Intuitivamente il caso pessimo per il quicksort si verifica quando il vettore è composto da elementi distinti già ordinati in modo crescente o decrescente, in entrambi i casi infatti la ricorrenza diviene:

$$T(n) = T(n-1) + \Theta(n) \quad se\ n \geq 2$$
$$= \Theta(1) \quad \text{altrimenti}$$

che si risolve facilmente in $\Theta(n^2)$.

Per dimostrare che la nostra intuizione è giusta basta dimostrare che per la seguente equazione vale $T(n) = O(n^2)$

$$T(n) = max_{0 \leq k \leq n-1} \{T(k) + T(n-1-k)\} + \Theta(n) \quad se\ n \geq 2$$
$$= \Theta(1) \quad \text{altrimenti}$$

Risolviamo di seguito l'equazione col metodo di sostituzione.

Considera l'equazione

$$T(n) = max_{0\le k\le n-1}\{T(k)+T(n-1-k)\}+bn \quad se\ n\ge 2$$
$$= a \qquad \text{altrimenti}$$

dove a e b sono costanti fissate e positive.

Dimostriamo col metodo di sostituzione che $T(n) \le cn^2 + c$ dove $c = \max\{a, b\}$.

Per i casi base $n \le 1$ l'ipotesi è certamente vera.
Assumiamola vera per $k < n$ e per ipotesi induttiva abbiamo:

$$T(n) \le max_{0\le k\le n-1}\left\{ck^2+c+c(n-1-k)^2+c\right\}+cn$$
$$= c\cdot max_{0\le k\le n-1}\left\{k^2+(n-1-k)^2+2\right\}+cn$$
$$\le c\cdot max_{0\le k\le n-1}\left\{(k+n-1-k)^2+2\right\}+cn \text{ uso } x^2+y^2<(x+y)^2$$
$$= c\cdot max_{0\le k\le n-1}\left\{(n-1)^2+2\right\}+cn$$
$$= c\cdot\left\{n^2-2n+4\right\}+cn$$
$$= cn^2+c-cn+3c$$
$$\le cn^2+c \qquad \text{poiché } -cn+c<0$$

intuitivamente il caso migliore del quicksort si verifica quando ad ogni passo, la dimensione dei due sotto-problemi è identica.

in questo caso l'equazione di ricorrenza diventa:

$$T(n) = 2T\left(\frac{n}{2}\right) + \Theta(n)$$

che si risolve facilmente in $T(n) = \Theta(n \log n)$

Per dimostrare che la nostra intuizione è giusta dobbiamo dimostrare che per la seguente equazione vale $T(n) = \Omega(n \log n)$

$$\begin{aligned} T(n) &= min_{0 \leq k \leq n-1} \{T(k) + T(n-1-k)\} + \Theta(n) && \textit{se } n \geq 2 \\ &= \Theta(1) && \text{altrimenti} \end{aligned}$$

l'equazione possiamo risolverla col metodo di sostituzione.

Considera l'equazione

$$T(n) = min_{0 \le k \le n-1} \{T(k) + T(n-1-k)\} + bn \quad se\ n \ge 2$$
$$= a \qquad \text{altrimenti}$$

dove a e b sono fissate costanti positive.

Dimostriamo col metodo di sostituzione che $T(n) \ge cn \log_e n$.
Per i casi base $n \le 1$ l'ipotesi è certamente soddisfatta
Assumiamola vera per $k < n$ e per ipotesi induttiva abbiamo:

$$T(n) \ge c \cdot \min_{0 \le k \le n-1} \left\{ k \log k + (n-1-k)\log(n-1-k) \right\} + bn$$

$\ge c \cdot f(x') + bn$ dove x' è il minimo che assume la funzione
$f(x) = x \log_e x + (n-1-x)\log_e(n-1-x)$ nell'intervallo $[0, n-1]$
per la derivata prima si ha $f'(x) = log_e x - log_e(n-1-x)$ che si annulla in
$\frac{n-1}{2}$ segue quindi:

$$T(n) \ge c \cdot f\left(\frac{n-1}{2}\right) + bn = c(n-1)\log_e \frac{n-1}{2} + bn \ge c(n-1)\log_e \frac{n}{4} + bn$$

$$= cn \log_e n - 2cn - c \log_e \frac{n}{4} + bn \ge cn \log_e n - 2cn - c\frac{n}{4} + bn \ge cn \log_e n - \frac{9cn}{4} + bn \le cn \log_e n$$

dove l'ultima diseguaglianza vale prendendo $c = \frac{4b}{9}$.

