# 期中考试参考答案

2022 年 5 月 2 日

## 一 ~ 五

A, C, B, B, ACD

## 六

### (1)

主方法不可用于求解 $T(n) = 8T\left(\frac{n}{2}\right) + 3n^3\log^3 n$。理由如下：
该递归式符合主方法考虑的 $T(n) = aT\left(\frac{n}{b}\right) + f(n)$ 的形式，其中 $a = 8, b = 2, f(n) = \Theta\left(n^3\log^3 n\right)$。这里 $n^{\log_b a} = n^{\log_2 8} = n^3$，满足 $f(n) = \Omega\left(n^{\log_b a}\right)$，但由于 $f(n)$ 渐进上只比 $n^{\log_b a}$ 大一个 $\log^3 n$ 的项，故不存在常数 $\varepsilon > 0$ 满足 $f(n) = \Omega\left(n^{\log_b a+\varepsilon}\right)$，从而该递归式不符合主方法的任何一种情况，不能使用主方法求解。

### (2)

下面用迭代法求解 $T(n) = 8T\left(\frac{n}{2}\right) + 3n^3\log^3 n$。假设 $n$ 是 2 的正整数次幂，有：

$$
\begin{aligned}
T(n) &= 8T\left(\frac{n}{2}\right) + 3n^3\log^3 n \\
&= 8\left(8T\left(\frac{n}{2^2}\right) + 3\left(\frac{n}{2}\right)^3\log^3\left(\frac{n}{2}\right)\right) + 3n^3\log^3 n \\
&= 8^2T\left(\frac{n}{2^2}\right) + 3n^3\left(\log^3 n + (\log n - 1)^3\right) \\
&= 8^2\left(8T\left(\frac{n}{2^3}\right) + 3\left(\frac{n}{2^2}\right)^3\log^3\left(\frac{n}{2^2}\right)\right) + 3n^3\left(\log^3 n + (\log n - 1)^3\right) \\
&= 8^3T\left(\frac{n}{2^3}\right) + 3n^3\left(\log^3 n + (\log n - 1)^3 + (\log n - 2)^3\right) \\
&= \cdots \\
&= 8^{\log n} + 3n^3\sum_{k=1}^{\log n} k^3 \\
&= \Theta\left(n^3\log^4 n\right).
\end{aligned}
$$

# 七

算法：利用最坏线性时间求中位数的算法求出 $n$ 个油井的 $y$ 坐标的中位数，作为主管道的位置。

正确性证明：

证明. 设 $n$ 个油井的 $y$ 坐标从小到大依次为 $y_1, y_2, \dots, y_n$，则需要确定主管道的位置 $y$，满足 $\sum_{k=1}^{n}|y-y_k|$ 最小。

当 $n$ 为偶数时，由绝对值不等式，可得

$$
\begin{aligned}
&\sum_{k=1}^{n}|y-y_k| \\
=&\sum_{k=1}^{n/2}\left(|y-y_k| + |y-y_{n+1-k}|\right) \\
\geq&\sum_{k=1}^{n/2}(y_{n+1-k}-y_k),
\end{aligned}
$$

当且仅当对任意 $1 \leq k \leq \frac{n}{2}$ 均有 $y_k \leq y \leq y_{n+1-k}$ 时，等号成立。而算法得到的 $y$ 为中位数 $y_{n/2}$，满足该条件。

类似地，当 $n$ 为奇数时，有

$$
\begin{aligned}
&\sum_{k=1}^{n}|y-y_k| \\
\geq &\sum_{k=1}^{\lfloor n/2\rfloor}\left(|y-y_k|+|y-y_{n+1-k}|\right) \\
\geq &\sum_{k=1}^{\lfloor n/2\rfloor}(y_{n+1-k}-y_k),
\end{aligned}
$$

当且仅当对任意 $1\leq k\leq\left\lfloor\frac{n}{2}\right\rfloor$ 均有 $y_k\leq y\leq y_{n+1-k}$，且 $y=y_{\lfloor n/2\rfloor+1}$ 时，等号成立。而算法得到的 $y$ 为中位数 $y_{\lfloor n/2\rfloor+1}$，满足条件。

