Algorithmen und Datenstrukturen 1

ALGO1 · SoSe-2021 · tcs.uni-frankfurt.de/algo1

⭐-Aufgabe: Zufällige Hashfunktionen

Revision 0e02c20 (2021-06-26)

Oft benutzen wir nicht etwa feste Hashfunktionen, sondern zufällige. Nur so können wir garantieren, dass die gewählte Hashfunktion die für Hashtabellen so wichtigen Eigenschaften auch wirklich haben, wie etwa die Annahme des einfachen gleichmäßigen Hashings.

Sei p eine Primzahl. Wir ziehen eine zufällige Zahl a gleichverteilt aus $\{1,\dots,p-1\}$ und unabhängig eine zufällige Zahl b gleichverteilt aus $\{0,\dots,p-1\}$. Sei $h_{a,b}\colon \mathbb{N} \to \{0,\dots,p-1\}$ die Funktion mit $h_{a,b}(x) = ax + b \bmod p$.

a) Sei x eine beliebige natürliche Zahl ungleich 0 und sei i ein beliebiges Element aus $\{0,\dots,p-1\}$. Finde eine kurze Formel in Abhängigkeit von p für die Wahrscheinlichkeit $\mathbb{P}_{a,b}(h_{a,b}(x) = i)$, und beweise deine Antwort. (Mit anderen Worten, beweise, dass die Annahme des einfachen gleichmäßigen Hashings für die zufällige Hashfunktion $h_{a,b}$ im Erwartungswert gilt.)

b) Sei $m \in \mathbb{N}$ mit $m < p$. Als eigentliche Hashfunktion $h'_{a,b}\colon \{0,\dots,p-1\} \to \{0,\dots,m-1\}$ definieren wir nun $h'_{a,b}(x) = (h_{a,b}(x) \bmod m)$. Beweise, dass für alle natürlichen x, y mit $x < y < p$ gilt:

$$\mathbb{P}_{a,b}(h'_{a,b}(x) = h'_{a,b}(y)) \leq \frac{1}{m}.$$

(Mit anderen Worten, beweise, dass die Kollisionswahrscheinlichkeit höchstens $\frac{1}{m}$ ist.)

c) Sei $0 \leq k \leq m$. Wir ziehen k zufällige Elemente $z_1,\dots,z_k \in \{0,\dots,m-1\}$ unabhängig und gleichverteilt aus $\{0,\dots,m-1\}$ (“mit Zurücklegen”). Sei q_k die Wahrscheinlichkeit, dass *keine* Kollision auftritt, das heißt:

$$q_k = \mathbb{P}_{z_1,\dots,z_k}\Big(\forall i,j \in \{1,\dots,k\}\colon i \neq j \implies z_i \neq z_j\Big).$$

Beweise, dass folgende Ungleichungen gelten:

$$\left(1 - \frac{k}{m}\right)^k \leq q_k \leq e^{-(k-1)k/(2m)}.$$

(Das heißt, selbst in dem idealen Fall, dass k Schlüssel *perfekt zufällig* auf m Buckets verteilt werden, müssen wir bereits für $k = \Theta(\sqrt{m})$ mit mindestens einer Kollision rechnen.)
Hinweis: Die Ungleichung $(1-x) \leq e^{-x}$ gilt stets und darf benutzt werden.

Hinweise zur Abgabe. Bitte schreib deine Beweise möglichst kurz und elegant auf. Den ⭐ erhälst du für die vollständige und korrekte Bearbeitung von Aufgabenteil a). Aufgabenteile b) und c) geben einen Bonus in der Bewertung.