MEDIANE E STATISTICHE D'ORDINE

- L' i-ESIMA STATISTICA D'ORDINE DI UN INSIEME DI n ELEMENTI E' L' i-ESIMO ELEMENTO PIÙ PICCOLO, CIOE' UN ELEMENTO x DELL'INSIEME TALE CHE I RIMANENTI $(n-1)$ ELEMENTI POSSONO ESSERE COSÌ SUDDIVISI:

$$\underbrace{\underbrace{\boxed{\qquad\qquad\qquad}}_{i-1} \le x \le \underbrace{\boxed{\qquad\qquad\qquad\qquad}}_{n-i}}_{n}$$

CASI PARTICOLARI

- $i = 1$ MINIMO
- $i = n$ MASSIMO
- $i = \left\lfloor \frac{n+1}{2} \right\rfloor$ MEDIANA (INFERIORE)
- $i = \left\lceil \frac{n+1}{2} \right\rceil$ MEDIANA SUPERIORE

PROBLEMA DELLA SELEZIONE

INPUT: - UN INSIEME A DI n NUMERI (DISTINTI)

- UN INTERO $1 \leq i \leq n$

OUTPUT: L' i-ESIMA STATISTICA D'ORDINE DI A

ALGORITMO BANALE:

- SI ORDINI A
- SI RESTITUISCA L'ELEMENTO $A[i]$

COMPLESSITA': $O(n \lg n)$

- PRESENTEREMO UN ALGORITMO LINEARE

ALCUNI CASI PARTICOLARI

```
MINIMUM(A)
min = A[1]
for i = 2 to A.length
    if min > A[i]
        min = A[i]
return min
```

\# CONFRONTI ESEGUITI = $n-1$

ESERCIZIO DIMOSTRARE CHE SONO NECESSARI $(n-1)$ CONFRONTI

- ANALOGAMENTE PER IL MASSIMO

CALCOLARE MINIMO E MASSIMO DI n ELEMENTI

I SOLUZIONE

- CALCOLARE IL MINIMO DI n ELEMENTI
- CALCOLARE IL MASSIMO DI $(n-1)$ ELEMENTI

\# CONFRONTI $= (n-1) + (n-2) = 2n-3$

- C'E' UNA SOLUZIONE PIÙ EFFICIENTE (IN TERMINI DI NUMERO DI CONFRONTI)?

II SOLUZIONE

- MANTENENDO min E max CORRENTI,
  - SI SUDDIVIDANO GLI n ELEMENTI IN $\frac{n}{2}$ COPPIE
  - PER CIASCUNA COPPIA SI CONFRONTINO I SUOI ELEMENTI, QUINDI SI CONFRONTI IL MINIMO CON min E IL MASSIMO CON max, AGGIORNANDO min E max SE NECESSARIO

  (SE RIMANE UN ELEMENTO SPAIATO, LO SI CONFRONTI DIRETTAMENTE CON min E max)

CONFRONTI:

- SE n E' PARI:

$$1 + 3 \cdot \frac{n-2}{2} = \frac{3n - 6 + 2}{2} = \frac{3n}{2} - 2$$

- SE n E' DISPARI:

$$1 + 3 \cdot \frac{n-3}{2} + 2 = \frac{3n - 9 + 2 + 4}{2}$$

$$= \frac{3n-3}{2} = \frac{3n+1}{2} - 2 = \left\lceil \frac{3n}{2} \right\rceil - 2$$

DUNQUE, IN DEFINITIVA:

$$\# \text{ CONFRONTI} = \left\lceil \frac{3n}{2} \right\rceil - 2$$

NOTA: SI DIMOSTRA CHE $\left\lceil \frac{3n}{2} \right\rceil - 2$ CONFRONTI SONO NECESSARI

SELEZIONE IN TEMPO ATTESO LINEARE

RANDOMIZED-SELECT(A, p, r, i)

```
if p == r
    return A[p]
q = RANDOMIZED-PARTITION(A, p, r)
k = q - p + 1
if i == k                // the pivot value is the answer
    return A[q]
elseif i < k
    return RANDOMIZED-SELECT(A, p, q - 1, i)
else return RANDOMIZED-SELECT(A, q + 1, r, i - k)
```

