Solution for Problem Set 7

201300035 方盛俊

Problem 1

(a)

由 Stirling's approximation 可知 $k! \sim \sqrt{2\pi n}(\frac{n}{e})^n$, 或 $\sqrt{c_1 n}(\frac{n}{e})^n \leqslant k! \leqslant \sqrt{c_2 n}(\frac{n}{e})^n$, 则 $\frac{\sqrt{c_1 k}}{k!} \leqslant (\frac{e}{k})^k$

即证 $Q_k = \binom{n}{k}\frac{1}{n^k}(1-\frac{1}{n})^{n-k} = \frac{n!}{k!(n-k)!} \cdot \frac{(n-1)^{n-k}}{n^n} \leqslant \frac{\sqrt{c_1 k}}{k!} \leqslant (\frac{e}{k})^k$

即证 $\frac{n!}{(n-k)!} \cdot \frac{(n-1)^{n-k}}{n^n} \leqslant \sqrt{c_1 k}$

即证 $\frac{\sqrt{c_2 n}(n-1)^{n-k}}{e^n (n-k)!} \leqslant \sqrt{c_1 k}$

可以看出, 当 k 递增时, 不等式左边递减, 右边递增, 则我们令 $k=1$, 即只需证明

$$\frac{\sqrt{n}(n-1)^{n-1}}{e^n (n-1)!} \leqslant \sqrt{\frac{c_1}{c_2}}$$

即证 $\sqrt{n}(n-1)^{n-1} \leqslant \sqrt{\frac{c_1}{c_2}} e^n \cdot \sqrt{c_1(n-1)}(\frac{n-1}{e})^{n-1} \leqslant \sqrt{\frac{c_1}{c_2}} e^n (n-1)!$

即证 $\sqrt{c_2 n} \leqslant e\sqrt{c_1^2(n-1)}$

即证 $\frac{n}{n-1} \leqslant e^2 \cdot \frac{c_1^2}{c_2}$

因为 c_1 和 c_2 是及其接近的一个数, 所以可以近似看作 $c_1 \approx c_2$, 那么就有 $\frac{c_1^2}{c_2} \approx 1$ 成立

则 $\frac{n}{n-1} \leqslant \frac{2}{2-1} = 2 \leqslant e^2 \cdot \frac{c_1^2}{c_2}$ 成立

则原式 $Q_k=\binom{n}{k}\frac{1}{n^k}(1-\frac{1}{n})^{n-k}\leqslant(\frac{e}{k})^k$ 成立.

(b)

记时间 A_i 为第 i 个位置刚好有 k 个键且该位置有最多的键. 则 $P(A_i)\leqslant Q_k$

由布尔不等式可知 $P_k=P(\bigcup_{i=1}^{n}A_i)\leqslant\bigcup_{i=1}^{n}Q_k=nQ_k$

(c)

即证 $P_k\leqslant nQ_k\leqslant n(\frac{e}{k})^k<\frac{1}{n^2}$

即证 $k\lg k-k\lg e>3\lg n$

当 $k\geqslant 3$ 时, k 递增, 则 $k\lg k-k\lg e$ 也递增

因为 $k\geqslant\frac{c\lg n}{\lg\lg n}$, 只需带入 $k=\frac{c\lg n}{\lg\lg n}$

则只需找出 c 使得

$$\frac{c\lg n}{\lg\lg n}(\lg\frac{c\lg n}{\lg\lg n}-\lg e)>3\lg n$$

即 $\frac{c}{\lg\lg n}(\lg\frac{\lg n}{\lg\lg n}+(\lg c-\lg e))>3$

令 $t=\lg\lg n$, 则 $2^t=\lg n$, 即

$$c(1-\frac{\lg t}{t}+\frac{\lg c-\lg e}{t})>3$$

令 $f(t)=1-\frac{\lg t}{t}+\frac{\lg c-\lg e}{t}$

则 $f'(t)=-\frac{\frac{1}{\ln 2}-\lg t}{t^2}-\frac{\lg c-\lg e}{t^2}=\frac{\lg t-\lg c}{t^2}$

