Coupon Collector Problem

Coupon Collector Problem: Wie oft muss man mit Zurücklegen aus $\{1, \ldots, n\}$ ziehen, bis alle Zahlen gezogen wurden?

Analyse Coupon Collector Problem:

- Sei X die Anzahl von Zügen, bis alle Zahlen gezogen wurden.
- X_i: Anzahl Züge, für die man exakt $i - 1$ verschiedene Zahlen hat.
- Offenbar ist $X = 1 + \sum_{i=1}^{n} X_i$. Jedes X_i ist geometrisch verteilt mit
$$p_i = \frac{n-(i-1)}{n}.$$
- Es folgt $\mathbb{E}[X_i] = \frac{1}{p_i} = \frac{n}{n-i+1}$ und

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}[1 + \sum_{i=1}^{n} X_i] = 1 + \sum_{i=1}^{n} \mathbb{E}[X_i] \\
&= 1 + \sum_{i=1}^{n} \frac{n}{n-i+1} = 1 + n \sum_{i=1}^{n} \frac{1}{i} = n \ln n + \Theta(n).
\end{aligned}$$

Quicksort

Algorithmus QUICKSORT

EINGABE: Menge S verschiedener $x_1, \ldots, x_n \subset \mathbb{Z}$

1. IF $|S| \leq 1$, Ausgabe S.
2. Wähle Pivotelement $x \in_R S$.
3. Partitioniere in $L = \{x_i \in S \mid x_i < x\}$ und $R = \{x_i \in S \mid x_i > x\}$.
4. Ausgabe QUICKSORT(L), x, QUICKSORT(R).

AUSGABE: Aufsteigende Sortierung von S.

Anmerkungen:

- Jede Iteration kostet für die Partitionierung $\mathcal{O}(n)$ Vergleiche.
- Im worst case benötigt man $\Omega(n)$ viele Rekursionen.
- Im best case benötigt man $\mathcal{O}(\log n)$ viele Rekursionen.
- Wir zeigen, dass man auch im average case $\mathcal{O}(\log n)$ benötigt.

Analyse Quicksort

Theorem Analyse Quicksort

QUICKSORT benötigt erwartet $2n \ln + \Theta(n)$ Vergleiche.

Beweis:

- Sei $y_1, \ldots, y_n$ die sortierte Reihenfolge von $x_1, \ldots, x_n$.
- IV X_{ij}= 1, falls y_i und y_j von QUICKSORT verglichen werden.
- Wir erhalten für die ZV X für die Gesamtzahl von Vergleichen
$$X = \sum_{1 \le i < j \le n} X_{ij} \text{ und } \mathbb{E}[X] = \sum_{1 \le i < j \le n} \mathbb{E}[X_{ij}].$$
- Es gilt $\mathbb{E}[X_{ij}] = \Pr(X_{ij} = 1)$.
- y_i, y_j werden verglichen gdw eines von beiden als erstes Pivot aus der Menge $\{y_i, \ldots, y_j\}$ ausgewählt werden. (Warum?)
- Dies geschieht mit $\Pr(X_{ij} = 1) = \frac{2}{j-i+1}$. Wir erhalten

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{1 \le i < j \le n} \frac{2}{j-i+1} = \sum_{i=1}^{n} \sum_{j=i+1}^{n} \frac{2}{j-i+1} = \sum_{i=1}^{n} \sum_{j=2}^{n-i+1} \frac{2}{j} = \sum_{j=2}^{n} \sum_{i=2}^{n+1-j} \frac{2}{j} = \\
&= \sum_{j=2}^{n} (n-1+j) \frac{2}{j} = \left((n+1) \sum_{j=2}^{n} \frac{2}{j}\right) - 2(n-1) = (2n+2) \sum_{j=1}^{n} \frac{1}{j} - 4n = 2n \ln n + \Theta(n).
\end{aligned}$$

Markov-Ungleichung

Satz Markov-Ungleichung

Sei $X \geq 0$ eine ZV. Dann gilt

$$\Pr(X \geq a) \leq \frac{\mathbb{E}[X]}{a} \text{ für alle } a > 0.$$

Beweis:

