Average number of compares for QUICKSORT with distinct keys

Recurrence from recursive program

$$C_N = N - 1 + \frac{1}{N} \sum_{1 \le j \le N} (C_{j-1} + C_{N-j})$$

Change j to $N+1-j$ in second sum

$$C_N = N - 1 + \frac{2}{N} \sum_{1 \le j \le N} C_{j-1}.$$

Multiply both sides by N

$$NC_N = N(N-1) + \frac{2}{N} \sum_{1 \le j \le N} C_{j-1}.$$

Subtract same equation for $N-1$

$$NC_N - (N-1)C_{N-1} = 2N + 2C_{N-1}$$

Rearrange terms

$$NC_N = (N+1)C_{N-1} + 2N$$

Divide by $N(N+1)$

$$\frac{C_N}{N+1} = \frac{C_{N-1}}{N} + \frac{2}{N+1}$$

Telescope

$$\frac{C_N}{N+1} = 2(H_{N+1} - 1)$$

Approximate

$$C_N \approx 2N \ln N$$

Average number of compares for QUICKSORT with equal keys

Recurrence for average number of comparisons

$$C(x_1,\ldots,x_n) = N + 1 + \frac{1}{N}\sum_{1\le j\le N} x_j(C(x_1,\ldots,x_{j-1}) + C(x_{j+1}\ldots x_n))$$

Multiply both sides by $N = x_1 + \ldots + x_n$

$$NC(x_1,\ldots,x_n) = N(N-1) + \sum_{1\le j\le N} x_j C(x_1,\ldots,x_{j-1}) + \sum_{1\le j\le N} x_j C(x_{j+1},\ldots,x_n).$$

Subtract same equation for $x_2,\ldots,x_n$ **(with** $D(x_1\ldots x_n) \equiv C(x_1,\ldots,x_n) - C(x_2,\ldots,x_n)$**)**

$$(x_1 + \ldots + x_n)D(x_1\ldots,x_n) = x_1^2 - x_1 + 2x_1(x_2 + \ldots + x_n) + \sum_{2\le j\le n} x_j D(x_1,\ldots,x_{j-1})$$

Subtract same equation for $x_1,\ldots,x_{n-1}$

$$(x_1 + \ldots + x_n)D(x_1,\ldots,x_n) - (x_1 + \ldots + x_{n-1})D(x_1,\ldots,x_{n-1}) = 2x_1x_n + x_nD(x_1,\ldots,x_{n-1})$$

Simplify, divide by N

$$D(x_1,\ldots,x_n) = D(x_1\ldots,x_{n-1}) + \frac{2x_1x_n}{x_1 + \ldots + x_n}$$

Telescope (twice)

$$C(x_1,\ldots,x_n) = N - n + 2\sum_{1\le k\le j\le n} \frac{x_kx_j}{x_k + \ldots + x_j}$$

Upper bound on QUICKSORT entropy

Quicksort entropy definition

$$Q = \sum_{1 \le k < j \le n} \frac{p_k p_j}{p_k + \ldots + p_j}$$

Separate double sum

$$Q = \sum_{1 \le k < n} p_k \sum_{k < j \le n} \frac{p_j}{p_k + \ldots + p_j}$$

Substitute $q_{ij} = (p_i + \ldots + p_j / p_i)$ **(note:** $1 = q_{ii} \le q_{i(i+1)} \le \ldots \le q_{in} < 1/p_i$**)**

$$Q = \sum_{1 \le k < N} p_k \sum_{k < j \le n} \frac{q_{kj} - q_{k(j-1)}}{q_{kj}}$$

Bound with integral

$$Q < \sum_{1 \le k < n} p_k \int_{q_{kk}}^{q_{kn}} \frac{1}{x} dx$$

Simplify

$$Q < \sum_{1 \le k < n} p_k \ln q_{kn} \le \sum_{1 \le k < n} p_k (-\ln p_k) = H \ln 2$$