Prof. Dr. Leif Kobbelt
Stefan Dollase, Ira Fesefeldt, Alexandra Heuschling, Gregor Kobsik

Lösung - Übung 6

Tutoraufgabe 1 (Gerrymandering):

Verschiedene Wahlsysteme haben verschiedene Stärken und Schwächen. Eine Anforderung an ein Wahlsystem ist, dass es nach der Wahl nicht nur einen Präsident für das ganze Land gibt, sondern auch einen Vertreter des lokalen Wahlkreises. Falls der Präsident jedoch nicht direkt, sondern von den Wahlkreisvertretern gewählt wird, so ergibt sich das Problem des Gerrymandering: Das Wahlergebnis hängt stark davon ab welche Personen welchem Wahlkreis zugeordnet sind. Die Person, welche bestimmt wer welchem Wahlkreis zugeordnet ist hat somit einen starken Einfluss auf den Wahlausgang. Wir betrachten dieses Problem nun in einem Zweiparteiensystem, das heißt jede Person kann entweder den Vertreter von Partei A oder den Vertreter von Partei B wählen. Die Wahlkreisvertreter können wiederum entweder den Präsidenten der Partei A oder den Präsidenten der Partei B wählen. Jeder Wahlkreisvertreter entscheidet sich für den Präsidenten der eigenen Partei.

a) Angenommen jeder Wahlkreis besteht aus 5 Personen. Wie viel Prozent der Wählerstimmen müssen mindestens für Vertreter von Partei A abgegeben werden, damit der Präsident von Partei A durch die Wahlkreisvertreter gewählt werden kann?

b) Angenommen jeder Wahlkreis besteht aus einer beliebigen aber festen Anzahl $k \in \mathbb{N}$ Personen.

i. Wie viel Prozent der Wählerstimmen müssen mindestens für Vertreter von Partei A abgegeben werden, damit der Präsident von Partei A durch die Wahlkreisvertreter gewählt werden kann?

ii. Wie muss k gewählt werden um mit einem möglichst geringen Prozentsatz trotzdem zu gewinnen.

Hinweise:

- Gehen Sie davon aus, dass in allen Wahlkreisen und während der Präsidentschaftswahl die Parteien nie gleich viele Stimmen bekommen.

Lösung

Sei $P = \{p_1, \ldots, p_n\}$ die Menge der Wähler und $v : P \to \{A, B\}$ die Funktion welche jedem Wähler seine Stimme zuordnet. Weiter sein $K = \{K_1, \ldots, K_m\}$ die Menge der Wahlkreise mit $K_1 \uplus \ldots \uplus K_m = P$. Die Wahlkreise partitionieren also die Menge der Wähler.
Die Stimmen der Wahlkreisvertreter lassen sich nun wie folgt berechnen:

$$v_p : K \to \{A, B\}, K_i \mapsto median(\{\!\{v(p) \mid p \in K_i\}\!\})$$

Sei nun $x = median(\{\!\{v_p(K_i) \mid K_i \in K\}\!\})$. Der Präsident der Partei x wird von den Wahlkreisvertretern gewählt.
Um die Wahl zu gewinnen muss der Präsident der Partei A mehr als 50% der Stimmen der Wahlkreisvertreter erhalten.

Hinweis:

- Die Notation $\{\!\{1, 1, 1\}\!\}$ stellt eine Multimenge mit drei Mal dem Wert 1 dar. Eine normale Menge würde natürlich nur einmal den Wert 1 enthalten. Für die Medianberechnung ist die Anzahl jedoch entscheidend. Daher nutzen wir hier Multimengen.

a) Um die Stimme eines Wahlkreisvertreters zu erhalten, muss in diesem Wahlkreis mindestens drei der fünf Stimmen (60%) für Partei A abgegeben worden sein.

