Οι διαφάνειες βασίζονται σε αυτές του ακόλουθου μαθήματος:

Introduction to Algorithms (6-046J), MIT

https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-046j-introduction-to-algorithms-sma-5503-fall-2005/

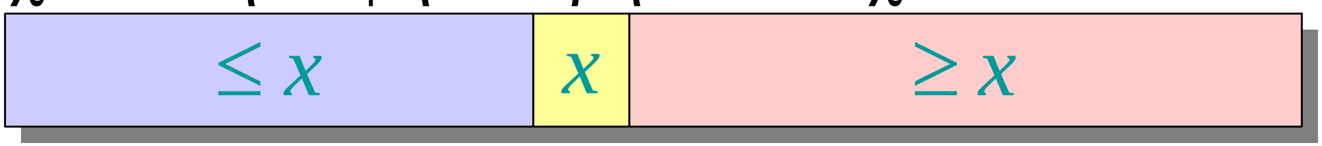
Ταχυταξινόμηση

- Επινοήθηκε από τον C.A.R. Hoare το 1962.
- Αλγόριθμος Διαίρει και Βασίλευε.
- Ταξινομεί «επί τόπου» (όπως η ενθετική ταξινόμηση σε αντίθεση με τη συγχωνευτική ταξινόμηση).
- Πολύ πρακτικός (αν ρυθμιστεί κατάλληλα).

Διαίρει και Βασίλευε

Ταχυταξινόμηση μίας συστοιχίας n στοιχείων:

1. ***Διαίρει:*** Διαμέριση της συστοιχίας σε δύο υπο-συστοιχίες γύρω από ένα ***στοιχείο-οδηγό*** x έτσι ώστε τα στοιχεία στη χαμηλότερη συστοιχία $\leq x \leq$ στοιχεία στην υψηλότερη συστοιχία.



2. ***Βασίλευε:*** Αναδρομική ταξινόμηση στις δύο υπο-συστοιχίες
3. ***Συνδύασε:*** Τετριμμένο.

Κλειδί: Υπορουτίνα διαμέρισης γραμμικού χρόνου.

Υπορουτίνα διαμέρισης

PARTITION(A, p, q) $\rhd$ $A[p\,..\,q]$
    $x \leftarrow A[p]$ $\rhd$ pivot = $A[p]$
    $i \leftarrow p$
    for $j \leftarrow p + 1$ **to** q
        do if $A[j] \leq x$
            then $i \leftarrow i + 1$
                exchange $A[i] \leftrightarrow A[j]$
    exchange $A[p] \leftrightarrow A[i]$
    return i

Χρόνος Εκτέλεσης $= O(n)$ για n στοιχεία.

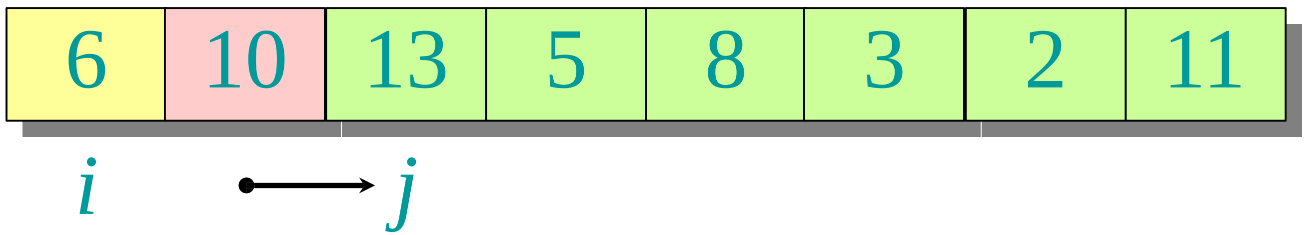
Invariant:

x	$\leq x$	$\geq x$	?
p	i	j	q

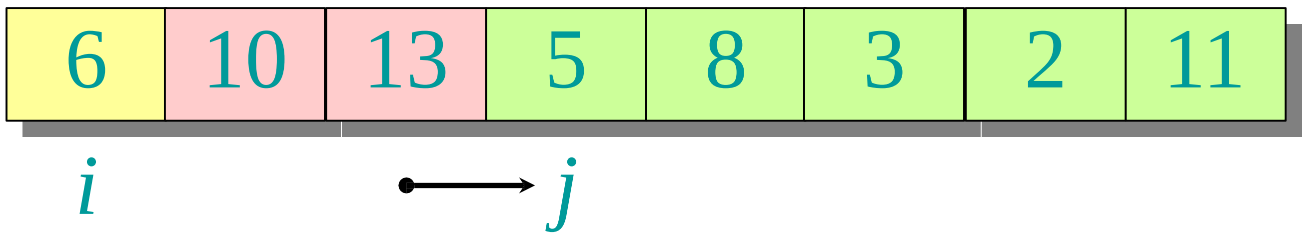
Παράδειγμα διαμέρισης

6	10	13	5	8	3	2	11
i	*j*						

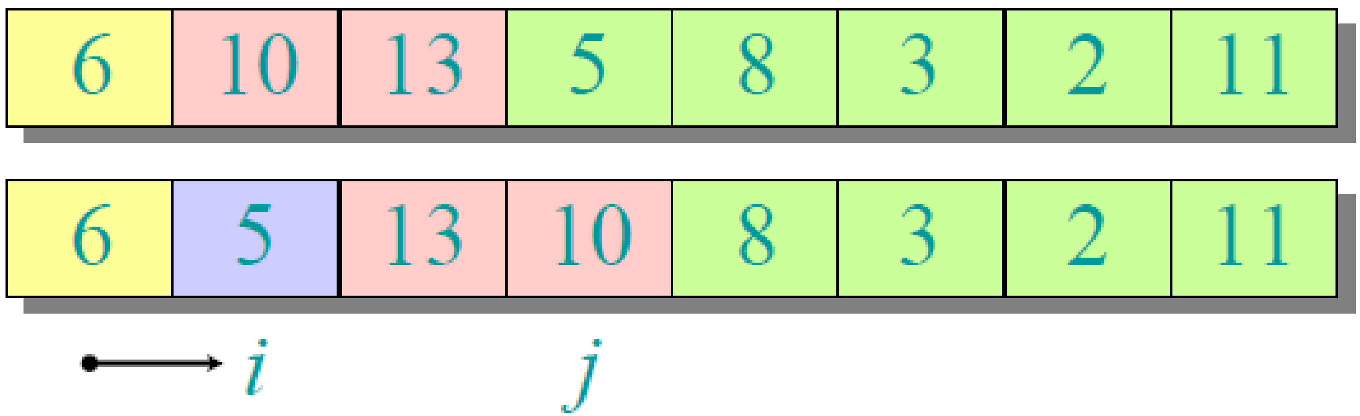
Παράδειγμα διαμέρισης



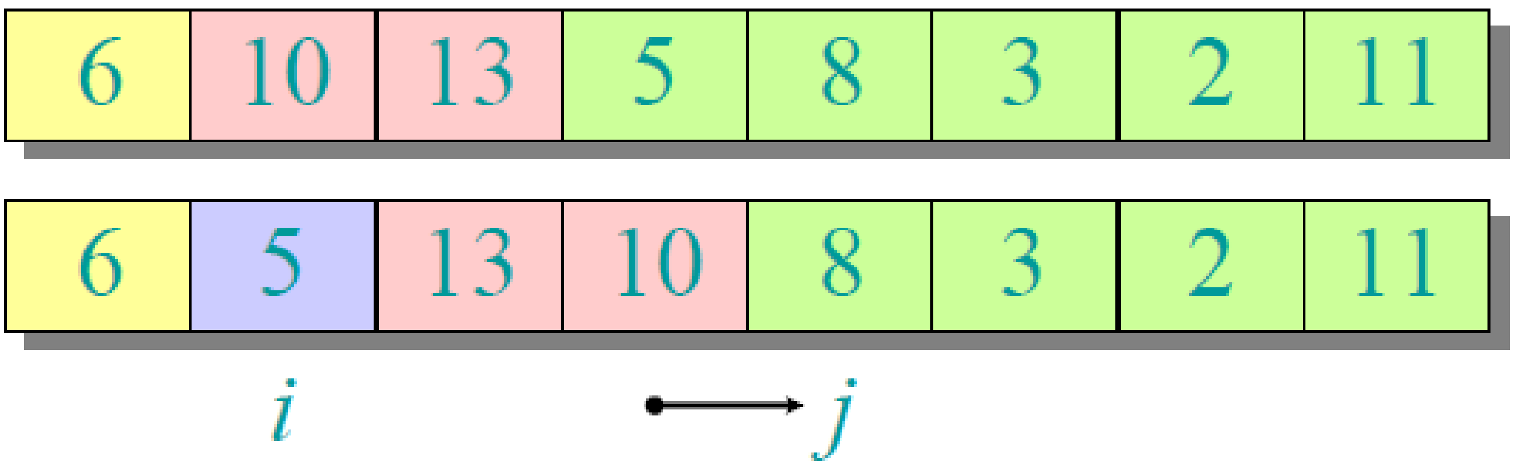
Παράδειγμα διαμέρισης



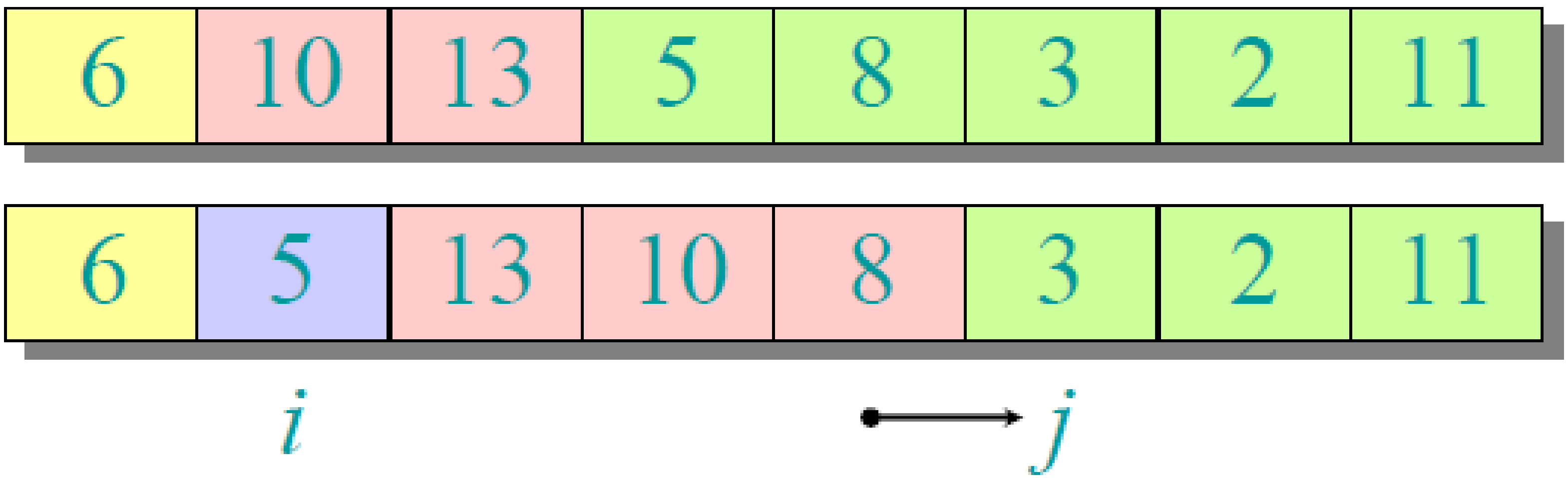
Παράδειγμα διαμέρισης



Παράδειγμα διαμέρισης



Παράδειγμα διαμέρισης



Παράδειγμα διαμέρισης

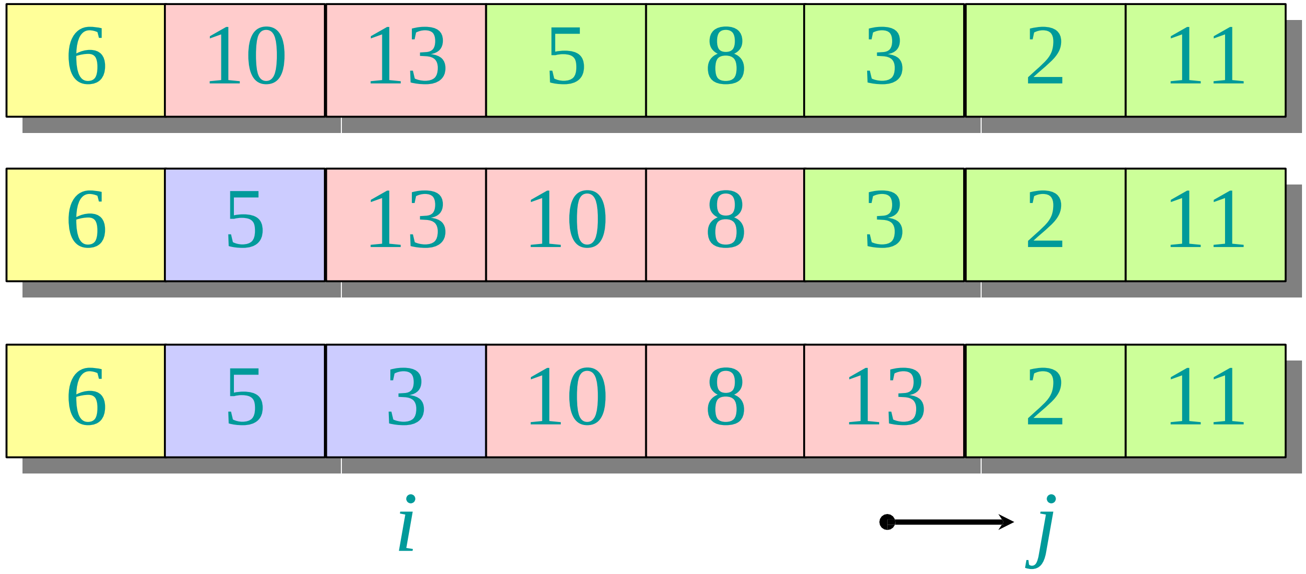
6	10	13	5	8	3	2	11

6	5	13	10	8	3	2	11

6	5	3	10	8	13	2	11

i j

Παράδειγμα διαμέρισης



