クィックソートの平均操作数を求める

02_PS_quicksort.pdf
真貝

教科書 [1]p37 コラム 6 で紹介した，クィックソートについて，n 個のものをクィックソートするときの平均操作数 Q_n を求めよう．

n 個のものに対して，初めの 1 つを基準に，残りの $n-1$ 個を 2 つに分けてゆく．残りが s 個と $n-1-s$ 個に分けられるとすると，

$$Q_n = n - 1 + Q_s + Q_{n-1-s} \tag{1}$$

と書くことができる．2 つの分けかたは，$\{Q_0, Q_{n-1}\}, \{Q_1, Q_{n-2}\}, \cdots, \{Q_{n-1}, Q_0\}$ の n パターンあって，そのどれもが等確率 $1/n$ で得られるとすれば，平均操作数 Q_n は

$$Q_n = n - 1 + \frac{1}{n}\sum_{s=0}^{n-1}(Q_s + Q_{n-1-s}) = n - 1 + \frac{2}{n}\sum_{s=0}^{n-1} Q_s \tag{2}$$

添え字を 1 減らしたものについては

$$Q_{n-1} = n - 2 + \frac{2}{n-1}\sum_{s=0}^{n-2} Q_s \tag{3}$$

となるので，$n\times(2) - (n-1)\times(3)$ より

$$\begin{aligned} nQ_n - (n-1)Q_{n-1} &= n(n-1) - (n-1)(n-2) - 2Q_{n-1} \\ nQ_n - (n+1)Q_{n-1} &= 2(n-1) \end{aligned}$$

となる．両辺を $n(n+1)$ で割ると，

$$\frac{Q_n}{n+1} - \frac{Q_{n-1}}{n} = \frac{2(n-1)}{n(n+1)} \tag{4}$$

となる．ここで，$P_n \equiv \dfrac{Q_n}{n+1}$ と新たに置くと，この式は

$$P_n - P_{n-1} = \frac{2(n-1)}{n(n+1)} \tag{5}$$

となって，P_n についての漸化式になる．なお，$P_0 = Q_0 = 0$ である．(5) の上限値を考えて，

$$P_n - P_{n-1} = \frac{2(n-1)}{n(n+1)} < \frac{2n}{n(n+1)} = \frac{2}{n+1} \tag{6}$$

として，1 つずつ添え字を減らした式を並べ，

$$\begin{aligned} P_n - P_{n-1} &< \frac{2}{(n+1)} \\ P_{n-1} - P_{n-2} &< \frac{2}{n} \\ &\vdots \qquad \vdots \\ P_1 - P_0 &< 2 \end{aligned}$$

これらの和をとると，

$$P_n - P_0 < 2\sum_{k=1}^{n+1}\frac{1}{k} \tag{7}$$

となる．右辺の和の式は，n を大きくすれば積分で書けて

$$\sum_{k=1}^{n+1}\frac{1}{k} = \int_1^{n+1}\frac{1}{x}dx = [\log x]_{x=1}^{x=n+1} = \log(n+1) \tag{8}$$

となることから，(7) は，

$$P_n - 0 < 2\log(n+1) \tag{9}$$

すなわち

$$\frac{Q_n}{n+1} < 2\log(n+1)$$

となる．したがって，

$$Q_n < 2(n+1)\log(n+1)$$

となって，

$$Q_n \sim n\log n \tag{10}$$

となる．

参考文献

[1] 真貝寿明，徹底攻略確率統計（共立出版，2012）