Partei A muss also in mehr als 50% der Wahlkreise mindestens 60% der Stimmen erhalten um gewinnen zu können. Insgesamt ist es also ausreichend mehr als $50\% \cdot 60\% = 30\%$ der Stimmen zu erhalten. Wie viel mehr als 30% der Stimmen benötigt werden hängt von der Anzahl der Wahlkreise ab. Je mehr Wahlkreise es gibt, desto geringer ist die Anzahl an zusätzlich benötigten Stimmen. Mit höchstens 30% der Stimmen ist es also nicht möglich die Wahl zu gewinnen.

b) Sei k die Anzahl der Personen in jedem Wahlkreis und m die Anzahl der Wahlkreise. Bei einer geraden Anzahl von Stimmen braucht es eine ganze Stimme mehr um zu gewinnen, bei einer ungeraden Anzahl von Stimmen reicht jedoch eine halbe Stimme mehr. Wir gehen daher nun davon aus, dass sowohl k als auch m ungerade ist, da so jeweils weniger Stimmen notwendig sind. Für gerade Stimmanzahlen funktioniert das Argument aber analog, mit einer etwas anderen Formel. Sei nun $k = 2k' + 1$ und $m = 2m' + 1$. Um die Wahl zu gewinnen sind mindestens $(k' + 1)(m' + 1)$ der $(2k' + 1)(2m' + 1)$ Stimmen notwendig, also ein Stimmanteil von

$$\frac{(k'+1)(m'+1)}{(2k'+1)(2m'+1)} > \frac{(k'+1)(m'+1)}{(2k'+2)(2m'+2)} = \frac{(k'+1)(m'+1)}{4(k'+1)(m'+1)} = \frac{1}{4} .$$

Dieser Wert ist immer größer als 25%. Mit ausreichend großen k' und m' ist es jedoch möglich beliebig nah an die 25% heranzukommen.

Tatsächlich konvergiert obige Formel gegen 30% für $k = 5$ und m gegen unendlich.

Wir stellen fest, dass diese Fragestellung stark verwandt ist mit der Frage wie weit der Median der Mediane vom tatsächlichen Median entfernt sein kann.
Hier einige Anmerkungen welche erklären, warum dieses Problem im deutschen Wahlsystem praktisch nicht auftritt, um das Vertrauen in das deutsche Wahlsystem nicht unnötig zu erschüttern:

- In Deutschland werden zwar Wahlkreisvertreter gewählt, welche auch im Parlament den Kanzler mitwählen. Das Stimmverhältnis im Parlament (und somit auch bei der Kanzlerwahl) wird jedoch per Veráltnis-wahlrecht über die Zweitstimmen festgelegt.
- In Deutschland sind die Wahlen geheim. Das bedeutet es gibt keine zentrale Stelle, welche vor der Wahl weiß welche Person welche Partei wählen wird. Somit ist eine Zuordnung der Stimmen zu den passenden Wahlkreisen praktisch unmöglich.
- In Deutschland bleiben die Wahlkreise über relativ lange Zeiträume unverändert. Es ist beispielsweise nicht üblich nach der Wahl die Wahlkreise neu festzulegen, um bei der nächsten Wahl eine bessere Gewinnchance zu haben.
- Deutschland hat kein Zweiparteiensystem.

Tutoraufgabe 2 (Selektionsalgorithmus mit 9er-Gruppen):

In der Vorlesung haben Sie gesehen, dass der Median der Mediane in linearer Zeit berechnet werden kann. Dazu wurde mithilfe des Medians der Mediane ein Pivot-Element gewählt. Bei der Berechnung des Medians der Mediane wurde zunächst die Eingabe in 5er-Gruppe unterteilt, anschließend jeweils der Median jeder 5er-Gruppe gebildet und schließlich der Median all dieser Mediane gebildet.
Angenommen die Eingabe wäre nicht in 5er-Gruppen eingeteilt worden, sondern in **9er-Gruppen**. Der restliche Selektionsalgorithmus bleibt hingegen unverändert.
Arbeitet der Selektionsalgorithmus weiterhin in linearer Zeit? Beweisen Sie Ihre Antwort.

Lösung

Der Selektionsalgorithmus benötigt mit 9er-Gruppen ebenfalls lineare Zeit.
Zunächst versuchen wir den Beweis der Linearität aus der Vorlesung zu übertagen.
Dafür haben wir, dass mindestens die Hälfte der $\lfloor \frac{n}{9} \rfloor$ Mediane, also $\lceil \frac{1}{2} \lfloor \frac{n}{9} \rfloor \rceil$ Mediane, größer gleich sind als der

Median of Medians. Mindestens die Hälfte der Elemente in einer Gruppe (das sind bei 9 genau 5) sind größer gleich deren Median. Also sind mindestens $5\lceil\frac{1}{2}\lfloor\frac{n}{9}\rfloor\rceil$ Elemente größer gleich der Median of Medians. Damit ist $|S_<| < n - 5\lceil\frac{1}{2}\lfloor\frac{n}{9}\rfloor\rceil < \lfloor\frac{7n}{9}\rfloor$ für $n \geq 50$. Symmetrisch ist damit auch $|S_>| < n - 5\lceil\frac{1}{2}\lfloor\frac{n}{9}\rfloor\rceil < \lfloor\frac{7n}{9}\rfloor$ für $n \geq 50$. Folgende Laufzeitabschätzung lesen wir aus dem Algorithmus ab.

