The average number of comparisons for quicksort is

$$S(n) = \begin{cases} 0 & \text{if } n = 0, 1 \\ \sum_{q=1}^{n} \frac{1}{n}[S(q-1) + S(n-q)] + n - 1 & \text{otherwise} \end{cases}$$

The average number of comparisons for quicksort is

$$S(n) = \begin{cases} 0 & \text{if } n = 0, 1 \\ \sum_{q=1}^{n} \frac{1}{n}[S(q-1) + S(n-q)] + n - 1 & \text{otherwise} \end{cases}$$

Theorem

$S(n) \leq an \lg n$ for some constant a and $n \geq 1$.

Proof by Constructive Induction.

Base case: $n = 1$: $S(1) = 0$ and $a \cdot 1 \cdot \lg 1 = 0$.

Induction Hypothesis:

Assume it holds for all positive integers less than n.
So, $S(k) \leq ak \lg k$ for $1 \leq k \leq n - 1$.

Induction step:

$$S(n) = \sum_{q=1}^{n} \frac{1}{n}[S(q-1) + S(n-q)] + n - 1$$

$$
\begin{aligned}
S(n) &= \sum_{q=1}^{n} \frac{1}{n}[S(q-1) + S(n-q)] + n - 1 \\
&= \frac{1}{n} \sum_{q=1}^{n} [S(q-1) + S(n-q)] + n - 1
\end{aligned}
$$

$$
\begin{aligned}
S(n) &= \sum_{q=1}^{n} \frac{1}{n}[S(q-1) + S(n-q)] + n - 1 \\
&= \frac{1}{n}\sum_{q=1}^{n}[S(q-1) + S(n-q)] + n - 1 \\
&= \frac{1}{n}\sum_{q=1}^{n} S(q-1) + \frac{1}{n}\sum_{q=1}^{n} S(n-q) + n - 1
\end{aligned}
$$

$$
\begin{aligned}
S(n) &= \sum_{q=1}^{n} \frac{1}{n}[S(q-1) + S(n-q)] + n - 1 \\
&= \frac{1}{n}\sum_{q=1}^{n}[S(q-1) + S(n-q)] + n - 1 \\
&= \frac{1}{n}\sum_{q=1}^{n} S(q-1) + \frac{1}{n}\sum_{q=1}^{n} S(n-q) + n - 1 \\
&= \frac{1}{n}\sum_{q=0}^{n-1} S(q) + \frac{1}{n}\sum_{q=0}^{n-1} S(q) + n - 1
\end{aligned}
$$

$$
\begin{aligned}
S(n) &= \sum_{q=1}^{n} \frac{1}{n}[S(q-1) + S(n-q)] + n - 1 \\
&= \frac{1}{n}\sum_{q=1}^{n}[S(q-1) + S(n-q)] + n - 1 \\
&= \frac{1}{n}\sum_{q=1}^{n} S(q-1) + \frac{1}{n}\sum_{q=1}^{n} S(n-q) + n - 1 \\
&= \frac{1}{n}\sum_{q=0}^{n-1} S(q) + \frac{1}{n}\sum_{q=0}^{n-1} S(q) + n - 1 \\
&= \frac{2}{n}\sum_{q=0}^{n-1} S(q) + n - 1 \;=\; \frac{2}{n}\sum_{q=1}^{n-1} S(q) + n - 1 \qquad \text{since } S(0) = 0
\end{aligned}
$$

$$
\begin{aligned}
S(n) &= \sum_{q=1}^{n} \frac{1}{n}[S(q-1)+S(n-q)] + n - 1 \\
&= \frac{1}{n}\sum_{q=1}^{n}[S(q-1)+S(n-q)] + n - 1 \\
&= \frac{1}{n}\sum_{q=1}^{n} S(q-1) + \frac{1}{n}\sum_{q=1}^{n} S(n-q) + n - 1 \\
&= \frac{1}{n}\sum_{q=0}^{n-1} S(q) + \frac{1}{n}\sum_{q=0}^{n-1} S(q) + n - 1 \\
&= \frac{2}{n}\sum_{q=0}^{n-1} S(q) + n - 1 \;=\; \frac{2}{n}\sum_{q=1}^{n-1} S(q) + n - 1 \qquad \text{since } S(0) = 0 \\
&\leq \frac{2}{n}\sum_{q=1}^{n-1} aq\lg q + n - 1 \qquad \text{by IH}
\end{aligned}
$$