- Definiere IV mit $I = 1$ gdw $X \geq a$. Wegen $X \geq 0$ gilt
$$I \leq \frac{X}{a} \text{ und } \mathbb{E}[I] \leq \mathbb{E}[\tfrac{X}{a}] = \frac{\mathbb{E}[X]}{a}.$$
- Da I eine ZV ist, folgt
$$\Pr(X \geq a) = \mathbb{E}[I] \leq \frac{\mathbb{E}[X]}{a}.$$

Bsp:

- n-facher Münzwürf: Obere Ws-Schranke für mehr als $\frac{3}{4}n$-mal Kopf.
- Sei X_i IV für Kopf im i-ten Wurf. Definiere $X = X_1 + \ldots + X_n$.
- Es gilt $\mathbb{E}[X_i] = \frac{1}{2}$ und $\mathbb{E}[X] = \frac{n}{2}$. Markov-Ungleichung liefert damit
$$\Pr(X \geq \tfrac{3}{4}n) \leq \frac{n/2}{3n/4} = \frac{2}{3}.$$

Momente einer ZV

Definition k-tes Moment

Das *k-te Moment* einer ZV X ist $\mathbb{E}[X^k]$.

Definition Varianz

Die Varianz einer ZV X ist definiert als

$$\mathbf{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

Die Standardabweichung von X ist $\sigma[X] = \sqrt{\mathbf{Var}[X]}$.

Bsp:

- Sei $X = k$ konstant. Dann ist $\mathbf{Var}[X] = \mathbb{E}[(k - \mathbb{E}[k])^2] = 0$.
- Sei k eine Konstante und
$$X = \begin{cases} k\mathbb{E}[X] & \text{mit Ws } \frac{1}{k} \\ 0 & \text{mit Ws } 1 - \frac{1}{k} \end{cases}, \text{ d.h. } X^2 = \begin{cases} k^2(\mathbb{E}[X])^2 & \text{mit Ws } \frac{1}{k} \\ 0 & \text{mit Ws } 1 - \frac{1}{k} \end{cases}.$$
- Es folgt $\mathbf{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = (k-1)(\mathbb{E}[X])^2$.
- D.h. $\mathbf{Var}[X]$ wird beliebig groß, falls X stark von $\mathbb{E}[X]$ abweicht.

Linearität + Kovarianz

Definition Kovarianz

Die *Kovarianz* von zwei ZV X, Y ist

$$\mathbf{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[X])].$$

Satz Linearität + Kovarianz

Für zwei ZV X, Y gilt

$$\mathbf{Var}[X + Y] = \mathbf{Var}[X] + \mathbf{Var}[Y] + 2\mathbf{Cov}(X, Y).$$

Beweis:

$$\begin{aligned}
\mathbf{Var}[X + Y] &= \mathbb{E}[(X + Y - \mathbb{E}[X + Y])^2] = \mathbb{E}[(X - \mathbb{E}[X] + Y - \mathbb{E}[Y])^2] \\
&= \mathbb{E}[(X - \mathbb{E}[X])^2 + (Y - \mathbb{E}[Y])^2 + 2(X - \mathbb{E}[X])(2Y - \mathbb{E}[Y])] \\
&= \mathbb{E}[(X - \mathbb{E}[X])^2] + E[(Y - \mathbb{E}[Y])^2] + 2\mathbb{E}[(X - \mathbb{E}[X])(2Y - \mathbb{E}[Y])] \\
&= \mathbf{Var}[X] + \mathbf{Var}[Y] + 2\mathbf{Cov}(X, Y).
\end{aligned}$$

Kovarianz

Satz Linearität + Kovarianz allgemein

Für ZV $X_1, \ldots, X_n$ gilt allgemein

$$\textbf{Var}[\sum_{i=1}^n X_i] = \sum_{i=1}^n \textbf{Var}[X_i] + \sum_{1 \le i < j \le n} 2\textbf{Cov}(X_i, X_j).$$

Beweis: analog

Satz Multiplikativität von $\mathbb{E}$ für unabhängige ZV

Für unabhängige ZV X, Y gilt $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$.

Beweis:

$$\begin{aligned}
\mathbb{E}[X \cdot Y] &= \sum_i \sum_j (i \cdot j) \cdot \Pr((X = i) \cap (Y = j)) \\
&= \sum_i \sum_j (i \cdot j) \cdot \Pr(X = i) \cdot \Pr(Y = j) \\
&= (\sum_i i \cdot \Pr(X = i)) \cdot (\sum_j j \cdot \Pr(Y = j)) = \mathbb{E}[X] \cdot \mathbb{E}[Y]
\end{aligned}$$

Kovarianz

Anmerkung: Für abhängige X, Y gilt i. Allg. $\mathbb{E}[X \cdot Y] \neq \mathbb{E}[X] \cdot \mathbb{E}[Y]$.