$$T(n) \leq \begin{cases} c_1, & \text{falls } n < 50 \\ c_2 n + T(\lfloor\frac{n}{9}\rfloor) + T(\lfloor\frac{7n}{9}\rfloor), & \text{falls } n \geq 50 \end{cases}$$

Wir benötigen $c_2 n$ Schritte für die Partitionierung, $T(\lfloor\frac{n}{9}\rfloor)$ Schritte um den Median der Mediane exakt zu bestimmen und müssen auf maximal $\frac{7n}{9}$ Werten im rekursiven Aufruf arbeiten.
Wir wollen zeigen, dass $T(n) \leq g(n)$ mit $g(n) = cn = O(n)$.

Induktionsanfang Angenommen $c_1 \leq c$, dann gilt $T(n) = c_1 \leq cn = g(n)$ für $n < 50$.

Induktionshypothese Angenommen es gilt $T(m) \leq cm$ für $m < n$.

Induktionsschritt Sei $n \geq 50$.

$$\begin{aligned} T(n) &\leq c_2 n + T(\lfloor\tfrac{n}{9}\rfloor) + T(\lfloor\tfrac{7n}{9}\rfloor) \\ &\leq c_2 n + c\lfloor\tfrac{n}{9}\rfloor + c\lfloor\tfrac{7n}{9}\rfloor && \text{(Induktionshypothese)} \\ &\leq c_2 n + c\tfrac{n}{9} + c\tfrac{7n}{9} \\ &= c_2 n + cn - c\tfrac{n}{9} \\ &= cn + n(c_2 - c\tfrac{1}{9}) \\ &\leq cn && \text{falls } c_2 - c\tfrac{1}{9} \leq 0 \end{aligned}$$

Die Zusatzubedingung lässt sich wie folgt in eine untere Grenze für c umformen.

$$c_2 - c\frac{1}{9} \leq 0 \iff c_2 \leq c\frac{1}{9} \iff 9c_2 \leq c$$

Die Induktion funktioniert also für $max(c_1, 9c_2) \leq c$. Damit benötigt der Selektionsalgorithmus für 9er-Gruppen ebenfalls lineare Zeit.

Tutoraufgabe 3 (Geschlossenes Hashing mit großen Hashtabellen):

In dieser Aufgabe beschäftigen wir uns mit der erwarteten Anzahl an Sondierungen in einer Hashtabelle bei Nutzung von geschlossenem Hashing. Dafür möchten wir zeigen, dass bei Nutzung einer Hashtabelle, die ausreichend groß ist (genauer, dass $n \leq \frac{m}{2}$ mit n die Anzahl der Eingaben und m die Größe der Hashtabelle) die Länge einer solche Sondierungssequenz im Averagecase lediglich logarithmisch groß in der Größe der Eingabe ist.
Dazu nehmen wir an, dass wir *uniformes hashing* nutzen, d.h. für eine zufällige Eingabe (Schlüssel) x und eine uniforme Hashfunktion mit $j$ter Sondierung $h(x, j)$ auf eine Menge der Größe m ist jede Sequenz $h(x, 1) \dots h(x, m)$ aus unterschiedlichen Hashwerten gleich wahrscheinlich. Insbesondere ist damit auch jeder Wert für $h(x, j)$ gleich wahrscheinlich.
Sei hierfür X_i die Anzahl der Sondierungen, die beim $i$ten Einfügen in die Hashtabelle vorgenommen werden müssen.

