht von der Form x^3, x^7 oder x^{11} sind.

(h) Es können $500 + 300 + 200 + 200 + 150 = 1350$ in allen Räumen sitzen. Es bleiben somit $1523 - 1350 = 173$ registrierte Teilnehmer übrig, die stehen müssen. Verteilt man diese gleichmäßig auf die Räume ergibt sich

$$\frac{173}{5} = 34,6.$$

Somit müssen mindestens 35 Teilnehmer in einem der Räume stehen.

(i) Es gibt insgesamt 150 001 Möglichkeiten der Anzahl der Haare. Bei Gleichverteilung ergibt sich

$$\frac{1\ 330\ 000\ 000}{150\ 001} \approx 8\ 866,6$$

Somit müssen mindestens 8 867 Chinesen die gleiche Anzahl an Haaren auf dem Kopf haben.

(j) Man muss mindestens n-mal würfeln, da im besten Fall in jedem Wurf die gleiche Zahl gewürfelt wird.

Kreativität: **10 Punkte**

Auf *Oneway Island* gibt es zwischen den Städten nur Einbahnstraßen. Jede Stadt ist mit jeder Stadt durch eine Einbahnstraße verbunden: Für zwei beliebige, unterschiedliche Städte A und B gilt, dass man entweder direkt von A nach B oder direkt von B nach A kommen kann, nicht aber beides direkt.

(a) Zeigen Sie, dass es eine Stadt geben muss, von der man direkt in mindestens die Hälfte aller anderen Städte kommen kann.

(b) Zeigen Sie, dass es eine Stadt geben muss, von der man mit höchstens einem Zwischenstopp zu jeder Stadt kommen kann.

Lösung:

(a) Wir beweisen die Aussage per Widerspruchsbeweis. Seien

- $S =_{\text{def}}$ Menge aller Städte
- $S \times S \supset E =_{\text{def}}$ Menge aller Straßen

Es gilt $\|E\| = \frac{1}{2}\left(\|S\|^2 - \|S\|\right)$, da jede Stadt mit jeder anderen verbunden ist, außer mit sich selbst, jedoch nur in einer Richtung.
Wir können die Menge E wie folgt partitionieren:

$$E_a =_{\text{def}} \{(a, a') : a' \in S \wedge (a, a') \in E\}$$

Es gilt für alle $a_1 \neq a_2 \in S$, dass $E_{a_1} \cap E_{a_2} = \emptyset$ und $\bigcup\limits_{a \in S} E_a = E$.

Gilt nun, dass $\|E_a\| < \frac{1}{2}(\|S\| - 1)$ für alle $a \in S$, also dass man von keiner Stadt in die Hälfte aller anderen Städte direkt gelangt, dann folgt

$$\begin{aligned}\frac{1}{2}\left(\|S\|^2 - \|S\|\right) = \|E\| = \sum_{a \in S} \|E_a\| \quad &< \quad \sum_{a \in S} \frac{1}{2}(\|S\| - 1) \\ &= \quad \|S\| \cdot \frac{1}{2}(\|S\| - 1) \\ &= \quad \frac{1}{2}\left(\|S\|^2 - \|S\|\right)\end{aligned}$$

Dies ist jedoch ein Widerspruch: $\frac{1}{2}\left(\|S\|^2 - \|S\|\right) \not< \frac{1}{2}\left(\|S\|^2 - \|S\|\right)$.
Somit muss es eine Stadt geben von der man direkt in mindestens die Hälfte aller anderen Städte gelangt. $\square$

(b) Wir beweisen die Aussage mittels vollständiger Induktion über die Anzahl n der Städte.

IA: $(n = 2)$ Es ex. immer eine Einbahnstraße zwischen 2 Städten, d.h. von der einen Stadt kommt man auf jeden Fall direkt zur anderen Stadt, somit ist die Behauptung für n=2 gegeben.

IV: Für eine Anzahl der Städte $\leq n - 1$ gilt, dass man mit höchstens 1 Zwischenstopp zu jeder anderen Stadt kommen kann.

