1.a: Lemma. 若数列 $a_1 \le b_1, ..., a_n \le b_n$，则两数列 a, b 升序排序后仍有 $a_{(1)} \le b_{(1)}, ..., a_{(n)} \le b_{(n)}$.

证明：考虑倒序取出排序后 b 中的元素，即从 n 到 1 依次考虑 $b_{(i)}$。反设 $a_{(i)} > b_{(i)}$，则可知 a 中至少有 $n-i+1$ 个数 $> b_{(i)}$，但由于 $a_i < b_i$ ，故 b 中也至少有 $n-i+1$ 个数 $> b_{(i)}$，与 $b_{(i)}$ 的定义矛盾（至多有 $n-i$ 个元素 $> b_{(i)}$）。故 $a_{(i)} \le b_{(i)}$ 恒成立。Q.E.D.

调用完 InsertionSort(h_2) 后，有 $A[k] \le A[k+h_2]$ 成立。

下证在调用 InsertionSort(h_1) 后，在新的序列 A' 中，$A'[k] \le A'[k+h_2]$ 仍成立。

对于当前的 k，取出 k 所在的 h_1 序列记为 $B[t] = \{i | 1 \le i \le n, i = k + t \cdot h_1, t \in \mathbb{Z}\}$，将 $k+h_2$ 所在的 h_1 序列记为 $C[t] = \{i | 1 \le i \le n, i = k + h_2 + t \cdot h_1, t \in \mathrm{Z}\}$。

记 $I_1[t] = \{i | 1 \le i \le n - h_2, i = k + t \cdot h_1, t \in \mathrm{Z}\}, I_2[t] = \{i + h_2 | i \in I_1\}$。可知 I_1 是 $B[t]$ 从开头开始向后的连续一段，而 I_2 是 $C[t]$ 从末尾开始向前的连续一段。

将 A 中以 I_1 为下标的子序列排序后记为 I_1'，I_2 对应下标的子序列排序后记为 I_2'，由 $A[i = k + th_1] \le A[i + h_2 = k + h_2 + th_1]$，由 Lemma 可知 I_1' 和 I_2' 满足 $I_1'[t] \le I_2'[t]$。

将 A 中以 B 为下标的子序列执行完 InsertionSort(h_1) 记为 B'，若 $B' - I_1'$ 在 A 中对应的元素均大于等于 $I_1'[0]$，则 $B[0] = B'[0]$ ，若其中有元素小于 $B'[0]$，则 $B[k] \le B'[0]$，故恒有 $A'[k] = B'[0] \le I_1'[0]$ 成立。同理有 $I_2'[0] \le C'[0] = A'[k + h_2]$。

从而执行 InsertionSort(h_1) 之后 $A'[k] = B'[0] \le I_1'[0] \le I_2'[0] \le C'[0] = A'[k + h_2]$。

1.b: 假设 $\{h_t\}$ ~~递减且~~或两两互质，否则容易构造反例。

Lemma. 当 a 和 b 互质，$c \ge ab$ 时，$ax + by = c$ 恒有非负整数解。

证明：不妨设 $a < b$，由裴蜀定理可知 $c = ax + by$ 存在整数解 (x_0, y_0) 满足 $0 \le x \le b - 1$，又 $c \ge ab$ 可知 $y_0 \ge 0$。Q.E.D.

由 (a) 可知，当执行完 InsertionSort(h_{t+1}) 和 InsertionSort(h_t) 后，有 $A[i] \le A[i + h_1]$ 和 $A[i] \le A[i + h_2]$ 恒成立。

由于 h_{t+1}, h_t 互质，则对任意 $c \ge h_t h_{t+1}$ 由引理可得存在非负整数解 (x, y) 使得 $h_{t+1}x + h_t y = c$，进而由归纳可得 $A[i] \le A[i + c]$ 恒成立。故在执行 InsertionSort(h_{t-1}) 时，任意位置 i 在插入排序的过程中，其对应的元素最多被往前移动到 $i - h_{t+1}h_t$ 的位置，进而移动次数至多为 $O(\frac{h_t h_{t+1}}{h_{t-1}})$。进而总移动次数为 $O(\frac{n h_t h_{t+1}}{h_{t-1}})$。

2.a:

$$T(n) = \frac{1}{n} \sum_{i=1}^{n} (T(i-1) + T(n-i) + O(n))$$

2.b:

设快速排序过程中存在常数 c 使得处理过程的时间复杂度 $O(n) \le cn$，以下归纳证明存在常数 $a > 2c \ln 2$ 使得 $T(n) \le an \log n$。

当 $n = 1$ 时，$T(1) = 0 \le 0$;