Hinweis:

- Aufgrund der Annahme $n \leq \frac{m}{2}$ können wir etwas bessere Abschätzungen angeben als in der Vorlesung.

a) Wir nehmen an, wir nutzen uniformes Hashing. Zeigen Sie, dass die Wahrscheinlichkeit, dass wir für die $i$te Eingabe mit $i \leq n$ in eine Hashtabelle der Größe m mit $n \leq \frac{m}{2}$ echt mehr als k viele Sondierungen vornehmen müssen, maximal $\frac{1}{2^k}$ ist. Formal ausgedrückt, dass $\mathbb{P}(X_i > k) < \frac{1}{2^k}$ gilt.

b) Zeigen Sie, dass die Wahrscheinlichkeit, dass die $i$te Eingabe mit $i \leq n$ in eine Hashtabelle der Größe m mit $n \leq \frac{m}{2}$ mehr als $2\log_2(n)$ viele Sondierungen benötigt, in $O(\frac{1}{n^2})$ ist. Formal ausgedrückt, dass $\mathbb{P}(X_i > 2\log_2(n)) = O(\frac{1}{n^2})$ gilt.

c) Sei $X = \max_{1 \leq i \leq n} X_i$ die maximale Anzahl an Sondierung aller n Eingaben in eine Hashtabelle der Größe m mit $n \leq \frac{m}{2}$. Zeigen Sie, dass die Wahrscheinlichkeit, dass X größer ist als $2\log_2(n)$ in $O(\frac{1}{n})$ ist. Formal ausgedrückt, dass $\mathbb{P}(X > 2\log_2(n)) = O(\frac{1}{n})$.

d) Sei $X = \max_{1 \leq i \leq n} X_i$ die maximale Anzahl an Sondierung aller n Eingaben in eine Hashtabelle der Größe m mit $n \leq \frac{m}{2}$. Zeigen Sie, dass die maximalen Sondierungen pro Eingabe X im Averagecase in $O(\log(n))$ ist. Formal ausgedrückt, dass $\mathbb{E}(X) = O(\log(n))$.

Lösung

a) Da $n \leq \frac{m}{2}$ gilt, ist stets mindestens die Hälfte aller Einträge frei. Damit eine Kollision entsteht, müssen wir aber bereits auf ein belegtes Feld landen. Da $h(x, j)$ aber uniform verteilt ist (sowohl für die Sequenz, als auch für jedes i), ist die Wahrscheinlichkeit eines der belegten Felder zu landen maximal $\frac{1}{2}$. Die Wahrscheinlichkeit, dass dies für eine Eingabe x aber nun k oft passiert ist damit auch maximal $\frac{1}{2^k}$.

b) Mithilfe der Ungleichung der vorherigen Teilaufgabe haben wir mit $k = 2\log_2(n)$ dass $\mathbb{P}(X_i > 2\log_2(n)) < \frac{1}{2^{2\log_2(n)}} = \frac{1}{2^{\log_2(n^2)}} = \frac{1}{n^2} = O\left(\frac{1}{n^2}\right)$.

c) Wir haben nach Definition von X und dem Maximum, dass

$$\mathbb{P}(X > 2\log_2(n)) = \mathbb{P}(\max_{1 \leq i \leq n} X_i > 2\log_2(n)) = \mathbb{P}\left(\bigvee_{1 \leq i \leq n} X_i > 2\log_2(n)\right) .$$

Wir können die Veroderung nach oben durch die Summe der einzel Wahrscheinlichkeiten abschätzen als

$$\mathbb{P}\left(\bigvee_{1 \leq i \leq n} X_i > 2\log_2(n)\right) \leq \sum_{1 \leq i \leq n} \mathbb{P}(X_i > 2\log_2(n)) .$$

Und schließlich mit der Lösung der vorhergehenden Aufgabe haben wir

$$\sum_{1 \leq i \leq n} \mathbb{P}(X_i > 2\log_2(n)) \leq \sum_{1 \leq i \leq n} \frac{1}{n^2} = \frac{n}{n^2} = \frac{1}{n} = O\left(\frac{1}{n}\right) .$$

d) Die längste Anzahl an Sondierungen möglich ist n, da wir auch nur n viele Einträge in der Hashtabelle haben werden.

Weiterhin wissen wir aus der vorherigen Aufgabe, dass $\mathbb{P}(X > 2\log_2(n)) \leq \frac{1}{n}$ gilt und da Wahrscheinlichkeiten maximal 1 sind, dass $\mathbb{P}(X \leq 2\log_2(n)) \leq 1$ gilt. Somit ist:

$$\begin{aligned}
&\mathbb{E}(X) \\
&\leq \mathbb{P}(X \leq 2\log_2(n)) \cdot 2\log_2(n) + \mathbb{P}(X > 2\log_2(n)) \cdot n \\
&\leq 1 \cdot 2\log_2(n) + \frac{1}{n} \cdot n \\
&= 2\log_2(n) + 1 = O(\log(n)) .
\end{aligned}$$