IS: $(n - 1 \to n)$ O.B.d.A. nummeriere die Städte von 1 bis n. Betrachte nun zunächst den Graphen der Städte 1 bis $n - 1$. Nach Induktionsvoraussetzung gibt es hier eine Stadt von der man mit höchstens einem Zwischenstopp zu jeder anderen Stadt kommen kann. Sei dies die Stadt k, wobei $1 \leq k \leq n - 1$.
Füge nun die n-te Stadt zum Graphen hinzu. Hierfür gibt es verschiedene Möglichkeiten:

1. Von der Stadt n führen nur Kanten weg zu allen anderen Städten. Dann erreicht man von dieser Stadt aus alle anderen Städte sogar direkt.
2. Zu der Stadt n führt eine Kante von Stadt k aus oder es führt eine Kante von einer von k aus direkt erreichbaren Stadt i zur Stadt n. Man erreicht von der Stadt k aus die Stadt n entweder direkt, oder durch einen Zwischenstopp über die Stadt i. Folglich erreicht über die Stadt k immer noch alle Städte einschließlich der Stadt n mit höchstens einem Zwischenschritt.
3. Keiner der ersten beiden Fälle tritt ein, d.h. es gibt Kanten die zur Stadt n hinführen, jedoch geht diese Kante nicht von der Stadt k bzw. von einer direkt

von k erreichbaren Stadt aus.

Somit sind diese Städte von n direkt erreichbar. Es gibt somit ausschließlich Kanten die von Städten ausgehen, die durch einen Zwischenstopp von k erreichbar sind, die zu n führen.

Betrachtet man nun den Graphen ohne k ist die Anzahl der Städte $n-1$. Hier muss es laut Induktionsvoraussetzung wieder eine Stadt geben, für welche die Aussage gilt. Sei hier die Stadt gerade x. Für diese Stadt x gilt also:

i. Stadt x ist direkt mit Stadt n verbunden und Stadt n erreicht alle Städte, die nicht im Subgraphen enthalten waren direkt, d.h. von Stadt x erreicht man alle Städte, die nicht im Subgraphen enthalten waren über höchstens 1 Zwischenstopp.

ii. Stadt x erreicht alle Städte im Subgraphen mit höchstens einem Zwischenstopp.

Da es sich hier um eine Partition aller Städte handelt, folgt, dass man von Stadt x aus über höchstens 1 Zwischenstopp zu allen anderen Städten kommen kann.

Somit existiert unter den n Städten eine Stadt, von der aus man alle anderen Städte mit höchstens einem Zwischenschritt erreicht. Es folgt somit die Behauptung.

Transfer: **10 Punkte**

Im Rahmen eines Projektes zum Aufbau eines Storage-Area-Networks interessieren Sie sich für Verfahren der *Datenkompression* von Bitstreams. Formal gesehen besteht ein derartiges Komprimierungsverfahren aus zwei Algorithmen: einem Komprimierungsalgorithmus K, der einen gegebenen Bitstring komprimiert, und einem Dekomprimierungsalgorithmus D, der den ursprünglichen Bitstring wieder herstellt. Damit das Verfahren korrekt ist, muss $x = D(K(x))$ für jedes endliche Wort x über dem Alphabet $\{0, 1\}$ gelten.

Bei Ihren Recherchen im Internet sind Sie auf eine Webseite gestoßen, auf der Sie ein Tool mit Namen *OmniZip* herunterladen können, das mit den Worten „*OmniZip* bekommt jede Datei klein“ beschrieben wird und eine Komprimierungsrate von 80% garantiert. Sie sind skeptisch und wollen beweisen, dass es so ein Programm gar nicht geben kann.

Dazu stellen Sie folgende Überlegungen an:

(a) Zeigen Sie, dass es für jede Funktion K, die Bitstrings der Länge n auf Bitstrings kürzerer Länge abbildet, mindestens zwei Bitstrings x und y der Länge n gibt, für die $K(x) = K(y)$ gilt.

