Алгоритмы. HW#6

Тураев Тимур, 504 (SE)

1.a *Зафиксируем хэш-значение x. Доказать, что вероятность, что k ключей будут иметь хэш x составляет*

$$Q_k = \binom{n}{k} \cdot \left(\frac{1}{n}\right)^k \cdot \left(1 - \frac{1}{n}\right)^{n-k}$$

Это очевидно следует из формулы вероятности для биномиального распределения: есть серия из n независимых испытаний, нужно найти вероятность того, что будет ровно k успехов (то есть хеш ключа равен точно x), вероятность успеха: $1/n$.

1.b *Пусть P_k – вероятность максимальной цепочки иметь длину k. Доказать, что $P_k \leqslant n \cdot Q_k$.*

Пусть X_i – длина цепочки в ячейке i. Тогда $Pr(X_i = k) = Q_k$ для любого i (из пункта a).

$$\begin{aligned}
P_k = Pr\left(\left(\max_{1 \leqslant i \leqslant n-1} X_i\right) = k\right) &\quad = \\
Pr(\exists i, s.t. X_i = k \; and \; \forall i X_i \leqslant k) &\quad \leqslant \\
Pr(\exists i, s.t. X_i = k) &\quad = \\
Pr(\bigcup_{1 \leqslant i \leqslant n} (X_i = k)) &\quad = \\
\sum_{1 \leqslant i \leqslant n} Pr(X_i = k) &\quad = n \cdot Q_k
\end{aligned}$$

1.c *Вывести из Стирлинга, что*

$$Q_k < \left(\frac{e}{k}\right)^k$$

1. Воспользуемся формулой Стирлинга чтобы оценить снизу факториал:

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \left(1 + O\left(\frac{1}{n}\right)\right)$$

$$n! \geqslant \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

$$n! \geqslant \left(\frac{n}{e}\right)^n$$

2. Теперь оценим биноминальный коэффициент:

$$\binom{n}{k} = \frac{n(n-1)(n-2)\dots(n-k+1)}{k!} \leqslant \frac{n^k}{k!}$$

Применим выведенную оценку для факториала из п.1:

$$\binom{n}{k} \leqslant \frac{n^k}{k!} \leqslant \frac{n^k \cdot e^k}{k^k}$$

3. Применим полученные оценки для выявления верхней границы Q_k

$$Q_k = \binom{n}{k} \cdot \left(\frac{1}{n}\right)^k \cdot \left(1 - \frac{1}{n}\right)^{n-k} \leqslant \binom{n}{k} \cdot \frac{1}{n^k} \cdot 1 \leqslant \frac{n^k \cdot e^k}{k^k} \cdot \frac{1}{n^k} = \frac{e^k}{k^k}$$

1.d *Показать, что для некоторого $c > 1$ верно $Q_k \leqslant 1/n^3$ при $k \geqslant c \cdot \frac{\log n}{\log \log n}$*

С учетом пункта с осталось показать, что

$$\left(\frac{e}{k}\right)^k \leqslant 1/n^3$$

для некоторого $k \geqslant c \cdot \frac{\log n}{\log \log n} = k_0$

Возьмем логарифм обеих частей:

$$\begin{aligned} k_0 \cdot \log\left(\frac{e}{k_0}\right) &\leqslant -3 \cdot \log n \\ k_0 \cdot (\log e - \log k_0) &\leqslant -3 \cdot \log n \\ c \cdot \frac{\log n}{\log \log n} \cdot (\log e - \log \frac{c \cdot \log n}{\log \log n}) &\leqslant -3 \cdot \log n \\ \frac{c}{\log \log n} \cdot (\log e - \log c - \log \log n + \log \log \log n) &\leqslant -3 \end{aligned}$$

Оценим $\log e - \log c$ сверху нулем, это верно для $c > e$

$$\begin{aligned} \frac{c}{\log \log n} \cdot (-\log \log n + \log \log \log n) &\leqslant -3 \\ \frac{c}{\log \log n} \cdot (\log \log n - \log \log \log n) &\geqslant 3 \\ c &\geqslant \frac{3 \cdot \log \log n}{\log \log n - \log \log \log n} \\ c &\geqslant \frac{3}{1 - \frac{\log \log \log n}{\log \log n}} \end{aligned}$$

Правая штука имеет глобальный максимум в точке e^{e^e} равный $\frac{3e}{e-1} \approx 4.7$. Поэтому можно указать такое $c \geqslant 5$ для которого доказываемая оценка верна.

1.e *Доказать оценку на матожидание максимальной длины цепочки*

По определению матожидания имеем

$$\begin{aligned} EM = \sum_{k=1}^{n} k \cdot Pr(M = k) &= \\ \sum_{k=1}^{k_0} k \cdot Pr(M = k) + \sum_{k=k_0+1}^{n} k \cdot Pr(M = k) &\leqslant \\ \sum_{k=1}^{k_0} k_0 \cdot Pr(M = k) + \sum_{k=k_0+1}^{n} n \cdot Pr(M = k) &= \\ k_0 \cdot Pr(M \leqslant k) + n \cdot Pr(M > k_0) &= \\ Pr\left[M > c \cdot \frac{\log n}{\log \log n}\right] \cdot n + Pr\left[M \leqslant c \cdot \frac{\log n}{\log \log n}\right] \cdot c \cdot \frac{\log n}{\log \log n} & \end{aligned}$$

Теперь можно вывести оценку:

$$\begin{aligned} EM \leqslant k_0 \cdot Pr(M \leqslant k) + n \cdot Pr(M > k_0) &= \\ k_0 \cdot 1 + n \cdot \sum_{k=k_0+1}^{n} Pr(M = k) &\leqslant \\ k_0 + n \cdot \sum_{k=k_0+1}^{n} P_k & \end{aligned}$$

Но оценку на P_k мы знаем из второго пункта. Поэтому имеем

$$
\begin{aligned}
EM \leqslant k_0 + n \cdot \sum_{k=k_0+1}^{n} P_k \leqslant \\
k_0 + n \cdot \sum_{k=k_0+1}^{n} n \cdot Q_k = \\
k_0 + n^2 \cdot \sum_{k=k_0+1}^{n} Q_k
\end{aligned}
$$

А оценку на Q_k берем из четвертого пункта:

$$
\begin{aligned}
EM \leqslant k_0 + n^2 \cdot \sum_{k=k_0+1}^{n} Q_k \leqslant \\
k_0 + n^2 \cdot \sum_{k=k_0+1}^{n} 1/n^3 \leqslant \\
k_0 + n^2 \cdot n \cdot 1/n^3 = k_0 + 1 = c \cdot \frac{\log n}{\log\log n} + 1 = O\left(\frac{\log n}{\log\log n}\right)
\end{aligned}
$$