Maximilian Katzmann
Hans-Peter Lehmann
Stefan Walzer

Übungsblatt 14 – Aktivsession

Randomisierte Algorithmik – Wintersemester 2023/2024

Folgende Aufgaben werden in der Aktiv-Session am 8.2.2024 gemeinsam bearbeitet. Gib eine Lösung zu zwei *der folgenden Aufgaben (deine Wahl) bis zum 15.2.2024 über Ilias ab.*

Aufgabe 1 – Deterministisches Schälen in Linearzeit

(a) Der Algorithmus constructByPeeling ist nichtdeterministisch in dem Sinne, dass nicht vorgeschrieben ist, welches i von der While-Schleife ausgewählt wird, wenn es mehrere zulässige Wahlen gibt. Zeige, dass der Erfolg des Algorithmus constructByPeeling nicht davon abhängt. Genauer noch: Für zwei mögliche Ausführungen von constructByPeeling bleibt die exakt gleiche Menge von Schlüsseln am Ende unplatziert.

(b) Verfeinere den Pseudocode so, dass ein deterministischer Algorithmus herauskommt, der erkennbar Laufzeit $\mathcal{O}(n)$ hat. Nutze zur Unterstützung geeignete Datenstrukturen deiner Wahl.

Aufgabe 2 – Ausgedünnte Poissonverteilung

Sei $\lambda \in \mathbb{R}_+$ und $p \in [0, 1]$. Betrachte das zweistufige Zufallsexperiment in dem wir

1. zunächst $X \sim \text{Pois}(\lambda)$ sampeln und
2. dann $Y \sim \text{Bin}(X, p)$ sampeln.

Zeige: $Y \sim \text{Pois}(\lambda p)$.

Hinweis: Es genügt folgende Zutaten zusammenzustöpseln und etwas zu rechnen.

(i) Definition der Poissonverteilung $\Pr_{Z\sim\text{Pois}(\lambda)}[Z = i] = e^{-\lambda}\frac{\lambda^i}{i!}$.

(ii) Definition der Binomialverteilung $\Pr_{Z\sim\text{Bin}(n,p)}[Z = i] = \binom{n}{i}p^i(1-p)^{n-i}$.

(iii) Satz von der totalen Wahrscheinlichkeit: $\Pr[Y = i] = \sum_j \Pr[Y = i \mid X = j]\Pr[X = j]$.

(iv) Siehe Wikipedia: Characterizations_of_the_exponential_function

Aufgabe 3 – Galton-Watson Prozesse und ein bisschen Biologie

Mitochondriale DNA wird von Müttern an ihre Kinder vererbt (der Vater spielt keine Rolle). In dieser Aufgabe betrachten wir die Wahrscheinlichkeit, dass eine bestimmte Mutation, die bei einer Frau (nennen wir sie Eva) auftritt, langfristig überlebt. Wir tun das in einem sehr einfachen Modell, das man Galton-Watson Prozess nennt: Wir nehmen an, dass jede Frau eine $\mathrm{Pois}(2\lambda)$-verteilte Anzahl von Kindern hat, die die Mutation erben.

Sei G_λ der (möglicherweise) unendliche Baum, der entsteht, wenn man mit einem Wurzelknoten startet und jeder Knoten (unabhängig von allen anderen Knoten) eine $\mathrm{Pois}(\lambda)$-verteilte Anzahl von Kindern erhält.

(a) Erkläre: Die Überlebenswahrscheinlichkeit der Mutation entspricht der Wahrscheinlichkeit, dass G_λ unendlich ist. Triff eine geeignete Annahme, um mit den Männern zu verfahren, und verwende Aufgabe 2.

(b) Sei p_i die Wahrscheinlichkeit, dass G_λ mindestens einen Knoten auf Ebene i hat. Drücke p_{i+1} in Abhängigkeit von p_i aus für $i \in \mathbb{N}_0$. Die Wurzelebene sei Ebene 0.

(c) Bestimme für $\lambda \in \{0, 0.5, 1, 1.1, 1.5\}$ jeweils eine Approximationen für die Wahrscheinlichkeit s_λ, dass G_λ unendlich ist. Es bietet sich an Geogebra oder Wolfram Alpha zu nutzen und ähnlich vorzugehen wie in der Vorlesung (als wie eine Rekurrenz der Überlebenswahrscheinlichkeiten beim Schälen betrachtet haben).

Aufgabe 4 – Verzerrte Binomialverteilung

Seien $\alpha, \beta \in \mathbb{R}_+$ sowie $c, d \in \mathbb{N}$ Konstanten und $X \sim \mathrm{Bin}(\lfloor \alpha m \rfloor - c, \beta/m)$. Zeige, dass sich X für $m \to \infty$ einer Poissonverteilung mit Parameter $\lambda = \alpha\beta$ annähert, das heißt zeige:

$$\forall i \in \mathbb{N} : \lim_{m\to\infty} \Pr_{X\sim\mathrm{Bin}(\lfloor \alpha m \rfloor - c, \beta/m)}[X = i] = \Pr_{Z\sim\mathrm{Pois}(\lambda)}[Z = i].$$

Hinweis: Siehe Hinweise (i), (ii), (iv) aus Aufgabe 2.