QuickSort nel caso medio

Valutiamo ora il costo computazionale nel caso medio quando gli n elementi da ordinare di A sono tutti distinti.

Lavoreremo sotto l'ipotesi che uno qualunque degli n numeri possa essere scelto come pivot.

Possiamo fare quest'assunzione nel caso in cui il pivot venga scelto in modo random tra gli n elementi di A. In questo caso si parla di **quicksort randomizzato** e per ottenere questo basterà nel codice visto prima far precedere la scelta del pivot al primo posto da queste semplici istruzioni di costo $O(1)$

```
i=random.randint(0,len(A)-1)
A[0],A[i] = A[i],A[0]
```

Questo significa che dalla sottosequenza di dimensione n con uguale probabilità $\frac{1}{n}$ vengono generate da ordinare due sottosequenze di dimensione k e n-1-k rispettivamente con $0 \leq k \leq n-1$. Abbiamo dunque

$$T(n) = \frac{1}{n}\sum_{k=0}^{n-1}\Big(T(k) + T(n-1-k)\Big) + \Theta(n)$$

Da

$$T(n) = \frac{1}{n}\sum_{k=0}^{n-1}\big(T(k) + T(n-1-k)\big) + \Theta(n)$$

notiamo che per ogni valore di q e $0 \leq q < n$ il termine $T(q)$ compare due volte nella sommatoria, la prima quando $k = q$ e la seconda quando $k = n - 1 - q$. Possiamo dunque scrivere

$$T(n) = \frac{2}{n}\sum_{q=0}^{n-1} T(q) + \Theta(n) \quad \text{per } n \geq 2$$

Infine poiché $T(0) = T(1) = \Theta(1)$ inglobando il contributo della sommatoria per $q = 0$ e $q = 1$ in $\Theta(n)$ otteniamo che la ricorrenza da studiare per il caso medio è

$$T(n) = \frac{2}{n}\sum_{q=2}^{n-1} T(q) + \Theta(n) \quad \text{per } n \geq 2$$

$$= \Theta(1) \text{ altrimenti}$$

Per risolvere la ricorrenza utilizziamo il metodo di sostituzione, e quindi eliminiamo per prima cosa la notazione asintotica:

$$T(n) = \frac{2}{n}\sum_{q=2}^{n-1} T(q) + \Theta(n) \quad \text{per } n \geq 2$$

$$= \Theta(1) \quad \text{altrimenti}$$

$$T(n) = \frac{2}{n}\sum_{q=2}^{n-1} T(q) + a \cdot n \quad \text{per } n \geq 2$$

$$= k \quad \text{altrimenti}$$

Ipotizziamo ora la soluzione $T(n) = O(n \log n)$. Per provarlo basterà dimostrare che esiste una costante c per cui si ha:

$$T(n) \leq c \cdot n \log_2 n \quad \text{per } n \geq 2$$

Sostituiamo la soluzione innanzi tutto nel caso base.

$$T(2) = \frac{2}{2}\sum_{q=2}^{1} T(q) + 2a = 2a$$

deve quindi aversi:

$$T(2) = 2a \;\; \leq c \cdot 2\log_2 2 = 2 \cdot c$$

che è vera per c opportunamente grande (basta ad esempio $c \geq a$).