所以带入 $t=2$, 则有 $c(1-\frac{\lg t}{t}+\frac{\lg c-\lg e}{t})=c(\frac{1}{2}+\frac{\lg c-\lg e}{2})>3$

取 $c=8$, 则 $c(\frac{1}{2}+\frac{\lg c-\lg e}{2})=4(1+3-\lg e)>3$

因为 $n\geqslant\frac{8\lg n}{\lg\lg n}$, 带入 $n=16$ 有 $16\geqslant\frac{8\times 4}{2}$ 恰好成立, 说明 $n\geqslant 16$, 即 $t\geqslant 2$

现在有 $f(t) = 1 - \frac{\lg t}{t} + \frac{3 - \lg e}{t}, f'(t) = \frac{\lg t - 3}{t^2}, t \geqslant 2$

则 $t = 8$ 时有最小值 $f(8) = 1 - \frac{\lg t}{t} + \frac{3 - \lg e}{t} = 1 - \frac{3}{8} + \frac{3 - \lg e}{8} = 1 - \frac{\lg e}{8} > 0$

可证明出, $k = 8$ 满足 $P_k < \frac{1}{n^2}$, 当 $k \geqslant \frac{c \lg n}{\lg \lg n}$

(d)

$$
\begin{aligned}
E(M) &= P(M < \frac{c\lg n}{\lg\lg n}) \cdot E(M|M < \frac{c\lg n}{\lg\lg n}) + \sum_{M=\frac{c\lg n}{\lg\lg n}}^{n} P(M) \cdot M \\
&\leqslant 1 \cdot O(\frac{c\lg n}{\lg\lg n}) + \sum_{M=\frac{c\lg n}{\lg\lg n}}^{n} \frac{1}{n^2} \cdot M \\
&\leqslant 1 \cdot O(\frac{c\lg n}{\lg\lg n}) + \sum_{M=1}^{n} \frac{1}{n^2} \cdot n \\
&= 1 \cdot O(\frac{c\lg n}{\lg\lg n}) + O(1) \\
&= \cdot O(\frac{c\lg n}{\lg\lg n})
\end{aligned}
$$

Problem 2

(a)

令 $y = \sum_{i=0}^{n-1} 2^{ip} y_i$, 则其的重新排列 $x = \sum_{j=0}^{n-1} 2^{jp} y_{i_j}$

$\because (2^p y_i \mod m) = (\sum_{k=1}^{y_i} (2^p \mod m) \mod m) = (\sum_{k=1}^{y_i} 1 \mod m) = (y_i \mod m)$

$$
\begin{aligned}
\therefore h(y) &= \sum_{i=0}^{n-1} 2^{ip} y_i \mod m \\
&= \sum_{i=0}^{n-1} (2^{ip} y_i \mod m) \mod m \\
&= \sum_{i=0}^{n-1} (y_i \mod m) \mod m
\end{aligned}
$$

同理 $h(x) = \sum_{i=0}^{n-1} (y_i \mod m) \mod m$

则 $h(y) = h(x)$

(b)

因为 $h_2(k)$ 和 m 有最小公因数 d, 所以原式可以写成

$h(k,i) = (h_1(k) + i \cdot \frac{h_2(k)}{d} \cdot d) \mod \frac{m}{d} \cdot d = [h_1(k) + d \cdot (i \cdot \frac{h_2(k)}{d} \mod \frac{m}{d})] \mod m$

令 $h(k,i) = h_1(k)$, 我们又有 $d \cdot (i \cdot \frac{h_2(k)}{d} \mod \frac{m}{d}) < m$,

所以只能 $(i \cdot \frac{h_2(k)}{d} \mod \frac{m}{d}) = 0$, 才有 $h(k,i) = h_1(k)$ 成立.