Der Vollständigkeit halber...

In der Vorlesung wurden noch folgende Aufgaben in Aussicht gestellt. Dafür reicht beim besten Willen leider die Zeit nicht. Für interessierte werden dennoch Lösungen bereitgestellt.

Aufgabe 5 – Peeling mit 2 Hashfunktionen

Angenommen wir verwenden den Schälalgorithmus constructByPeeling in einem Setting mit nur zwei Hashfunktionen h_1 und h_2. Wir gehen vereinfachend davon aus, dass $h_1(x) \neq h_2(x)$ für alle $x \in D$ gilt und unter dieser Einschränkung die Paare $(h_1(x), h_2(x))$ uniform zufällig und für verschiedene $x \in D$ unabhängig sind.[1] Zeige:

(a) Es werden alle Schlüssel platziert genau dann wenn der Graph

$$G = ([m], \{\{h_0(x), h_1(x)\} \mid x \in S\})$$

keine Kreise enthält.
Beachte: G ist als Multigraph zu verstehen, der mehrere Kopien derselben Kante enthalten darf (aber nicht muss).

(b) Falls $n = \Omega(m)$ ist, gibt es einen Kreis in G mit Wahrscheinlichkeit $\Omega(1)$.
Hinweis: Es genügt Kreise der Länge $k = 2$ zu betrachten (das heißt doppelte Kanten).

Aufgabe 6 – Der Schälalgorithmus bleibt nicht erst spät stecken[2]

Für eine Schlüsselmenge $S \subseteq D$ der Größe $n = |S|$ und Hashfunktionen $h_1, h_2, h_3 \sim \mathcal{U}([m]^D)$ betrachten wir den Cuckoo Graphen wie in der Vorlesung:

$$G = G_{S,h_1,h_2,h_3} = (S, [m], \{(x, h_i(x)) \mid x \in S, i \in [3]\})$$

Wir nehmen lediglich $\alpha = \frac{n}{m} \leq 1$ an. Sei $S' \subseteq S$ die Menge der Schlüssel, die vom Schälalgorithmus nicht entfernt werden können (es gilt $|S'| \in \{0, \ldots, n\}$). Wir wollen zeigen, dass $\Pr[|S'| \in \{1, \ldots, \delta m\}] = \mathcal{O}(1/m)$ ist für eine Konstante $\delta > 0$ (die später gewählt wird).

Intution: Entweder gilt $S' = \emptyset$ oder $|S'|$ ist $\Omega(m)$; alles dazwischen ist unwahrscheinlich.

(a) Zeige: $|S'| = 1$ ist nicht möglich.

(b) Zeige: $|N(S')| \leq \frac{3}{2}|S'|$. Hierbei ist $N(S')$ die Menge aller Nachbarn von S' in G.
Hinweis: Betrachte den Subgraphen von G, der von S' und $N(S')$ induziert ist. Wieviele Kanten gibt es? Was ist der Grad der $N(S')$-Knoten?

---

[1] Um das sicherzustellen können wir zum Beispiel $h_2(x) := (h_1(x) + h_{\text{diff}}(x)) \bmod m$ definieren für ein $h_{\text{diff}} : D \to [m-1]$.

[2] Diese Aufgabe ist etwas aufwändig.

(c) Zeige, dass es eine Konstante C gibt, sodass für $s \in \{2, \ldots, n\}$ folgendes gilt:

$$p_s := \Pr[\exists X \subseteq S, |X| = s : \exists Y \subseteq [m], |Y| = \lfloor \tfrac{3}{2}s \rfloor : N(X) \subseteq Y] \leq \left(C \cdot \frac{s}{m}\right)^{s/2}.$$

Hinweis: Einfach brutal Union-Bound über alle Möglichkeiten von X und Y verwenden. Nützlich ist außerdem die Abschätzung $\binom{n}{k} \leq (\frac{ne}{k})^k$ für Binomialkoeffizienten. Ignoriere die Gaussklammern, das machen alle so.

(d) Zeige (i) $p_2 + p_3 + p_4 + p_5 = O(1/m)$, (ii) $\sum_{s=6}^{\sqrt{m}} p_s = O(1/m)$, (iii) $\sum_{s=\sqrt{m}}^{m/(2C)} p_s = O(1/m)$.

(e) Wähle $\delta = 2C$ und zeige $\Pr[|S'| \in \{1, \ldots, \delta n\}] \leq \sum_{s=2}^{\delta n} p_s = O(1/m)$.