多路快速排序的分析

李嘉图 lijt19@mails.tsinghua.edu.cn

引理 1. 对于 $0 < B < n$，有

$$\sum_{1\le i\le B} i\binom{n-i-1}{k-1} = \binom{n}{k+1} - \frac{Bk+n}{k+1}\binom{n-B-1}{k}. \tag{1}$$

证明.

$$\begin{aligned}&\sum_{1\le i\le B} i\binom{n-i-1}{k-1}\\ =&\sum_{1\le i\le B} n\binom{n-i-1}{k-1} - \sum_{1\le i\le B}(n-i)\binom{n-i-1}{k-1}\end{aligned} \tag{2}$$

其中第一项可以使用上指标求和法[1]处理，即

$$\begin{aligned}&\sum_{1\le i\le B} n\binom{n-i-1}{k-1}\\ =&\, n\sum_{1\le n-i-1\le B}\binom{i}{k-1}\\ =&\, n\sum_{n-B-1\le i\le n-2}\binom{i}{k-1}\\ =&\, n\left[\sum_{i\le n-2}\binom{i}{k-1} - \sum_{i\le n-B-2}\binom{i}{k-1}\right]\\ =&\, n\left[\binom{n-1}{k} - \binom{n-B-1}{k}\right].\end{aligned} \tag{3}$$

第二项使用提取、吸收恒等式[1]经过转化后也可以类似处理。

$$\begin{aligned}&\sum_{1\le i\le B}(n-i)\binom{n-i-1}{k-1}\\ =&\sum_{1\le i\le B} k\binom{n-i}{k}\\ =&\, k\sum_{1\le n-i\le B}\binom{i}{k}\\ =&\, k\sum_{n-B\le i\le n-1}\binom{i}{k}\\ =&\, k\left[\sum_{i\le n-1}\binom{i}{k} - \sum_{i\le n-B-1}\binom{i}{k}\right]\\ =&\, k\left[\binom{n}{k+1} - \binom{n-B}{k+1}\right].\end{aligned} \tag{4}$$

考虑将上面两部分带入 (2)，得到

$$
\begin{aligned}
(2) &= (3) - (4) \\
&= n\left[\binom{n-1}{k} - \binom{n-B-1}{k}\right] - k\left[\binom{n}{k+1} - \binom{n-B}{k+1}\right] \\
&= n\left[\binom{n-1}{k} - \binom{n-B-1}{k}\right] - \\
&\quad (k+1)\left[\binom{n}{k+1} - \binom{n-B}{k+1}\right] + \\
&\quad \left[\binom{n}{k+1} - \binom{n-B}{k+1}\right] \\
&= n\left[\binom{n-1}{k} - \binom{n-B-1}{k}\right] - \\
&\quad \left[n\binom{n-1}{k} - (n-B)\binom{n-B-1}{k}\right] + \\
&\quad \left[\binom{n}{k+1} - \binom{n-B}{k+1}\right] \\
&= -B\binom{n-B-1}{k} + \binom{n}{k+1} - \binom{n-B}{k+1} \\
&= \binom{n}{k+1} - \left[B\binom{n-B-1}{k} + \frac{n-B}{k+1}\binom{n-B-1}{k}\right] \\
&= \binom{n}{k+1} - \frac{Bk+n}{k+1}\binom{n-B-1}{k}.
\end{aligned} \tag{5}
$$

□

引理 2. 当 $n > k > 1$ 时，有

$$
\frac{1}{\binom{n}{k}}\left[\sum_{1\le i\le n-2}(n-1-i)i\binom{n-i-2}{k-2} + \sum_{1\le i\le n-1} 2i\binom{n-i-1}{k-1}\right] = n-k. \tag{6}
$$

证明. 考虑下面的组合问题：给定 n 个初始为白色的有标号小球，从中等概率随机的选择 k 个染为黑色，求白色小球的期望个数。很明显答案为 $n-k$，等于上式右侧。下面证明左侧也具有同样的组合意义。

我们称一个染色方案为 C，一个不可向两侧延伸的白色连续段为一个极长白色段，一个染色方案所有极长白色段的集合为 $\mathcal{W}(C)$，那么剩余白球的期望个数为

