Algorithmen und Datenstrukturen 2

ALGO2 · WiSe-2023/24 · tcs.uni-frankfurt.de/algo2/ · 2024-02-05 · 7ae310c

Übungen zu Woche 6: Randomisierte Algorithmen II

Dienstag

6.1 Erwartungswerte.

a) Sei X eine Zufallsvariable, die die Werte 2, 5 und 8 mit folgenden Wahrscheinlichkeiten annimmt:

$$\Pr[X = 2] = \tfrac{1}{3}, \qquad \Pr[X = 5] = \tfrac{1}{2}, \qquad \Pr[X = 8] = \tfrac{1}{6}.$$

Was ist der Erwartungswert von X?

b) Sei Y eine Zufallsvariable, die den Wert 2^i mit Wahrscheinlichkeit $2^{-(i+1)}$ für alle $i \in \{0, 1, 2, \dots\}$ annimmt. Was ist der Erwartungswert von Y?

moodle **6.2 Weihnachtsfeier.** Während der Weihnachtsfeier des Instituts für Informatik möchte Professor Regloh wissen, wer die meisten Kekse im „Bing or Ding"- Spiel gewonnen hat. Er schlägt den folgenden Algorithmus vor:

Algorithm 1: Finde Student:in mit den meisten Keksen

```
max ← −∞
s ← null
Sei S[1...n] ein Array mit Student:innen, zunächst initialisiert mit S[i] = i.
Permutiere das Array S nun gleichverteilt zufällig, sodass eine zufällige Reihenfolge entsteht.
Sei K[i] die von Student:in S[i] gewonnene Anzahl an Keksen.
for i = 1,...,n do
    if K[i] > max then
        max ← K[i] und s ← S[i]        (∗)
    end
end
return s
```

Nimm an, dass alle Student:innen eine unterschiedliche Anzahl an Keksen gewonnen haben, das heißt wir haben $K[i] \neq K[j]$ für alle i, j mit $i \neq j$.

a) Was ist die Wahrscheinlichkeit dafür, dass die im Code mit $(*)$ markierte Zeile in der letzten Iteration der for-Schleife ausgeführt wird?

b) Sei X_i eine Zufallsvariable. Sie nimmt den Wert 1 an, wenn die $(*)$-Zeile in Iteration i ausgeführt wird, und 0 wenn nicht. Berechne $\Pr(X_i = 1)$. *Tipp:* X_i kennst du schon aus der ersten Teilaufgabe

c) Wie oft wird die $(*)$-Zeile ausgeführt? Berechne den Erwartungswert für die Anzahl der Ausführungen. *Tipp:* Benutze die Zufallsvariablen X_i.

6.3 Analyse von Selection. In Abschnitt 13.5 (**KT**) wurde die erwartete Laufzeit von Selection analysiert. Anstatt eine Phase wie im Buch zu definieren, definieren wir eine Phase wie folgt: Der Algorithmus ist in Phase j, wenn die Größe g der betrachteten Menge im folgenden Intervall liegt

$$n\left(\frac{2}{3}\right)^{j+1} < g \leq n\left(\frac{2}{3}\right)^{j}.$$

Führe die Analyse von Selection mit der neuen Phasendefinition durch. Erreicht man mit dieser Analyse im Erwartungswert immer noch eine lineare Laufzeit?

Donnerstag

moodle **6.4 Bierkisten.** Von n Bierkisten $B_1, \ldots, B_n$ enthalten genau k Kisten eine Flasche Bier ($k \leq n$), während die restlichen Kisten leer sind. Du kannst mit bloßem Auge nicht erkennen, welche der Kisten leer sind und welche nicht. Das Ziel ist es, eine Flasche Bier in einer Kiste zu finden. Dafür schlagen wir einen deterministischen und zwei randomisierte Algorithmen vor:

> FindeBier:
> Öffne die Kisten $B_1, \ldots, B_n$ der Reihenfolge nach, wobei das „Öffnen" einer Kiste einem Rechenschritt entspricht. Der Algorithmus terminiert, sobald ein Bier gefunden wurde.
>
> FindeZufälligBier:
> Solange noch kein Bier gefunden wurde, wähle eine uniform zufällige Zahl $i \in \{1, 2, \ldots, n\}$ und öffne die Kiste B_i.
>
> FindeZufälligUngeöffnetesBier:
> Solange noch kein Bier gefunden wurde, wähle uniform zufällig eine bisher noch nicht geöffnete Kiste aus und öffne diese.