Παράδειγμα διαμέρισης

6	10	13	5	8	3	2	11
6	5	13	10	8	3	2	11
6	5	3	10	8	13	2	11
6	5	3	2	8	13	10	11
			i			*j*	

Παράδειγμα διαμέρισης

6	10	13	5	8	3	2	11

6	5	13	10	8	3	2	11

6	5	3	10	8	13	2	11

6	5	3	2	8	13	10	11
			i				j

Παράδειγμα διαμέρισης

6	10	13	5	8	3	2	11

6	5	13	10	8	3	2	11

6	5	3	10	8	13	2	11

6	5	3	2	8	13	10	11

i j

Παράδειγμα διαμέρισης

6	10	13	5	8	3	2	11
6	5	13	10	8	3	2	11
6	5	3	10	8	13	2	11
6	5	3	2	8	13	10	11
2	5	3	6	8	13	10	11
			i				

Ψευδοκώδικας ταχυταξινόμησης

QUICKSORT(A, p, r)
    if $p < r$
        then $q \leftarrow$ PARTITION(A, p, r)
            QUICKSORT$(A, p, q–1)$
            QUICKSORT$(A, q+1, r)$

Αρχική Κλήση: QUICKSORT$(A, 1, n)$

Ανάλυση Ταχυταξινόμησης

- Υποθέτουμε ότι όλα στοιχεία εισόδου είναι διαφορετικά.
- Έστω $T(n) =$ ο χρόνος εκτέλεσης χειρότερης περίπτωσης σε μία συστοιχία n στοιχείων.

Χειρότερη περίπτωση Ταχυταξινόμησης

- Η είσοδος είναι ταξινομημένη σε αντίστροφη σειρά.
- Διαμέριση γύρω από το ελάχιστο ή μέγιστο στοιχείο.
- Μία πλευρά της διαμέρισης δεν έχει στοιχεία.