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{w\in\mathcal{W}(C)}|w|\right] \\
&= \sum_{w}|w|\Pr\{\mathrm{w}\in\mathcal{W}(\mathrm{C})\}.
\end{aligned} \tag{7}
$$

极长白色段的“边界”有两种情况，第一种是两侧都有黑色球（即不在最左侧也不在最右侧），这种情况下每一个连续段作为极长白色段的概率都是 $\binom{n-|w|-2}{k-2}/\binom{n}{k}$；第二种情况是一侧靠边界、另一侧有黑色球，这种情况下每一个连续段作为极长白色段的概率都是 $\binom{n-|w|-1}{k-1}/\binom{n}{k}$，由于可以靠左、右边界，这一项需要乘上 2 的系数。由于 $k > 1$，极长白色段不可能两侧都靠边界。将以上两

种情况求和，我们便得到

$$\frac{1}{\binom{n}{k}}\left[\sum_{1\le i\le n-2}(n-1-i)i\binom{n-i-2}{k-2}+\sum_{1\le i\le n-1}2i\binom{n-i-1}{k-1}\right] \tag{8}$$

这正是等式的左边，因此我们证明了等式成立。 □

引理 3. 不妨设算法拥有 $2k$ 元比较器，则算法的时间复杂度可以表示为

$$T(n)=\frac{1}{\binom{n}{k}}\left[\sum_{0=i_0<i_1<i_2<\cdots<i_d<i_{d+1}=n+1}\ \sum_{1\le j\le d+1}T(i_j-i_{j-1}-1)\right]+O\left(\frac{n}{k}\right). \tag{9}$$

证明. 由于多路快速排序等概率随机的选择 k 个元素作为分割点，期望下每一小段比较次数之和等于每一个划分方案的比较次数乘上这个方案的概率。前者正是枚举所有的划分点 $i_1, i_2, \ldots, i_k$ 并递归处理每一小段，后者为 $1/\binom{n}{k}$。再加上划分所需要的比较次数 $O(n/k)$ 便得到了期望比较次数。

□

引理 4.

$$T(n)=\frac{1}{\binom{n}{k}}\left[\sum_{1\le i\le n-2}(n-1-i)T(i)\binom{n-i-2}{k-2}+\sum_{1\le i\le n-1}2T(i)\binom{n-i-1}{k-1}\right]+O\left(\frac{n}{k}\right). \tag{10}$$

证明. 注意到选择分割点和引理 2 中选择小球染黑具有相同的组合意义，因此用相同的方法将引理 3 中的求和式展开，便可以证明这一引理。 □

引理 5.

$$\binom{n}{k+1}-\frac{Bk+n}{k+1}\binom{n-B-1}{k}=\Theta\left(\binom{n}{k+1}\right) \tag{11}$$

证明. 不妨设

$$\begin{aligned} f(n,k)&=\binom{n}{k+1},\\ g(n,k)&=\frac{Bk+n}{k+1}\binom{n-B-1}{k}. \end{aligned} \tag{12}$$

考察 $n\to+\infty$ 的情形。注意到 $f(n,k)$ 和 $g(n,k)$ 均为 n 的 $k+1$ 次多项式，为此只需证明 $f(n,k)$ 的最高次项系数大于 $g(n,k)$ 的最高次项系数。其中前者为 $1/(k+1)!$，后者为

$$\begin{aligned} &\frac{2}{(k+1)k!}\times\left(\frac{k-1}{k}\right)^k\\ =&\frac{2}{(k+1)!}\left(1-\frac{1}{k}\right)^k\\ \le&\frac{2}{e(k+1)!} \end{aligned} \tag{13}$$

其中 $\frac{2}{e}\approx 0.7358<1$，这便完成了证明。

□

定理 1. 多路快速排序算法的期望比较次数是 $O(n\log n/k\log k)$.