Per il passo induttivo possiamo scrivere:

$$
\begin{aligned}
T(n) &= \frac{2}{n}\sum_{q=2}^{n-1} T(q) + an \\
&\le \frac{2}{n}\sum_{q=2}^{n-1} c\cdot q\cdot \log_2 q + an \\
&= \frac{2c}{n}\sum_{q=2}^{n-1} q\cdot \log_2 q + an
\end{aligned}
$$

dimostreremo che che: $$\sum_{q=1}^{n-1} q\cdot\log_2 q \le \frac{n^2}{2}\log_2 n - \frac{n^2}{8}$$

possiamo scrivere:

$$
\begin{aligned}
T(n) &\le \frac{2c}{n}\left(\frac{n^2}{2}\log_2 n - \frac{n^2}{8}\right) + a\cdot n \\
&= c\cdot n\cdot \log n - \frac{cn}{4} + a\cdot n \\
&\le c\cdot n\cdot \log n
\end{aligned}
$$

dove l'ultima diseguaglianza segue se prendiamo $c \ge 4a$ in modo da avere $-\frac{c\cdot n}{4} + a\cdot n \le 0.$

Resta solo da dimostrare che $\sum_{q=1}^{n-1} q \cdot \log_2 q \le \frac{n^2}{2}\log_2 n - \frac{n^2}{8}$.

Valutiamo ora la sommatoria $\sum_{q=1}^{n-1} q \log q$, *spezzandola in due nel punto* $\left\lceil \frac{n}{2} \right\rceil - 1$ *:*

$$\sum_{q=1}^{n-1} q \cdot \log q = \sum_{q=1}^{\lceil \frac{n}{2} \rceil - 1} q \cdot \underbrace{\log q}_{\le \log \frac{n}{2} = \log n - 1} + \sum_{q=\lceil \frac{n}{2} \rceil}^{n-1} q \cdot \underbrace{\log q}_{\le \log n}$$

Abbiamo dunque

$$\sum_{q=1}^{n-1} q \cdot \log q \le \sum_{q=1}^{\lceil \frac{n}{2} \rceil - 1} q \cdot (\log n - 1) + \sum_{q=\lceil \frac{n}{2} \rceil}^{n-1} q \cdot \log n$$

Dunque possiamo scrivere:

$$\sum_{q=1}^{n-1} q \cdot \log q \;\leq \sum_{q=1}^{\lceil \frac{n}{2} \rceil -1} q \cdot (\log n - 1) + \sum_{q=\lceil \frac{n}{2} \rceil}^{n-1} q \cdot \log n$$

$$= \sum_{q=1}^{\lceil \frac{n}{2} \rceil -1} q \log n - \sum_{q=1}^{\lceil \frac{n}{2} \rceil -1} q + \sum_{q=\lceil \frac{n}{2} \rceil}^{n-1} q \log n$$

$$= \log n \sum_{q=1}^{n-1} q - \sum_{q=1}^{\lceil \frac{n}{2} \rceil -1} q$$

Ora valutiamo le due sommatorie,

$$\log n \sum_{q=1}^{n-1} q - \sum_{q=1}^{\lceil \frac{n}{2} \rceil - 1} q = \frac{(n-1)n}{2} \log_2 n - \frac{\left(\left\lceil \frac{n}{2} \right\rceil - 1 \right) \left\lceil \frac{n}{2} \right\rceil}{2}$$

$$\leq \log_2 n \frac{(n-1)n}{2} - \frac{\left(\frac{n}{2} - 1 \right) \frac{n}{2}}{2}$$

$$= \frac{n^2}{2} \log_2 n - \frac{n}{2} \log_2 n - \frac{n^2}{8} + \frac{n}{4}$$

$$\leq \frac{n^2}{2} \log_2 n - \frac{n^2}{8} \quad \text{(perché } -\frac{n}{2} \log_2 n + \frac{n}{4} \leq 0 \text{ per } n \geq 2\text{)}$$

Ricapitolando:

$$\sum_{q=1}^{n-1} q \cdot \log_2 q \leq \log_2 n \sum_{q=1}^{n-1} q - \sum_{q=1}^{\lceil \frac{n}{2} \rceil - 1} q \leq \frac{n^2}{2} \log_2 n - \frac{n^2}{8}$$

Abbiamo appena dimostrato che nel caso medio si ha $T(n) = O(n \log n)$.

Inoltre in media non si può avere meglio dell'ottimo che sappiamo essere $\Theta(n \log n)$ da questo deduciamo che per il caso medio vale anche $T(n) = \Omega(n \log n)$.

Possiamo dunque concludere che nel caso medio il Quicksort ha un costo computazionale:

$$T(n) = \Theta(n \log n)$$