$$T(n) = T(0) + T(n-1) + \Theta(n)$$

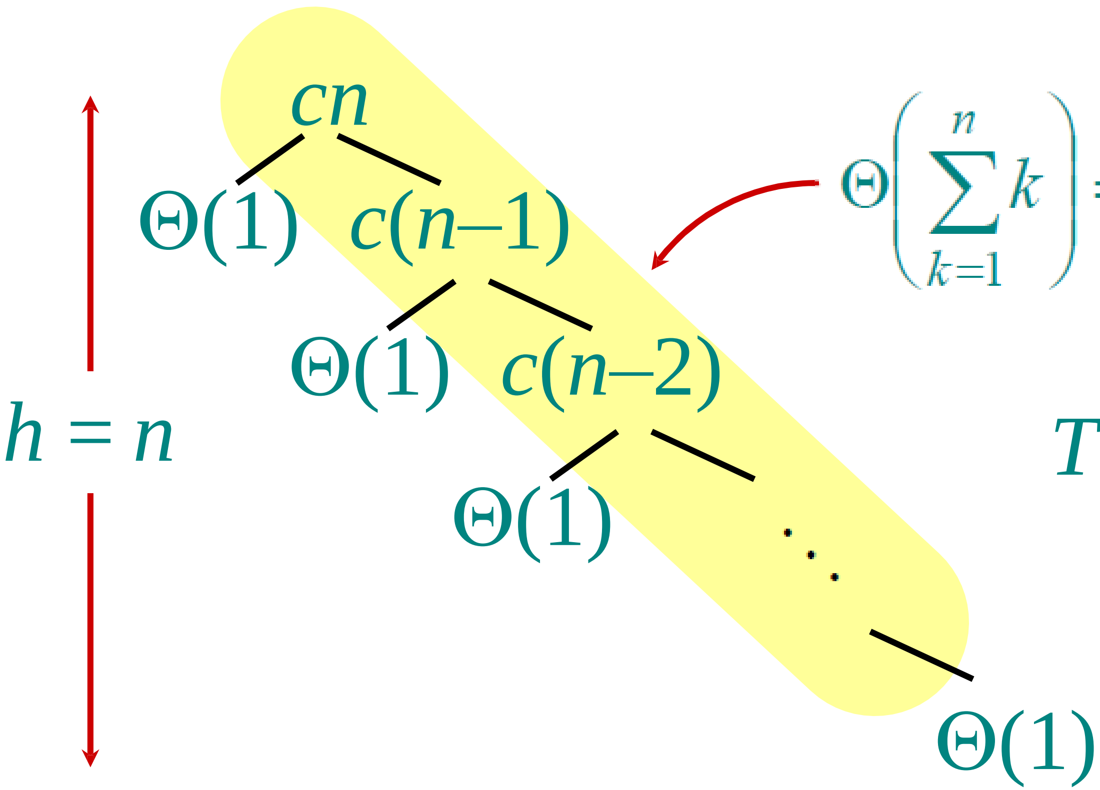
$$= \Theta(1) + T(n-1) + \Theta(n)$$

$$= T(n-1) + \Theta(n)$$ ***(αριθμητική πρόοδος)***

$$= \Theta(n^2)$$

Δένδρο αναδρομής χειρότερης περίπτωσης

$$T(n) = T(0) + T(n-1) + cn$$



$$\Theta\left(\sum_{k=1}^{n} k\right) = \Theta(n^2)$$

$$T(n) = \Theta(n) + \Theta(n^2)$$
$$= \Theta(n^2)$$

Ανάλυση Καλύτερης περίπτωσης

(Διαίσθηση μόνο)

Αν είμαστε τυχεροί, η PARTITION διαιρεί τη συστοιχία στη μέση:

$$T(n) = 2T(n/2) + \Theta(n) = \Theta(n \lg n)$$

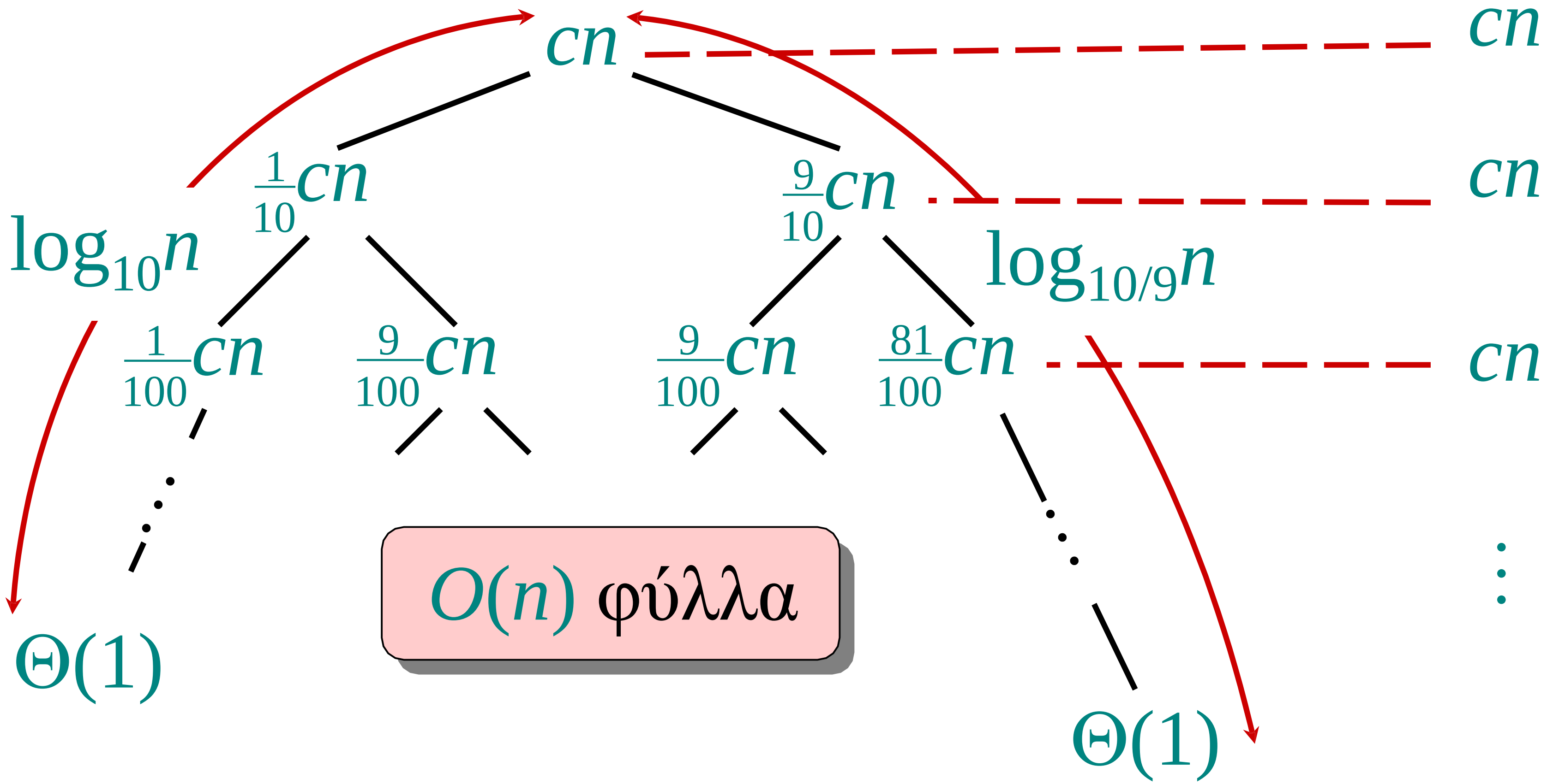
(όπως η συγχωνευτική ταξινόμηση)

Αν η διαίρεση είναι πάντα 1/10:9/10

$$T(n) = T\left(\tfrac{1}{10}n\right) + T\left(\tfrac{9}{10}n\right) + \Theta(n)$$

Ποια είναι η λύση της αναδρομής;

Ανάλυση της «σχεδόν καλύτερης» περίπτωσης



$\Theta(n \lg n)$

$$cn \log_{10} n \le T(n) \le cn \log_{10/9} n + O(n)$$

Περισσότερη Διαίσθηση

Ας υποθέσουμε ότι εναλλάξ είμαστε τυχεροί, άτυχοι, τυχεροί, άτυχοι, ….

$$L(n) = 2U(n/2) + \Theta(n) \quad \textbf{\textit{Τυχεροί}}$$

$$U(n) = L(n-1) + \Theta(n) \quad \textbf{\textit{άτυχοι}}$$

Επιλύοντας:

$$\begin{aligned} L(n) &= 2(L(n/2-1) + \Theta(n/2)) + \Theta(n) \\ &= 2L(n/2-1) + \Theta(n) \\ &= \Theta(n \lg n) \quad \textbf{\textit{Τυχεροί!}} \end{aligned}$$

Πως μπορούμε να εξασφαλίσουμε ότι θα είμαστε συνήθως τυχεροί;

Τυχαιοκρατική ταχυταξινόμηση

Ιδέα: Διαμέριση γύρω από τυχαίο στοιχείο.