当 $n > 1$ 时，

$$
\begin{aligned}
T(n) &\leq cn + \frac{2}{n}a\sum_{i=1}^{n-1} i\log i \\
&\leq cn + \frac{2}{n}a\int_2^n x\log x\mathrm{dx} \\
&= cn + \frac{2}{n}a\int_2^n \mathrm{d}(\frac{1}{2}x^2\log x - \frac{x^2}{4\ln 2}) \\
&= cn + an\log n - \frac{an}{2\ln 2} - O(\frac{1}{n}) \\
&\leq an\log n \\
&= O(n\log n)
\end{aligned}
$$

2.c: 在快速排序的过程中，每次往 k 对应的那边找（$k < pivot$ 就往左，否则往右，$k = pivot$ 则已找到）。

不妨假设 $k < \frac{n}{2}$，由于元素 i, j 可能被比较的情况为：

- $i < k < j$ 且 i 或 j 在 $[i, j]$ 中被先选到；
- $i < j < k$ 且 i 或 j 要在 $[i, k]$ 之间先被选到
- $k < i < j$ 且 i 或 j 要在 $[k, j]$ 之间先被选到

故

$$
\begin{aligned}
E[\#\text{ of compares}] &= \sum_{i<k<j}\frac{2}{j-i+1} + \sum_{i<j<k}\frac{2}{k-i+1} + \sum_{k<i<j}\frac{2}{j-k+1} \\
&= \sum_{len=3}^{n}\frac{2\min(len-2, k-1, n-len)}{len} + \sum_{i<k}\frac{2(k-i-1)}{k-i+1} + \sum_{k<j}\frac{2(j-k-1)}{j-k+1} \\
&\leq 2n + 2n + 2n \\
&= O(n)
\end{aligned}
$$

3.a: 确定性的排序算法可以用决策树表示，对于深度为 n 的决策树，叶子数量为 2^n，而高度小于 $0.5n\log n$ 的叶子数量至多为 $2^{0.5n\log n} = n^{0.5n}$，故当 n 足够大时，满足运行时间小于 $0.5n\log n$ 的排列数量至多为

$$
\frac{n^{0.5n}}{n!} \leq \frac{n^{0.5n}\cdot e^n}{n^n} \leq \frac{e^n}{\sqrt{n}^n} < 0.01
$$

3.b: 假设命题不成立，即超过 50% 的排列 σ 满足：

$$
E_r[\mathrm{Time}(A(\sigma, r))] < 0.4n\log n.
$$

定义集合 $S = \{\sigma \mid E_r[\mathrm{Time}(A(\sigma, r))] < 0.4n\log n\}$，则 $|S| > 0.5n!$。

对每个 $\sigma \in S$，至少 50% 的随机串 r 满足（由 *markov* 不等式推出）：

$$\text{Time}(A(\sigma, r)) < 0.8n \log n.$$

因此，定义 “快排对” 为满足 $\text{Time}(A(\sigma, r)) < 0.8n \log n.$ 的 (σ, r) 组合，则 “快排对” (σ, r) 的总数至少为：

$$|S| \cdot 0.5 \cdot 2^{l(n)} > 0.5n! \cdot 0.5 \cdot 2^{l(n)} = \frac{n! \cdot 2^{l(n)}}{4}.$$

对每个固定 r，算法 $A(\cdot, r)$ 的决策树中深度小于 $0.8n \log n$ 的叶节点数至多为：

$$2^{0.8n \log n} = n^{0.8n}.$$

因此，所有 $2^{l(n)}$ 个可能的 r 对应的快排对总数至多为：

$$2^{l(n)} \cdot n^{0.8n}.$$

由假设，应有

$$2^{l(n)} \cdot n^{0.8n} \geq \frac{n! \cdot 2^{l(n)}}{4} \implies n^{0.8n} \geq \frac{n!}{4}.$$

根据斯特林公式 $n! \approx \left(\frac{n}{e}\right)^n$，当 n 充分大时：

$$n^{0.8n} = e^{0.8n \ln n}, \quad \frac{n!}{4} \approx \frac{1}{4}\left(\frac{n}{e}\right)^n = e^{n \ln n - n - \ln 4}.$$

显然 $e^{0.8n \ln n} \ll e^{n \ln n}$，矛盾。

4: 当 $\frac{a_i}{b_i} \neq \frac{a_j}{b_j}$ 时，有 $\left|\frac{a_i}{b_i} - \frac{a_j}{b_j}\right| = \left|\frac{a_i b_j - a_j b_i}{b_i b_j}\right| > \frac{1}{n^4}$，故我们可将所有 $\frac{a_i}{b_i}$ 映射到 val_i，其中 $val_i = argmax\{k | \frac{k}{n^4} \leq \frac{a_i}{b_i}\}$，且 $\frac{a_i}{b_i} < \frac{a_j}{b_j}$ 当且仅当 $val_i < val_j$。

而 $val_i \in [0, n^6]$ 且是整数，可将其视为 6 位 n 进制数并使用基数排序，时间复杂度为 $O(6n) = O(n)$。