(b) Welchen relativen Verlust können Sie für alle Wörter der Länge n bestenfalls erwarten, wenn jedes Wort auf 80% seiner Länge komprimiert wird?

Hinweis: Aus Teilaufgabe (a) folgt, dass bei einer 80%-igen Komprimierungsrate die Dekomprimierung verlustbehaftet sein muss. Der absolute Verlust bei der Komprimierung/Dekomprimierung eines Wortes x der Länge n ist die Anzahl der Bits, in denen sich x und $D(K(x))$ unterscheiden. Der relative Verlust ist das Verhältnis von absolutem Verlust zu Wortlänge. – Um die Aufgabe zu lösen, sollten Sie sich zunächst überlegen, wie viele Wörter der Länge n bestenfalls auf 80% ihrer Länge verlustfrei komprimiert werden können.

Lösung:

(a) Es sei K eine Funktion, die Bitstrings der Länge n auf Bitstrings der Länge $< n$ abbildet, also

$$K : \{0, 1\}^n \to \bigcup_{k=0}^{n-1} \{0, 1\}^k$$

Nach dem Schubfachschluss gibt es wegen

$$\|\{0, 1\}^n\| = 2^n$$

und

$$\left\| \bigcup_{k=0}^{n-1} \{0, 1\}^k \right\| = \sum_{k=0}^{n-1} 2^k = 2^n - 1$$

stets mindestens zwei Bitstrings x und y mit $K(x) = K(y)$.

(b) Für die möglichen Bitstrings steht bestenfalls die ganze Menge $\bigcup\limits_{k=0}^{\lfloor \frac{8}{10}n \rfloor} \{0,1\}^k$ zur Verfügung. Dies sind genau

$$\left\| \bigcup_{k=0}^{\lfloor \frac{8}{10}n \rfloor} \{0,1\}^k \right\| = 2^{\lfloor \frac{8}{10}n \rfloor + 1} - 1 =: m$$

viele Wörter.
Im Besten Fall ist das Maximum der Anzahl an Wörtern am Kleinsten, falls Gleichverteilung gegeben ist. Sonst gäbe es die komprimierten Wörter w und v, für die gilt: $\|D(w)\| < \|D(v)\| + 1$ und man würde nicht mehr den Besten Fall betrachten und der relative Verlust wäre auch größer.
Wir können somit alle Wörter der Länge n in Gruppen aufteilen und somit partitionieren. Dies tun wir, indem wir die 2^n Wörter, die es zu komprimieren gilt auf die Menge der verfügbaren Wörter (gerade m) aufteilen, sodass in jeder Menge genau

$$\frac{2^n}{2^{\lfloor \frac{8}{10}n \rfloor + 1} - 1}$$

Elemente liegen.
Dies sind dann

$$\frac{2^n}{2^{\lfloor \frac{4}{5}n \rfloor + 1} - 1} \leq \frac{2^n}{2^{\lfloor \frac{4}{5}n \rfloor}} = 2^{n - \lfloor \frac{4}{5}n \rfloor} = 2^{\lceil \frac{n}{5} \rceil}$$

Die Anzahl an Bits, die sich unterscheiden, ist gerade $\log(2^{\lceil \frac{n}{5} \rceil}) = \lceil \frac{n}{5} \rceil$. Es ergibt sich somit für den relativen Verlust

$$\frac{\lceil \frac{n}{5} \rceil}{n} \leq \frac{n+1}{5n} = \frac{1}{5} + \frac{1}{5n}$$

Der Bruch $\frac{1}{5n}$ verschwindet bei wachsendem n. Somit kann man für den relativen Verlust den Wert $\approx 0.2 = 20\%$ als obere Schranke setzen.