综上，无论 $n$ 为偶数还是奇数，算法均能返回最优结果。

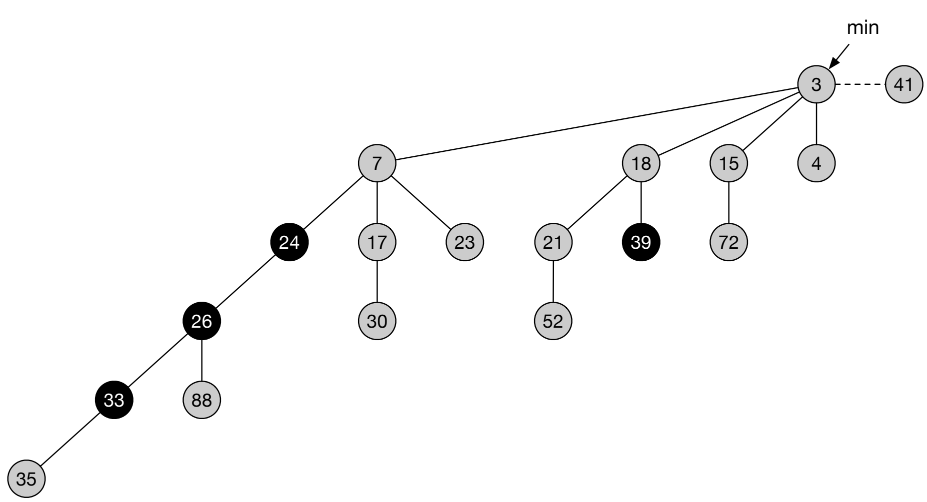
□

时间复杂度：最坏 $O(n)$。

# 八

## (1)

最终结果（答案不唯一）：



真实代价：min 节点的儿子节点个数加上根链表中其它节点的个数，为 6。

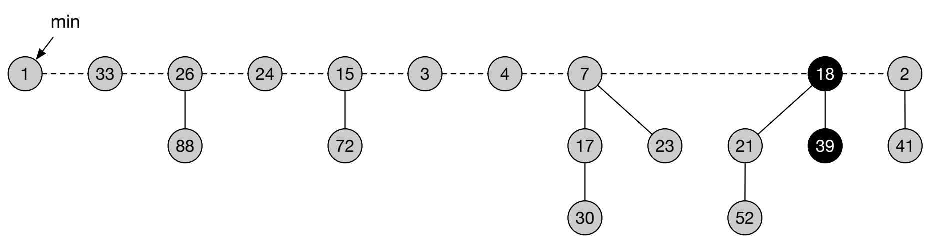
势能函数表达式：$t(H)+2m(H)$，其中 $t(H)$ 为根链表中节点的个数，$m(H)$ 为被标记的节点个数。

势能变化：根链表中节点个数减少 4 个，标记节点个数减少 1 个，故势能变化为 $-4-2\times 1=-6$。

摊还代价：$6+(-6)=0$（这里真实代价实际上是没法准确计算的，大致正确即可）。

### (2)

最终结果：




真实代价：执行切割的次数，为 4。

势能变化：根链表中节点个数增加 4 个，标记节点个数减少 3 个，故势能变化为 $4-2\times 3=-2$。

摊还代价：$4+(-2)=2$。

## 九

### (1)

设二进制计数器 $A$ 的势能函数为 $\Phi(A)$ 等于满足 $A[i]=1$ 的 $i$ 的个数。该势能函数满足 $\Phi(A)\geq 0$ 恒成立。由本小题的假设知，$\Phi(A)$ 的初始值为 0。故可将单次操作的实际代价与势能差之和作为摊还代价。

设一次INCREMENT操作中将 $c$ 个 $A[i]$ 由 1 重置为了 0。由伪代码知该次INCREMENT操作的实际代价为 $O(c)$，势能差为 $1-c$，故摊还代价为

$O(c)+1-c=O(1)$，因为我们可以设置势能的单位来抵消前面 $O(c)$ 中的常数。

### (2)

势能函数的定义同上，此时势能函数仍满足 $\Phi(A) \geq 0$ 恒成立，但不满足初始值 $\Phi(A_0)=0$。

记 $n$ 次`INCREMENT`的实际代价分别为 $c_1, c_2, \ldots, c_n$，实际代价加上势能差分别为 $\hat{c}_1, \hat{c}_2, \ldots, \hat{c}_n$。由第 (1) 小问知 $\hat{c}_i = O(1), \forall 1 \leq i \leq n$。由 $\hat{c}_i$ 的定义知 $\sum_{i=1}^n \hat{c}_i = \sum_{i=1}^n c_i + \Phi(A_n) - \Phi(A_0)$，其中 $A_n, A_0$ 分别为 $n$ 次操作结束后和初始时的计数器。则有

$$
\begin{aligned}
\sum_{i=1}^n c_k &= \sum_{i=1}^n \hat{c}_i + \Phi(A_0) - \Phi(A_n) \\
&\leq \sum_{i=1}^n \hat{c}_i + \Phi(A_0) \\
&\leq \sum_{i=1}^n \hat{c}_i + k \\
&= O(n+k).
\end{aligned}
$$