证明. 施归纳于 n，用代入法证明存在常数 c 使得 $T(n) \le cn\log n/k\log k$，即证明

$$\frac{1}{\binom{n}{k}k\log k}\left[\sum_{1\le i\le n-2} ci(n-1-i)\binom{n-i-2}{k-2}\log i + \sum_{1\le i\le n-1} 2ci\binom{n-i-1}{k-1}\log i\right] + \frac{c_1 n}{k} \le \frac{cn\log n}{k\log k}. \tag{14}$$

考虑对 $\log i$ 的项做一些放缩。取 $B = n/k$，将中括号中第一部分的和式分成大小两部分，并分别用 $\log B$ 和 $\log n$ 放缩，利用引理 1，得到

$$\begin{aligned}
&\sum_{1\le i\le n-2} ci(n-1-i)\binom{n-i-2}{k-2}\log i \\
\le\ &\sum_{1\le i\le B} ci(n-1-i)\binom{n-i-2}{k-2}\log\frac{n}{k} + \sum_{B+1\le i\le n-2} ci(n-1-i)\binom{n-i-2}{k-2}\log n \\
=\ &c\log n\left[\sum_{1\le i\le n-2} i(n-1-i)\binom{n-i-2}{k-2}\right] - c\log k\sum_{1\le i\le B} i(n-1-i)\binom{n-i-2}{k-2} \\
=\ &c\log n\left[\sum_{1\le i\le n-2} i(n-1-i)\binom{n-i-2}{k-2}\right] - c\log k\sum_{1\le i\le B} i(k-1)\binom{n-i-1}{k-1} \\
=\ &c\log n\left[\sum_{1\le i\le n-2} i(n-1-i)\binom{n-i-2}{k-2}\right] - c(k-1)\log k\left[\binom{n}{k+1} - \frac{Bk+n}{k+1}\binom{n-B-1}{k}\right].
\end{aligned} \tag{15}$$

根据引理 5，右边中括号可以用 $c_2\binom{n}{k+1}$ 代替。

$$c\log n\left[\sum_{1\le i\le n-2} i(n-1-i)\binom{n-i-2}{k-2}\right] - cc_2(k-1)\binom{n}{k+1}\log k. \tag{16}$$

对于式 (14) 中括号中第二项，我们用 $\log n$ 放缩 $\log i$，(14) 中括号中的项的上界是

$$\begin{aligned}
&c\log n\left[\sum_{1\le i\le n-2} i(n-1-i)\binom{n-i-2}{k-2} + \sum_{1\le i\le n-1} 2i\binom{n-i-1}{k-1}\right] - cc_2(k-1)\binom{n}{k+1}\log k \\
\le\ &c(n-k)\binom{n}{k}\log n - cc_2(k-1)\binom{n}{k+1}\log k.
\end{aligned} \tag{17}$$

将这个上界带入 (14) 中，得到

$$\begin{aligned}
T(n) &\le \frac{1}{\binom{n}{k}k\log k}\left[c(n-k)\binom{n}{k}\log n - cc_2(k-1)\binom{n}{k+1}\log k\right] + \frac{c_1 n}{k} \\
&\le \frac{cn\log n}{k\log k} + \left[\frac{c_1 n}{k} - \frac{cc_2(k-1)\log k}{\binom{n}{k}k\log k}\binom{n}{k+1}\right] \\
&= \frac{cn\log n}{k\log k} + \left[\frac{c_1 n}{k} - cc_2\times\frac{(k-1)(n-k)\log k}{k^2\log k}\right] \\
&\le \frac{cn\log n}{k\log k} + \left[\frac{c_1 n}{k} - cc_2c_3\times\frac{n}{k}\right] \\
&\le \frac{cn\log n}{k\log k},
\end{aligned} \tag{18}$$

只需取 $c = \frac{c_1}{c_2c_3}$ 便可以完成归纳。 □

Acknowledgement: 感谢范致远、王之栋同学帮助计算了引理 1，感谢吴梦迪同学发现并改正了定理证明中的错误，感谢胡子晗同学帮助验算了引理 2。

参考文献

[1] Edward A Bender, R L Graham, Donald E Knuth, and Oren Patashnik. *Concrete Mathematics: A Foundation for Computer Science, Second Edition.* Post & Telecom Press, 2013.