- Ο χρόνος εκτέλεσης είναι ανεξάρτητος της διάταξης της εισόδου.
- Δεν χρειάζεται να γίνει καμία υπόθεση σχετικά με την κατανομή εισόδου.
- Καμία είσοδος δεν προκαλεί συμπεριφορά χειρότερης περίπτωσης.
- Η χειρότερη περίπτωση καθορίζεται μόνο από την έξοδο της ψευδογεννήτριας τυχαίων αριθμών.

Τυχαιοκρατική ταξινόμηση

```
ΤΥΧΑΙΟΚΡΑΤΙΚΗ ΔΙΑΜΕΡΙΣΗ(A, p, r)
i = RANDOM(p, r)
εναλλάσσουμε το A[p] με το A[i]
επιστροφή ΔΙΑΜΕΡΙΣΗ (A, p, r)
```

```
ΤΥΧΑΙΟΚΡΑΤΙΚΗ ΤΑΧΥΤΑΞΙΝΟΜΗΣΗ(A, p, r)
αν p < r
    q = ΤΥΧΑΙΟΚΡΑΤΙΚΗ ΔΙΑΜΕΡΙΣΗ (A, p, r)
    ΤΥΧΑΙΟΚΡΑΤΙΚΗ ΤΑΧΥΤΑΞΙΝΟΜΗΣΗ (A, p, q – 1)
    ΤΥΧΑΙΟΚΡΑΤΙΚΗ ΤΑΧΥΤΑΞΙΝΟΜΗΣΗ (A, q + 1, r)
```

Ανάλυση Τυχαιοκρατικής Ταχυταξινόμησης

Έστω $T(n)$ = η τυχαία μεταβλητή για το χρόνο εκτέλεσης της τυχαιοκρατικής ταξινόμησης σε μία είσοδο μεγέθους n, υποθέτοντας ότι οι τυχαίοι αριθμοί είναι ανεξάρτητοι μεταξύ τους.

Για $k = 0, 1, \ldots, n-1$, ορίζουμε τη δείκτρια μεταβλητή

$$X_k = \begin{cases} 1 & \text{αν η PARTITION παράγει μία } k : n-k-1 \text{ διαμέριση,} \\ 0 & \text{αλλιώς.} \end{cases}$$

$E[X_k] = \Pr\{X_k = 1\} = 1/n$, αφού όλες οι διαμερίσεις είναι εξίσου πιθανές, υποθέτοντας ότι τα στοιχεία είναι διαφορετικά.

Ανάλυση (συν.)

$$T(n) = \begin{cases} T(0) + T(n-1) + \Theta(n) & \text{αν } 0 : n-1 \text{ διαμέριση,} \\ T(1) + T(n-2) + \Theta(n) & \text{αν } 1 : n-2 \text{ διαμέριση,} \\ \vdots & \\ T(n-1) + T(0) + \Theta(n) & \text{αν } n-1 : 0 \text{ διαμέριση} \end{cases}$$

$$= \sum_{k=0}^{n-1} X_k \left(T(k) + T(n-k-1) + \Theta(n)\right)$$

Υπολογισμός Αναμενόμενης Τιμής

$$E[T(n)] = E\left[\sum_{k=0}^{n-1} X_k\big(T(k) + T(n-k-1) + \Theta(n)\big)\right]$$

$$= \sum_{k=0} E\left[X_k\big(T(k) + T(n-k-1) + \Theta(n)\big)\right]$$

$$= \sum_{k=0}^{n-1} E\left[X_k\right] \cdot E\left[T(k) + T(n-k-1) + \Theta(n)\right]$$

$$= \frac{1}{n}\sum_{k=0}^{n-1} E\left[T(k)\right] + \frac{1}{n}\sum_{k=0}^{n-1} E\left[T(n-k-1)\right] + \frac{1}{n}\sum_{k=0}^{n-1} \Theta(n)$$

$$= \frac{2}{n}\sum_{k=1}^{n-1} E\left[T(k)\right] + \Theta(n)$$

Δύσκολη Αναδρομή

$$E[T(n)] = \frac{2}{n}\sum_{k=2}^{n-1} E[T(k)] + \Theta(n)$$

(Οι όροι για $k = 0, 1$ απορροφώνται στο $\Theta(n)$.)