因此，可将 $O\left(\frac{n+k}{n}\right) = O\left(1+\frac{k}{n}\right)$ 作为单次`INCREMENT`的摊还代价。当 $n = O(1)$ 时，该代价为 $O(k)$；当 $n = \Omega(k)$ 时，该代价为 $O(1)$。

## 十

### (1)

证法 *1.* 对 $1 \leq i < j \leq n$，设随机变量 $X_{ij}$ 为事件 “$P[i]=P[j]$” 的指示器随机变量，则 $\mathrm{E}[X_{ij}] = \Pr\left[P[i]=P[j]\right] = \frac{1}{n^3}$。

设随机变量 $X$ 为满足 $P[i]=P[j], i<j$ 的 $i,j$ 的个数，则 $X = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} X_{ij}$，根据期望的线性性，有 $\mathrm{E}[X] = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \mathrm{E}[X_{ij}] = \binom{n}{2}/n^3 \leq \frac{1}{n}$。

根据马尔可夫不等式，有

$$\Pr[X \geq 1] = \Pr\left[X \geq \frac{1}{\mathrm{E}[X]} \cdot \mathrm{E}[X]\right] \leq \left(\frac{1}{\mathrm{E}[X]}\right)^{-1} = \mathrm{E}[X] \leq \frac{1}{n},$$

即 $P[1], P[2], \ldots, P[n]$ 中有相同值的概率不超过 $\frac{1}{n}$，故它们均不相同的概率至少为 $1 - \frac{1}{n}$。

$\square$

*证法 2.* 对 $1 \leq i < j \leq n$，设 $A_{ij}$ 为事件 “$P[i] = P[j]$”，则 $\Pr[A_{ij}] = \frac{1}{n^3}$。由合集不等式，有

$$\begin{aligned}
&\Pr\left[\bigcup_{i=1}^{n-1} \bigcup_{j=i+1}^{n} A_{ij}\right] \\
\leq &\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \Pr[A_{ij}] \\
= &\binom{n}{2} \Big/ n^3 \\
\leq &\frac{1}{n},
\end{aligned}$$

即 $P[1], P[2], \ldots, P[n]$ 中有相同值的概率不超过 $\frac{1}{n}$，故它们均不相同的概率至少为 $1 - \frac{1}{n}$。

$\square$

### (2)

用较好的方法实现该算法的时间复杂度是 $O(n)$ 的。理由如下：
首先，算法除对 $P$ 数组进行排序外均可在 $O(n)$ 时间内运行。而注意到 $P$ 数组中的数均为 $[1, n^3]$ 中的整数，可以将它们减去 1 之后视为不超过 3 位的 $n$ 进制数，来进行基数排序，即 3 次复杂度为 $O(n+n) = O(n)$ 的计数排序，总时间复杂度为 $O(n)$。

# 十一

### (1)

证明. 对 $i = 1, 2, \ldots, n$，设随机变量 $X_i$ 为事件“投 $n$ 个球后第 $i$ 个盒子为空”的指示器随机变量，则 $\mathrm{E}[X] = \Pr[\text{投}n\text{个球后第}i\text{个盒子为空}] = \left(1 - \frac{1}{n}\right)^n$，因为每次投球未投到第 $i$ 个盒子的概率均为 $1 - \frac{1}{n}$，而 $n$ 次投球是独立的。

设随机变量 $X$ 为投 $n$ 个球后空盒子所占的比例，则 $X = \frac{1}{n}\sum_{i=1}^{n} X_i$，由期望的线线性有 $\mathrm{E}[X] = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[X_i] = \frac{1}{n} \cdot n \cdot \left(1 - \frac{1}{n}\right)^n = \left(1 - \frac{1}{n}\right)^n$，进一步有

$$\lim_{n\to\infty}\mathrm{E}[X] = \lim_{n\to\infty}\left(1 - \frac{1}{n}\right)^n = \lim_{n\to\infty}\left(\left(1 + \frac{1}{-n}\right)^{-n}\right)^{-1} = \frac{1}{e},$$