- SI DIMOSTRA CHE IL TEMPO ATTESO DI RANDOMIZED-SELECT È $O(n)$

(CASO PESSIMO: $O(n^2)$)

SELEZIONE IN TEMPO LINEARE (NEL CASO PEGGIORE)

$SELECT(A, i)$

- SI DIVIDA A IN $\lfloor \frac{n}{5} \rfloor$ GRUPPI DI 5 ELEMENTI, PIÙ UN EVENTUALE GRUPPO CON I RESTANTI $n \bmod 5$ ELEMENTI
- SI DETERMININO LE MEDIANE DEGLI $\lceil \frac{n}{5} \rceil$ GRUPPI E SIA M LA SEQUENZA DI TALI MEDIANE
- SI EFFETTUI LA CHIAMATA RICORSIVA $m = SELECT\left(M, \left\lceil \frac{|M|}{2} \right\rceil\right)$
- SI PARTIZIONI A IN A_1, A_2, A_3 DOVE

$$\left.\begin{aligned} A_1 &= [x \in A : x < m] \\ A_2 &= [x \in A : x = m] \\ A_3 &= [x \in A : x > m] \end{aligned}\right\} \text{SOTTOSEQUENZE DI } A$$

- **if** $|A_1| \geq i$

  then **return** SELECT(A_1, i)

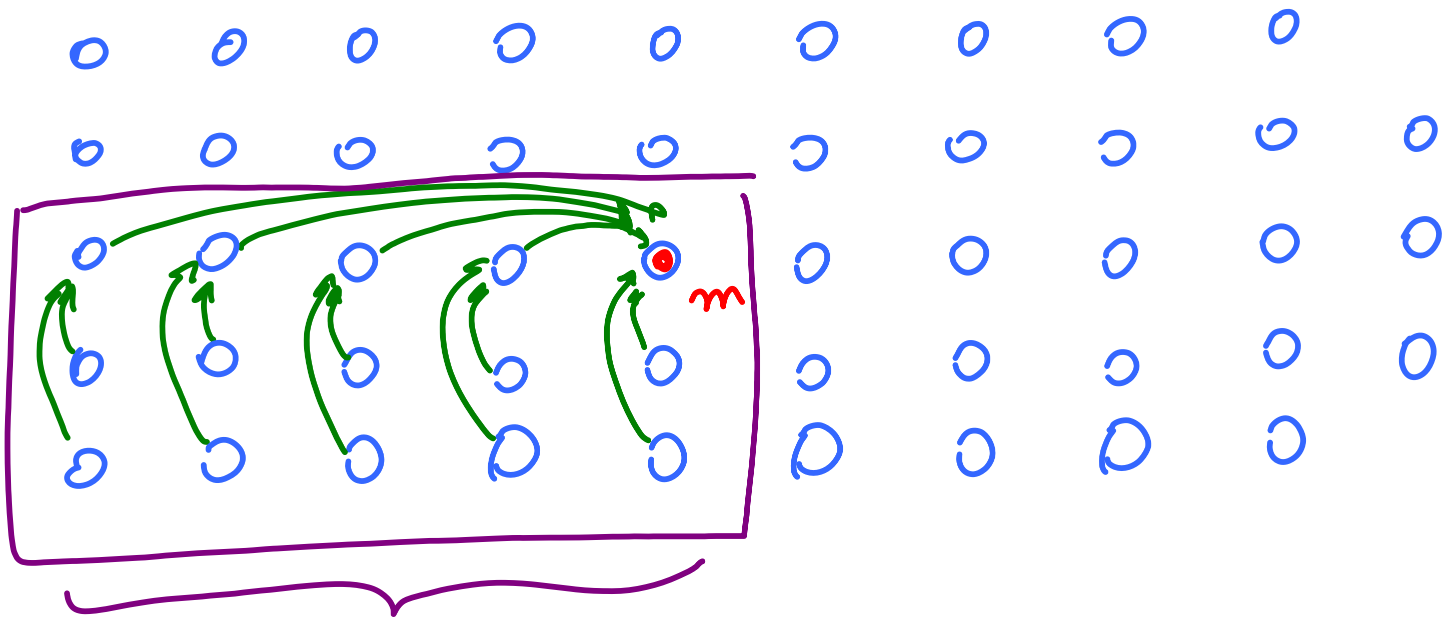
  else **if** $(|A_1| + |A_2| \geq i)$

  then **return** m

  else **return** SELECT$(A_3, i - |A_1| - |A_2|)$

CORRETTEZZA: PER INDUZIONE SU $|A|$

ANALISI DI COMPLESSITA'

