DEFINITION 5.19 (Median)

(1) Für n Zahlen $x_1, \dots, x_n$ ist ein Median eine Zahl m mit den Eigenschaften

$$|\{x_i \mid x_i \leq m\}| \geq \frac{n}{2}$$

$$|\{x_i \mid x_i \geq m\}| \geq \frac{n}{2}$$

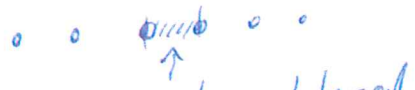
(2) Für Verteilungen:

Mit $\int_{-\infty}^{\infty} f(x)\,dx =: A$ ist m ein Median, wenn

$$\int_{-\infty}^{m} f(x)\,dx \geq \frac{A}{2}$$

und $$\int_{m}^{\infty} f(x)\,dx \geq \frac{A}{2}$$

(3) Für eine ungerade Menge von Zahlen ist ein Median eindeutig, für eine gerade Menge kann (!) es mehr als einen geben.



Median-Intervall

(4) Verallgemeinert:

m hat Rang k in X, wenn

$$|\{x_i \mid x_i \leq m\}| \geq k$$

$$|\{x_i \mid x_i \geq m\}| \geq n-k$$

PROBLEM 5.20 (Median)

Gegeben: n Zahlen $\{x_1, \dots, x_n\} =: X$

Gesucht: Ein Median n von X

Allgemeiner:

PROBLEM 5.21 (Rang-k-Element)

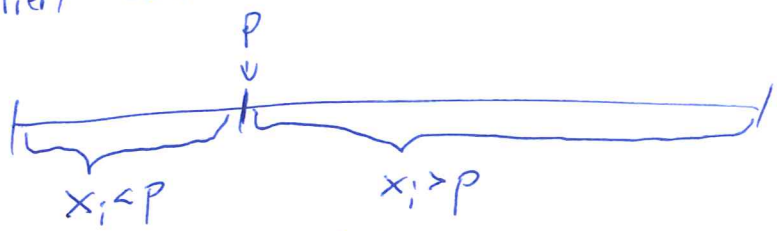
Gegeben: n Zahlen $\{x_1, \dots, x_n\} =: X$, $k \in \{1, \dots, n\}$

Gesucht: Ein Element von Rang k in X

Man kann beide Probleme lösen, indem man sortiert und dann abzählt – Laufzeit $\Theta(n \log n)$.

Geht das schneller?

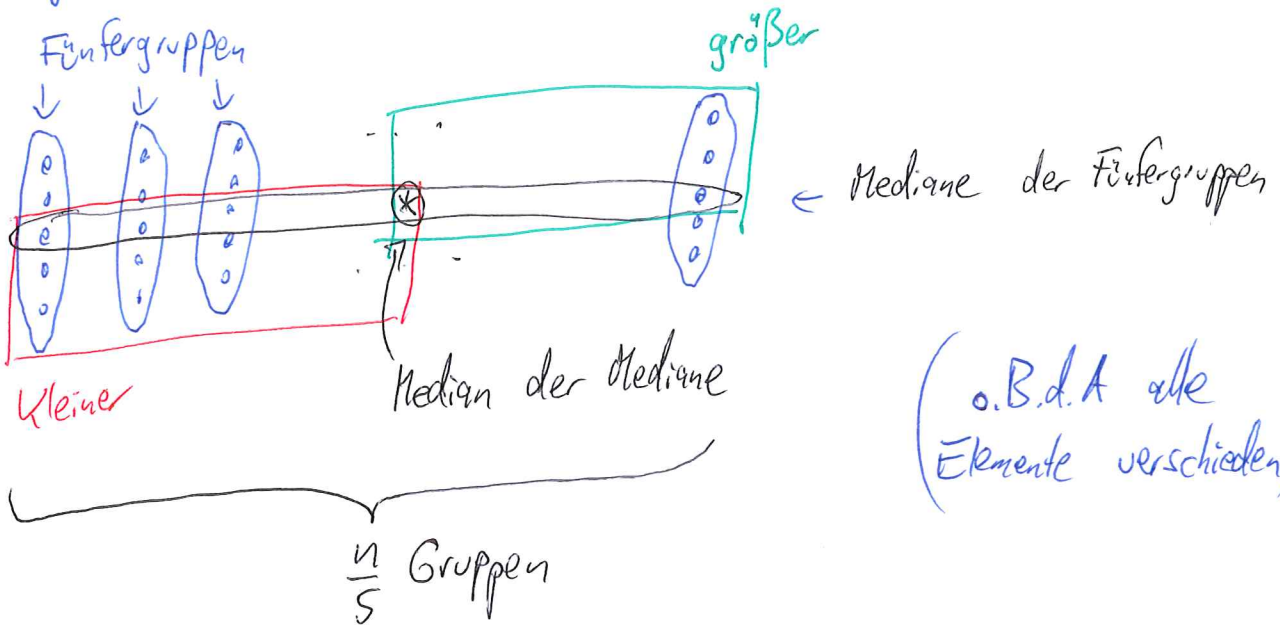
Idee inspiriert von QUICKSORT!