即为欲证结论。

□

### (2)

证明. 考虑将该扔球的过程划分为 $n$ 个阶段 $S_1, S_2, \ldots, S_n$，其中第 $i(1 \leq i \leq n)$ 个阶段为恰好有 $i-1$ 个盒子有球时进行的所有扔球，包括其中最后一次使得有 $i$ 个盒子有球的那次扔球。对 $1 \leq i \leq n$，设随机变量 $X_i$ 为第 $i$ 个阶段中扔球的个数。

注意到，阶段 $S_i(1 \leq i \leq n)$ 中，每次扔球前恰有 $i-1$ 个盒子有球，而 $X_i$ 为该阶段中第一次将球扔到空盒子中时所扔的球数，故 $X_i$ 服从成功概率为 $1 - \frac{i-1}{n} = \frac{n-i+1}{n}$ 的几何分布，从而有 $\mathrm{E}[X] = \frac{n}{n-i+1}$。

由题意有 $X = \sum_{i=1}^{n} X_i$，由期望的线线性得到

$$
\begin{aligned}
\mathrm{E}[X] &= \sum_{i=1}^{n} \mathrm{E}[X_i] \\
&= \sum_{i=1}^{n} \frac{n}{n-i+1} \\
&= n\sum_{i=1}^{n} \frac{1}{i} \\
&= O(n \log n),
\end{aligned}
$$

即为欲证结论。

□

# 十二

## (1)

证明. 设 $x, y$ 为任意满足 $x, y \in \{0,1\}^w$ 且 $x \neq y$ 的数。下面考虑计算 $H$ 中满足 $h(x) = h(y)$ 的 $h$ 的个数，即求关于 $a$ 的方程

$$\left\lfloor \frac{ax \bmod 2^w}{2^{w-l}} \right\rfloor = \left\lfloor \frac{ay \bmod 2^w}{2^{w-l}} \right\rfloor \tag{1}$$

的满足 $a \in \{0,1\}^w$ 且 $a$ 为奇数的解的个数。

注意到，方程(1)的解的个数不超过一系列方程

$$ax \bmod 2^w = ay \bmod 2^w + k \tag{2}$$

的解的个数之和，其中 $k \in \left[-\left(2^{w-l}-1\right), 2^{w-l}-1\right]$。进一步地，这些方程(2)的解的个数之和不超过一系列方程

$$a(x-y) \bmod 2^w = t \tag{3}$$

的解的个数之和的两倍，其中 $t \in \{0,1\}^{w-l}$。

不妨设 $x > y$，并设 $x - y = 2^r \cdot s$，其中 $2 \nmid s$。由于 $a$ 是奇数，且 $r < w$，故若 $2^r \nmid t$，则方程(3)无解，否则其等价于

$$as \bmod 2^{w-r} = \frac{t}{2^r}.$$

注意到 $t$ 的范围为 $\{0,1\}^{w-l}$, 若 $w-l<r$, 则方程解的个数为 0; 否则, 由于奇数在模 $2^w$ 意义下存在唯一乘法逆元，故对于一个 $\frac{t}{2^r} \in \{0,1\}^{w-l-r}$, 至多只有一个奇数 $a \in \{0,1\}^{w-r}$ 满足方程(3)，从而至多只有 $2^r$ 个奇数 $a \in \{0,1\}^w$ 满足方程(3)，故对所有 $t$，方程(3)解的个数之和不超过 $2^r \cdot 2^{w-l-r} = 2^{w-l}$。

注意到 $|H| = 2^{w-1}$，哈希表的大小为 $m = 2^l$，故对任意 $x, y \in \{0,1\}^w$ 且 $x \neq y$，冲突概率不超过 $\frac{2 \cdot 2^{w-l}}{2^{w-1}} = \frac{4}{2^l} = \frac{4}{m}$，故 $H$ 是全域的。

$\square$

### (2)

本题给出的哈希函数族计算更高效，因为对 $2^w$ 取模以及除以 $2^{w-l}$ 再取下取整的运算均可通过位运算实现。