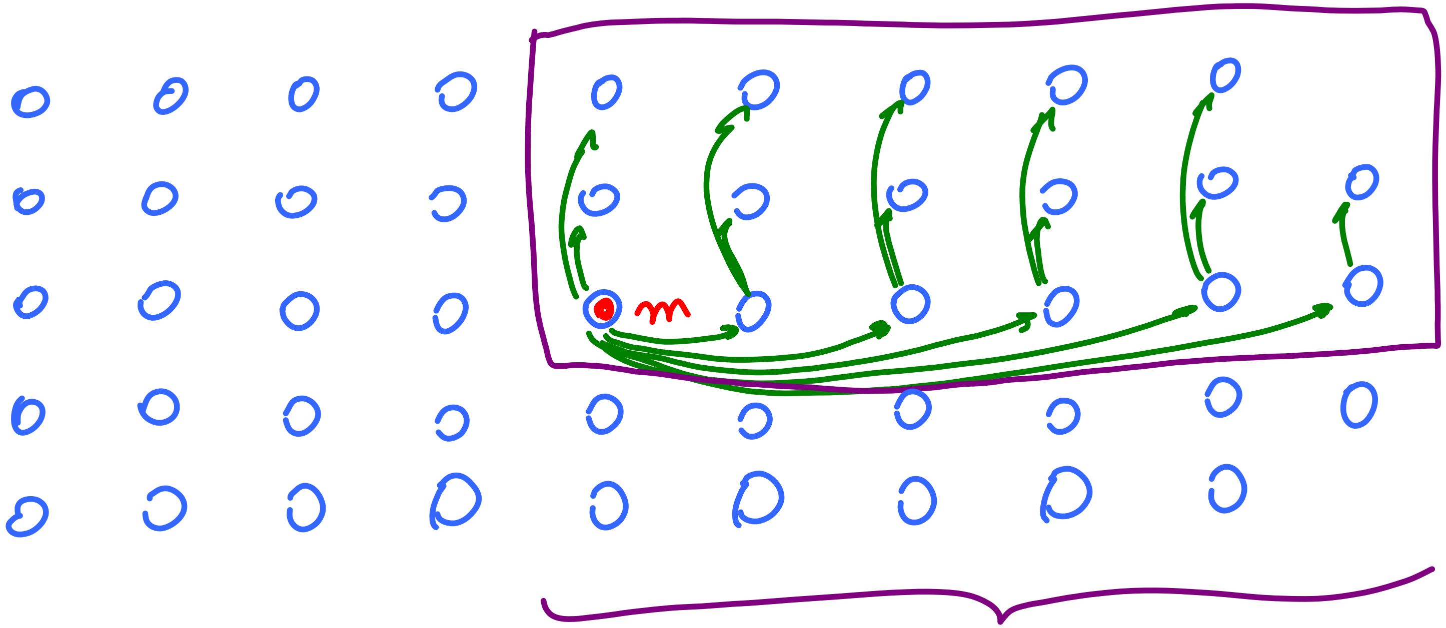


$$\left\lceil \left\lceil \frac{n}{5} \right\rceil / 2 \right\rceil$$

- LIMITE INFERIORE SUL NUMERO DI ELEMENTI $\leq$ m

$$|A_1|+|A_2| \geq 3\cdot\left\lceil \left\lceil \frac{n}{5} \right\rceil / 2 \right\rceil - 2 = 3\left\lceil \frac{n}{10} \right\rceil - 2 \geq \frac{3n}{10} - 2$$

- QUINDI: $|A_3| = n - (|A_1|+|A_2|) \leq n - \frac{3n}{10} + 2 = \frac{7n}{10} + 2$



$$\left\lfloor \left\lceil \frac{n}{5} \right\rceil / 2 \right\rfloor + 1 \qquad \le \frac{7n}{10} + 2$$