- Wir betrachten einen 2-fachen Münzwurf.
- IV X für Kopf im 1. Wurf, ZV Y für Anzahl Kopf in beiden Würfen.
- Es gilt $\mathbb{E}[X] \cdot \mathbb{E}[Y] = \frac{1}{2} \cdot 1 = \frac{1}{2}$, aber $\mathbb{E}[X \cdot Y] = \frac{3}{4}$.

Satz Linearität von **Var** für unabhängige ZV

Für unabhängige ZV X, Y gilt

$$\mathbf{Cov}(X, Y) = 0 \text{ und damit } \mathbf{Var}[X + Y] = \mathbf{Var}[X] + \mathbf{Var}[Y].$$

Beweis:

$$\begin{aligned}
\mathbf{Cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[X - \mathbb{E}[X]] \cdot \mathbb{E}[Y - \mathbb{E}[Y]] \\
&= (\mathbb{E}[X] - \mathbb{E}[X])(\mathbb{E}[Y] - \mathbb{E}[Y]) = 0
\end{aligned}$$

Korollar Linearität von **Var** für unabhängige ZV

Für paarweise unabhängige ZV $X_1, \ldots X_n$ gilt

$$\mathbf{Var}[\textstyle\sum_{i=1}^n X_i] = \sum_{i=1}^n \mathbf{Var}[X_i].$$

Chebyshev-Ungleichung

Bsp: Varianz der Binomialverteilung $B(n,p)$

- Seien $X_1,\ldots,X_n$ ZV aus $B(1,p)$ und $X = X_1 + \ldots + X_n$. Es gilt
  $\mathbf{Var}[X_i] = \mathbb{E}[(X_i - \mathbb{E}[X_i])^2] = p(1-p)^2 + (1-p)(-p)^2 = p(1-p)$.
- Damit ist $\mathbf{Var}[X] = np(1-p)$.

Satz Chebyshev-Ungleichung

Für eine ZV X und alle $a > 0$ gilt

$$\Pr(|X - \mathbb{E}[X]| \geq a) \leq \frac{\mathbf{Var}[X]}{a^2}.$$

Beweis:

- Es gilt $\Pr(|X - \mathbb{E}[X]| \geq a) = \Pr((X - \mathbb{E}[X])^2 \geq a^2)$.
- Da $(X - \mathbb{E}[X])^2 > 0$ liefert die Markov-Ungleichung

$$\Pr((X - \mathbb{E}[X])^2 \geq a^2) \leq \frac{\mathbb{E}[(X-\mathbb{E}[X])^2]}{a^2} = \frac{\mathbf{Var}[X]}{a^2}.$$

Chebyshev-Ungleichung

Korollar

Für eine ZV X und alle $t > 1$ gilt

1. $\Pr(|X - \mathbb{E}[X]| \geq t \cdot \sigma) \leq \frac{1}{t^2}$,
2. $\Pr(|X - \mathbb{E}[X]| \geq t \cdot \mathbb{E}[X]) \leq \frac{\mathbf{Var}[X]}{t^2(\mathbb{E}[X])^2}$.

Bsp: *n*-facher Münzwurf

- Schranke für die Ws, dass $\geq \frac{3}{4}n$-mal Kopf auftritt.
- Sei X_i IV für Kopf. Es gilt $\mathbb{E}[(X_i^2)] = \mathbb{E}[X_i] = \frac{1}{2}$ und damit
$$\mathbf{Var}[X_i] = \mathbb{E}[(X_i)^2] - (\mathbb{E}[X_i])^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}.$$
- Für $X = X_1 + \ldots + X_n$ folgt $\mathbf{Var}[X] = \frac{n}{4}$. Chebyshev liefert
$$\Pr(X \geq \tfrac{3}{4}n) \leq \Pr(|X - \mathbb{E}[X]| \geq \tfrac{n}{4}) \leq \frac{\mathbf{Var}[X]}{(n/4)^2} = \frac{4}{n}.$$