a) Was ist die *worst-case* Laufzeit von FindeBier?
b) Was ist die *best-case* Laufzeit von FindeBier?

c) Was ist die *worst-case* Laufzeit von FindeZufälligBier?
d) Was ist die erwartete Laufzeit von FindeZufälligBier?

e) Was ist die *worst-case* Laufzeit von FindeZufälligUngeöffnetesBier?
f) Was ist die erwartete Laufzeit von FindeZufälligUngeöffnetesBier?
Betrachte hierzu folgende Hilfestellung: Sei X die Anzahl der vom Algorithmus geöffneten Kisten. Sei X_i eine Indikatorvariable für jede *leere* Kiste B_i. Wenn die Kiste geöffnet wurde, ist $X_i = 1$, sonst ist $X_i = 0$.

- f_1) Stelle X durch die X_i dar, in der Form $X = \ldots\ldots$
- f_2) Was ist der Erwartungswert von X_i?
- f_3) Benutze die Erwartungswerte der X_i, um den Erwartungswert von X anzugeben.
- f_4) Was ist die erwartete Laufzeit von FindeZufälligUngeöffnetesBier?

6.5 Schrauben und Muttern. Beim Ausmisten deines Kellers hast du genau N Muttern und N Schrauben gefunden. Du erinnerst dich, dass jede Mutter zu *genau* einer Schraube und jede Schraube zu *genau* einer Mutter passt. Aber es nicht möglich ist, ein Paar von Mutter und Schraube mit bloßem Auge zu vergleichen. Um einen Vergleich durchzuführen, musst du die Mutter an der Schraube anbringen und herausfinden, welches der beiden Teile größer ist.

Für dieses Problem ist nur ein sehr komplizierter deterministischer Algorithmus mit Laufzeit $O(N \log N)$ bekannt. Randomisierte Algorithmen dafür sind deutlich einfacher. Gib einen randomisierten Algorithmus an, der mit möglichst wenigen Vergleichen herausfindet, welche Paare zusammenpassen.

Tipp: Modifiziere QuickSort für dieses Problem

Weitere Aufgaben und Projekte

6.6 Randomisierte Pivotwahl. Die Laufzeit von *Quicksort* und *Quickselect* hängt ganz entscheidend von der Wahl des Pivotelements ab. Das Pivotelement sollte das zu sortierende Array in zwei möglichst gleichgroße Teilarrays aufspalten.

Gegeben sei ein unsortiertes Array $A[1..n]$ mit n paarweise verschiedenen Elementen $A[1], \ldots, A[n]$. Wir nehmen an, dass n ganzzahlig durch 4 teilbar ist. Weiterhin sei der *Rang* $r(x)$ *von* x die Position des Elements x im sortierten Array. Zum Beispiel heißt $r(x) = 3$, dass x das drittkleinste Element von A ist.

Eine mögliche Strategie für die Pivotwahl ist Folgende:

SamplePivot(A):

- Wähle uniform zufällig 7 Elemente $p_1, p_2, \ldots, p_7$ aus dem Array A aus. Hierbei können Elemente mehrmals gewählt werden („Ziehen mit Zurücklegen“).
- Liefere den Median dieser 7 Elemente zurück.

a) Was ist die Laufzeit von SamplePivot? (Ohne Begründung.)

b) Berechne $\Pr\left(\frac{n}{4} < r(p_i) \leq \frac{3n}{4}\right)$ für alle $i \in \{1, \ldots, 7\}$ und begründe deine Antwort.

c) Sei

$$X_i = \begin{cases} 1, & \text{wenn } \frac{n}{4} < r(p_i) \leq \frac{3n}{4}; \\ 0, & \text{sonst.} \end{cases}$$

Sei $X = \sum_{i=1}^{7} X_i$. Berechne $\Pr(X \geq 4)$ und begründe deine Antwort.

d) Berechne $\Pr\left(\frac{n}{4} < r(\textsc{SamplePivot}(A)) \leq \frac{3n}{4}\right)$ und begründe deine Antwort.

e) Wir wiederholen SamplePivot(A) jetzt unabhängig so oft, bis das Ereignis

$$\frac{n}{4} < r(\textsc{SamplePivot}(A)) \leq \frac{3n}{4}$$

zum ersten Mal eintritt. Sei W die Anzahl der Wiederholungen. Was ist der Erwartungswert von W? Begründe deine Antwort.