$$= \frac{2a}{n} \sum_{q=1}^{n-1} q \lg q + n - 1$$

$$
\begin{aligned}
&= \frac{2a}{n} \sum_{q=1}^{n-1} q \lg q + n - 1 \\
&\leq \frac{2a}{n} \int_{1}^{n} x \lg x dx + n - 1 \qquad \text{by integral bound}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{2a}{n}\sum_{q=1}^{n-1} q\lg q + n - 1 \\
&\leq \frac{2a}{n}\int_{1}^{n} x\lg x dx + n - 1 \qquad \text{by integral bound} \\
&= \frac{2a}{n}\left[\frac{x^2\lg x}{2} - \frac{x^2\lg e}{4}\right]\Big|_1^n + n - 1
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{2a}{n}\sum_{q=1}^{n-1} q \lg q + n - 1 \\
&\le \frac{2a}{n}\int_{1}^{n} x \lg x dx + n - 1 \qquad \text{by integral bound} \\
&= \frac{2a}{n}\left[\frac{x^2 \lg x}{2} - \frac{x^2 \lg e}{4}\right] \Big|_1^n + n - 1 \\
&= \frac{2a}{n}\left[\left(\frac{n^2 \lg n}{2} - \frac{n^2 \lg e}{4}\right) - \left(\frac{1^2 \lg 1}{2} - \frac{1^2 \lg e}{4}\right)\right] + n - 1
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{2a}{n}\sum_{q=1}^{n-1} q \lg q + n - 1 \\
&\leq \frac{2a}{n}\int_{1}^{n} x \lg x dx + n - 1 \qquad \text{by integral bound} \\
&= \frac{2a}{n}\left[\frac{x^2 \lg x}{2} - \frac{x^2 \lg e}{4}\right]\Big|_1^n + n - 1 \\
&= \frac{2a}{n}\left[\left(\frac{n^2 \lg n}{2} - \frac{n^2 \lg e}{4}\right) - \left(\frac{1^2 \lg 1}{2} - \frac{1^2 \lg e}{4}\right)\right] + n - 1 \\
&= an \lg n - \frac{an \lg e}{2} + \frac{a \lg e}{2n} + n - 1
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{2a}{n}\sum_{q=1}^{n-1} q \lg q + n - 1 \\
&\leq \frac{2a}{n}\int_{1}^{n} x \lg x dx + n - 1 \qquad \text{by integral bound} \\
&= \frac{2a}{n}\left[\frac{x^2 \lg x}{2} - \frac{x^2 \lg e}{4}\right] \Big|_1^n + n - 1 \\
&= \frac{2a}{n}\left[\left(\frac{n^2 \lg n}{2} - \frac{n^2 \lg e}{4}\right) - \left(\frac{1^2 \lg 1}{2} - \frac{1^2 \lg e}{4}\right)\right] + n - 1 \\
&= an \lg n - \frac{an \lg e}{2} + \frac{a \lg e}{2n} + n - 1 \\
&= an \lg n + \left[1 - \frac{a \lg e}{2}\right] n + \frac{a \lg e}{2n} - 1
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{2a}{n}\sum_{q=1}^{n-1} q\lg q + n - 1 \\
&\le \frac{2a}{n}\int_{1}^{n} x\lg x dx + n - 1 \qquad \text{by integral bound} \\
&= \frac{2a}{n}\left[\frac{x^2\lg x}{2} - \frac{x^2\lg e}{4}\right] \Big|_1^n + n - 1 \\
&= \frac{2a}{n}\left[\left(\frac{n^2\lg n}{2} - \frac{n^2\lg e}{4}\right) - \left(\frac{1^2\lg 1}{2} - \frac{1^2\lg e}{4}\right)\right] + n - 1 \\
&= an\lg n - \frac{an\lg e}{2} + \frac{a\lg e}{2n} + n - 1 \\
&= an\lg n + \left[1 - \frac{a\lg e}{2}\right] n + \frac{a\lg e}{2n} - 1 \\
&\le an\lg n \qquad \text{for the induction to hold}
\end{aligned}
$$

Need

$$an \lg n + \left[1 - \frac{a \lg e}{2}\right] n + \frac{a \lg e}{2n} - 1 \ \leq \ an \lg n$$

Need

$$an \lg n + \left[1 - \frac{a \lg e}{2}\right] n + \frac{a \lg e}{2n} - 1 \;\leq\; an \lg n$$

Need

$$1 - \frac{a \lg e}{2} \leq 0 \quad \Longrightarrow \quad a \geq \frac{2}{\lg e}$$

Set $a = \frac{2}{\lg e} \approx 1.39$

Need

$$an \lg n + \left[1 - \frac{a \lg e}{2}\right] n + \frac{a \lg e}{2n} - 1 \;\leq\; an \lg n$$

Need

$$1 - \frac{a \lg e}{2} \leq 0 \quad \Longrightarrow \quad a \geq \frac{2}{\lg e}$$

Set $a = \frac{2}{\lg e} \approx 1.39$

Need

$$\frac{a \lg e}{2n} - 1 \leq 0 \quad \Longleftrightarrow \quad \frac{2 \lg e}{(\lg e) 2n} - 1 \leq 0 \quad \Longleftrightarrow \quad \frac{1}{n} - 1 \leq 0$$

which always holds since $n \geq 1$.

Need

$$an\lg n + \left[1 - \frac{a\lg e}{2}\right] n + \frac{a\lg e}{2n} - 1 \;\le\; an\lg n$$

Need

$$1 - \frac{a\lg e}{2} \le 0 \quad \Longrightarrow \quad a \ge \frac{2}{\lg e}$$

Set $a = \frac{2}{\lg e} \approx 1.39$

Need

$$\frac{a\lg e}{2n} - 1 \le 0 \quad \Longleftrightarrow \quad \frac{2\lg e}{(\lg e)2n} - 1 \le 0 \quad \Longleftrightarrow \quad \frac{1}{n} - 1 \le 0$$

which always holds since $n \ge 1$.

So,

$$S(n) \;\approx\; 1.39n\lg n$$

Now that we are finished, we realize that

$$S(n) \;\leq\; an \lg n \;=\; \frac{2n \lg n}{\lg e} \;=\; 2n \ln n$$

which is a more natural formula.

Now that we are finished, we realize that

$$S(n) \; \leq \; an \lg n \; = \; \frac{2n \lg n}{\lg e} \; = \; 2n \ln n$$

which is a more natural formula.

Could go back and do the Constructive Induction with $S(n) \leq an \ln n$, which would simplify the algebra.