- Wähle ein Pivotelement
- Teile die Menge in größere und kleinere Elemente
- Bei QUICKSORT: Beide Teile rekursiv weiter unterteilen
  Hier: Nur ein Teil! Der mit Median (bzw. Rang k)
- Aber: Bei schlechter Pivotauswahl bleibt nur der größere Teil übrig (Pivot am Rand!), es dauert also $\Omega(n^2)$.
- Ziel: Vermeide Worst Case, indem man jeweils Mindestgröße für beide Teile erreicht.

Umsetzung

(siehe Kapitel 9.3 im Cormen)



(o.B.d.A alle Elemente verschieden)

(1) Teile n Elemente in $\lceil \frac{n}{5} \rceil$ Fünfergruppen (Ggf. kleinere Gruppe für den Rest – Runden wird im weiteren ignoriert, geht in O-Notation unter) } Zeit $\Theta(n)$

(2) Bestimme jeweils Mediane der Gruppen } $\Theta(n)$ (jeweils $O(1)$)

(3) Bestimme (rekursiv) Median der Mediane x } $T(\frac{n}{5})$

(4) Verwende Median der Mediane x als Pivot. Das liefert $i-1$ kleinere, $n-i$ größere Elemente } $\Theta(n)$

(5) Falls $i = k$: fertig
Falls $i < k$: $(k-i)$-tes Element rechts
Falls $i > k$: k-tes Element links } $T(??)$ ↑ Größe des Restarrays

↑ Algorithmus

↓ Analyse

(6) Ist x rechts, werden mindestens die jeweils $\frac{n}{10}$ Mediane eliminiert, die kleiner sind als x – es fallen mitsamt der kleineren Elemente aus der Gruppe jeweils 3 Elemente weg, insgesamt mindestens $\frac{3n}{10}$.

Analog wenn x links liegt:
Es fallen mindesten $\frac{3n}{10}$ größere Elemente weg.

Also hat der Restarray höchstens Größe $\frac{7n}{10}$.

(7) Damit bekommt man die Rekursionsgleichung

$$T(n) = T\left(\frac{n}{5}\right) + T\left(\frac{7n}{10}\right) + \Theta(n)$$

(8) Nach Mastertheorem mit

$$k=1, \quad \alpha_1 = \frac{1}{5}, \quad \alpha_2 = \frac{7}{10}$$

ist $\left(\frac{1}{5}\right)^1 + \left(\frac{7}{10}\right)^1 = \frac{9}{10} < 1$.

Also ist $T(n) \in \Theta(n)$.

<u>SATZ 5.22</u> (Blum, Floyd, Pratt, Rivest, Tarjan 1973 [BFPRS])
„R" von RSA

Median der Mediane berechnet ein Element von Rang k in $O(n)$

<u>BEOBACHTUNG 5.23</u>

$\Omega(n)$ ist unvermeidbar für die Berechnung eines Medians, da man alle Elemente anschauen muss. Das Problem hat also Komplexität $\Theta(n)$.

<u>FRAGE 5.24</u>

Was passiert, wenn man Dreier- statt Fünfergruppen verwendet?
Was passiert bei Siebenergruppen?