Eine untere Schranke lässt sich folgendermaßen zeigen:
Für jeden beliebigen Komprimierungsalgorithmus K gilt nach dem erweiterten Schubfachschluss, dass es eine Zeichenkette y gibt, auf die mindestens $\dfrac{2^n}{2^{\lfloor 0.8n \rfloor + 1} - 1} \geq 2^{0.2n-1}$ Zeichenketten abgebildet werden. Sei $x := D(y)$. Wenn der absolute Verlust bei der Abbildung auf y höchstens k beträgt, dann haben alle Urbilder $K^{-1}(y)$ höchstens Hammingdistanz k von x. Es gibt insgesamt $\binom{n}{i}$ Zeichenketten, die Hammingdistanz i haben. Daher gibt es insgesamt $\sum_{i=0}^{k} \binom{n}{i}$ Zeichenketten, die Hammingdistanz höchstens k von x haben. Wenn der absolute Verlust von $K^{-1}(y)$ also k ist, dann muss notwendigerweise $\sum_{i=0}^{k} \binom{n}{i} \geq 2^{0.2n-1}$ gelten. Der absolute Verlust muss daher mindestens $\min\{k \in \mathbb{N}_+ | \sum_{i=0}^{k} \binom{n}{i} \geq 2^{0.2n-1}\}$ sein.

Leider lässt sich diese Summe nicht ohne Weiteres berechnen, deswegen schätzen wir diese noch einmal nach oben ab. Es gilt:

$$
\begin{aligned}
\sum_{i=0}^{k} \binom{n}{i} &\overset{(*)}{<} \binom{n}{k} \frac{n-k+1}{n-2k+1} \\
&\leq 2\binom{n}{k} \qquad\qquad (\text{für } k \leq \frac{n}{3})
\end{aligned}
$$

Darüber hinaus lässt sich folgern, dass

$$
\begin{aligned}
\binom{n}{k} &= \frac{n!}{k!(n-k)!} \\
&\leq \frac{n^k}{k!} \\
&< \frac{n^k}{k!} \cdot \frac{e^k k!}{k^k} \\
&= \left(\frac{ne}{k}\right)^k
\end{aligned}
$$

da $e^k = \sum_{i=0}^{\infty} \frac{k^i}{i!} > \frac{k^k}{k!}$.

Mithin gilt $\sum_{i=0}^{k} \binom{n}{i} \leq 2\left(\frac{ne}{k}\right)^k$. Mit dieser Ungleichung können wir nun den minimalen absoluten Verlust $k_{\min}$ nach unten abschätzen, denn es gilt

$$
\begin{aligned}
k_{\min} &\geq \min\left\{k \in \mathbb{N}_+ | \sum_{i=0}^{k} \binom{n}{i} \geq 2^{0.2n-1}\right\} \\
&\geq \min\left\{k \in \mathbb{N}_+ | 2\left(\frac{ne}{k}\right)^k \geq 2^{0.2n-1}\right\}
\end{aligned}
$$

Durch Lösen dieser Ungleichung ergibt sich, dass der relative Verlust für jeden Kompressionsalgorithmus mindestens $\frac{k_{\min}}{n} \geq \exp\left(1 + W_{-1}\left(\frac{\log(4)}{ne} - \frac{-0.2\log(2)}{e}\right)\right)$ beträgt, wobei $W_{-1}(z)$ den unteren Funktionsast der Lambert'schen W-Funktion bezeichnet. Aus dieser Gleichung lässt sich folgern, dass der relative Verlust $\frac{k_{\min}}{n}$ mindestens 2% ist für $n \geq 35$ und 3% für $n \geq 404$.

Beweis von (*) für $k \leq \frac{n}{2}$:

$$
\begin{aligned}
\sum_{i=0}^{k} \binom{n}{i} &= \binom{n}{k} \sum_{i=0}^{k} \frac{\binom{n}{k-i}}{\binom{n}{k}} \\
&= \binom{n}{k} \sum_{i=0}^{k} \frac{k^{\underline{i}}}{(n-k+1+i)^{\underline{i}}} \\
&\leq \binom{n}{k} \sum_{i=0}^{k} \frac{k^{i}}{(n-k+1)^{i}} \\
&< \binom{n}{k} \sum_{i=0}^{\infty} \left(\frac{k}{n-k+1}\right)^{i} \\
&= \binom{n}{k} \frac{n-k+1}{n-2k+1}
\end{aligned}
$$