我们又有 $\frac{h_2(k)}{d}$ 与 $\frac{m}{d}$ 互素,

当 $0 < i < \frac{m}{d}$ 时, 假设 $(i \cdot \frac{h_2(k)}{d} \mod \frac{m}{d}) = 0$

则 i 能被 $\frac{m}{d}$ 整除, 这与 $0 < i < \frac{m}{d}$ 矛盾, 所以 $(i \cdot \frac{h_2(k)}{d} \mod \frac{m}{d}) \neq 0$

当 $i = \frac{m}{d}$ 时, 恰有 $(i \cdot \frac{h_2(k)}{d} \mod \frac{m}{d}) = 0$

说明此时 $h(k,i) = h_1(k)$

所以原命题成立.

Problem 3

(a)

对于任意两个元素来说, 他们碰撞的概率为 $\frac{1}{m}$

所以碰撞的预期值为 $E(X)=\binom{n}{2}\cdot\frac{1}{m}=\frac{n(n-1)}{2m}$

(b)

概率为 $p=\frac{m}{m}\cdot\frac{m-1}{m}\cdot\cdots\cdot\frac{m-n+1}{m}=\frac{m!}{m^n\cdot n!}$

(c)

设 Y 为找到一个完美哈希函数时, 已经尝试了多少个不同的函数.

则 $P(Y=k)=p(1-p)^{k-1}$

$\therefore E(Y)=\sum_{k=1}^{\infty}kp(1-p)^{k-1}=\frac{1}{p}=\frac{m^n\cdot n!}{m!}$

(d)

前 N 个函数均不是完美的概率是 $p'=(1-p)^N=(1-\frac{m!}{m^n\cdot n!})^N$

(e)

令 $1-p'=1-\frac{1}{n}$

$\therefore (1-\frac{m!}{m^n\cdot n!})^N=\frac{1}{n}$

$\therefore N[\lg(m^n\cdot n!-m!)-\lg(m^n\cdot n!)]=-\lg n$

$\therefore N=\frac{\lg n}{n\lg m+\lg(n!)-\lg(m^n\cdot n!-m!)}$

所以至少要找 $N=\frac{\lg n}{n\lg m+\lg(n!)-\lg(m^n\cdot n!-m!)}$ 个哈希函数

Problem 4

我们认为 $\hat{c}_i=3$, 下面是证明:

这一系列的操作, 对应的开销分别是:

```
1 2 1 4 1 1 1 8 ...
```

对应序号:

```
1 2 3 4 5 6 7 8 ...
```

我们将第 2^k 位置的开销 2^k 均摊到 $2^{k-1}+1$ 到 2^k 位置的开销中, 可以看出, 每一位需要多均摊 `2` 的开销.

并且, 随着 k 的变化, 每一位最多只会均摊一次 `2`, 不会有重叠的部分, 所以我们可以保证 $\sum_{i=1}^{n} c_i \leqslant \sum_{i=1}^{n} \hat{c}_i$

所以可得均摊开销是 $O(1)$.

Problem 5

我们令 $\hat{c}_i = 9$

目标: 证明 $\sum_{i=1}^{k} c_i \leqslant \sum_{i=1}^{k} \hat{c}_i$, 对于任何 $i \in \mathbb{N}^+$

策略: 使用归纳法, 保证余额永远非负.

Basis: 在第一次操作之前, 余额为 0.

I.H.: 在第 i 次操作之前, 余额非负.

I.S.:

消耗最大的路径是 $\frac{3}{4}n \xrightarrow{-n} \frac{3}{8}n \xrightarrow{+(\hat{c}_i-1)\cdot\frac{1}{8}n} \frac{1}{4}n \xrightarrow{-n} \frac{1}{2}n \xrightarrow{+(\hat{c}_i-1)\cdot\frac{1}{4}n} \frac{3}{4}n \to \cdots$

可以看出, 只要我们确保 $(\hat{c}_i - 1)\cdot\frac{1}{8}n \geqslant n$, 即 $\hat{c}_i \geqslant 9$, 就可以保证余额仍然非负.

由以上分析可知, 均摊消耗仍然是 $O(1)$