- LIMITE INFERIORE SUL NUMERO DI ELEMENTI $\geq m$

$$|A_2| + |A_3| \geq 3\left(\left\lfloor \left\lceil \frac{n}{5} \right\rceil / 2 \right\rfloor + 1\right) - 2 \geq 3 \cdot \left\lceil \left\lceil \frac{n}{5} \right\rceil / 2 \right\rceil - 2 = 3 \left\lceil \frac{n}{10} \right\rceil - 2$$

$$\geq \frac{3n}{10} - 2$$

- QUINDI : $|A_1| = n - (|A_2| + |A_3|) \leq n - \frac{3n}{10} + 2 = \frac{7n}{10} + 2$

- QUINDI IL TEMPO DI ESECUZIONE $T(n)$ SODDISFA LA SEGUENTE RICORRENZA:

$$T(n) \leq T\left(\left\lceil \frac{n}{5} \right\rceil\right) + T\left(\frac{7n}{10} + 2\right) + O(n)$$

- DIMOSTRIAMO PER INDUZIONE CHE

$T(n) \leq cn$, PER n SUFF. GRANDE E PER QUALCHE $c > 0$

$$T(n) \leq c\left\lceil \frac{n}{5} \right\rceil + c\left(\frac{7n}{10} + 2\right) + an$$

$$< c\frac{n}{5} + c + \frac{7cn}{10} + 2c + an = \frac{9cn}{10} + 3c + an$$

$$= cn + \left(-\frac{cn}{10} + 3c + an\right)$$

- IN PARTICOLARE, SE $c > 40a$, ALLORA PER OGNI $n \geq 40$ SI HA:

$$-\frac{cn}{10} + 3c + an = \left(-\frac{c}{10} + a\right) n + 3c \leq \left(-\frac{c}{10} + a\right) 40 + 3c$$

$$= -\frac{c}{10} \cdot 40 + 40a + 3c = -4c + 40a + 3c = -c + 40a < 0$$

E QUINDI

$$T(n) \leq cn, \quad \text{CON} \quad c > \max\left(40a, T(1), \frac{T(2)}{2}, \ldots, \frac{T(39)}{39}\right)$$

E $n \in \mathbb{N}^+$

ESERCIZI

9.3-3
Show how quicksort can be made to run in $O(n \lg n)$ time in the worst case, assuming that all elements are distinct.

9.3-5
Suppose that you have a "black-box" worst-case linear-time median subroutine. Give a simple, linear-time algorithm that solves the selection problem for an arbitrary order statistic.

9.3-6
The kth ***quantiles*** of an n-element set are the $k - 1$ order statistics that divide the sorted set into k equal-sized sets (to within 1). Give an $O(n \lg k)$-time algorithm to list the kth quantiles of a set.

9.3-7
Describe an $O(n)$-time algorithm that, given a set S of n distinct numbers and a positive integer $k \leq n$, determines the k numbers in S that are closest to the median of S.

9.3-8

Let $X[1 \, . \, . \, n]$ and $Y[1 \, . \, . \, n]$ be two arrays, each containing n numbers already in sorted order. Give an $O(\lg n)$-time algorithm to find the median of all $2n$ elements in arrays X and Y.

9.3-9

Professor Olay is consulting for an oil company, which is planning a large pipeline running east to west through an oil field of n wells. The company wants to connect a spur pipeline from each well directly to the main pipeline along a shortest route (either north or south), as shown in Figure 9.2. Given the x- and y-coordinates of the wells, how should the professor pick the optimal location of the main pipeline, which would be the one that minimizes the total length of the spurs? Show how to determine the optimal location in linear time.

9-1 *Largest i numbers in sorted order*

Given a set of n numbers, we wish to find the i largest in sorted order using a comparison-based algorithm. Find the algorithm that implements each of the following methods with the best asymptotic worst-case running time, and analyze the running times of the algorithms in terms of n and i.

a. Sort the numbers, and list the i largest.

b. Build a max-priority queue from the numbers, and call EXTRACT-MAX i times.

c. Use an order-statistic algorithm to find the ith largest number, partition around that number, and sort the i largest numbers.