Απόδειξε: $E[T(n)] \le an \lg n$ για μία σταθερά $a > 0$.

- Επίλεξε a αρκετά μεγάλο έτσι ώστε το $an \lg n$ να κυριαρχεί επί των $E[T(n)]$ για επαρκώς μικρά $n \ge 2$.

Χρησιμοποιούμε το γεγονός:

$$\sum_{k=2}^{n-1} k \lg k \le \tfrac{1}{2} n^2 \lg n - \tfrac{1}{8} n^2$$

Μέθοδος Αντικατάστασης

$$E[T(n)] \le \frac{2}{n}\sum_{k=2}^{n-1} ak\lg k + \Theta(n)$$

$$= \frac{2a}{n}\left(\frac{1}{2}n^2\lg n - \frac{1}{8}n^2\right) + \Theta(n)$$

$$= an\lg n - \left(\frac{an}{4} - \Theta(n)\right)$$

$$\le an\lg n,$$

αν το a επιλεχθεί αρκετά μεγάλο έτσι ώστε το $an/4$ επικρατεί του $\Theta(n)$.

Διαφορετική ανάλυση της τυχαιοκρατικής ταξινόμησης

Έστω z_i από $z_1, z_2, \ldots, z_n$ το *i-στο* μικρότερο στοιχείο

Έστω Z_{ij} είναι το σύνολο των στοιχείων $Z_{ij} = z_i, z_{i+1}, \ldots, z_j$

$$\text{Έστω } X_{ij} = I\{z_i \text{ συγκρίνεται με } z_j\} = \begin{cases} 1 & \alpha\nu\ z_i \text{ συγκρίνεται με } z_j \\ 0 & \text{αλλιώς} \end{cases}$$

- Πόσες φορές το z_i συγκρίνεται με το z_j;
- Το πολύ μία φορά. Γιατί;

Συνολικό πλήθος συγκρίσεων:

$$X = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij}$$

$$E[X] = E\left[\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} X_{ij}\right]$$

$$= \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} E[X_{ij}]$$

$$= \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} p\{z_i \text{ συγκρίνεται με } z_j\}$$

$$p\{z_i \text{ συγκρίνεται με } z_j\} \text{ ;}$$

- Το στοιχείο οδηγός διαχωρίζει το σύνολο των στοιχείων σε δύο σύνολα
- Τα στοιχεία από το ένα σύνολο δεν θα συγκριθούν με στοιχεία του άλλου συνόλου.
- Αν ένα στοιχείο οδηγός x επιλεχθεί τέτοιο ώστε $z_i < x < z_j$ τότε τα z_i και z_j δεν θα συγκριθούν ποτέ.

$p\{z_i$ συγκρίνεται με $z_j\}$

$$
\begin{aligned}
&= \Pr\{z_i \text{ ή } z_j \text{ επιλέγεται πρώτο οδηγός από το } Z_{ij}\} \\
&= \Pr\{z_i \text{ επιλέγεται πρώτο οδηγός από το } Z_{ij}\} \\
&\quad + \Pr\{z_j \text{ επιλέγεται πρώτο οδηγός από το } Z_{ij}\} \\
&= \frac{1}{j-i+1} + \frac{1}{j-i+1} \\
&= \frac{2}{j-i+1} \,. \qquad (7.3)
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}[X] &= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j-i+1} \\
&= \sum_{i=1}^{n-1} \sum_{k=1}^{n-i} \frac{2}{k+1} \\
&< \sum_{i=1}^{n-1} \sum_{k=1}^{n} \frac{2}{k} \\
&= \sum_{i=1}^{n-1} O(\lg n) \\
&= O(n \lg n) \,.
\end{aligned}
$$