Given an unsorted list $S = [s_1, \ldots, s_n]$ for odd n,
find the median m

Easy alg: $O(n \log n)$ time by sorting S.

Clever deterministic alg.: $O(n)$ divide & conquer algorithm.

Easy randomized alg.: $O(n)$ expected running time

QuickSelect approach:

Find a "good" pivot p

Partition S into $S_{<p}, S_{=p}, S_{>p}$

Then recurse in one of 3 sublists.

Good pivot means: $|A_{<p}| \leq \frac{3}{4} n$ & $|A_{>p}| \leq \frac{3}{4} n$

Then $T(n) \leq T\left(\frac{3}{4} n\right) + O(n) = O(n)$

How to get a good pivot?

Chose a random element r of S.

With prob. $= \frac{1}{2}$, r is a good pivot.

Check if it is. If so use it &
if not repeat. $O(1)$ trials in expectation.

Goal: simple randomized alg. with $O(n)$ running time with high probability (whp)

whp means Prob. $\geq 1 - \frac{1}{n^c}$ for constant $c > 0$.

in this case, $c = \frac{1}{4}$. (can boost with slower run time)

Idea: Find ℓ & u in S where

$\ell \leq m \leq u$ &

ℓ and u are close to $\frac{n}{2}^{nd}$ largest.

C=center

Then, let C be those $x \in S$ where $\ell \leq x \leq u$

By we know $|C|$ is small

So sort C, check $|S_{<\ell}|$

& output $\left(\frac{n}{2} - |S_{<\ell}|\right)^{th}$ smallest of C.

Need $|C| = o(n)$, we'll aim for $|C| = n^{3/4}$

How to find "good" ℓ & u?

Choose a random subset R of S.

Let R be $n^{3/4}$ random elements of S, chosen with replacement

for simplicity, so R may be multiset.

Ideally $m = \text{median}(S) \stackrel{?}{=} \text{median}(R)$

in which case it's the $\frac{n^{3/4}}{2}$ nd smallest of R.

To account for the stddev,

set $\ell = \frac{1}{2}n^{3/4} - \sqrt{n}$ & $u = \frac{1}{2}n^{3/4} + \sqrt{n}$

That defines the algorithm.

Input: $S = [s_1, \ldots, s_n]$ for odd n

Output: Median m of S.

1. Pick $n^{3/4}$ random elements of S (with replacement). Let R be this multiset.
2. Sort R
3. Let ℓ be $\frac{1}{2}n^{3/4} - \sqrt{n}$ smallest of R, & u be $\frac{1}{2}n^{3/4} + \sqrt{n}$ smallest of R.
4. Let $C = \{x \in S : \ell \leq x \leq u\}$ & $S_{<\ell}$ & $S_{>u}$ Find by 1 scan of S.
5. If $|S_{<\ell}| > \frac{n}{2}$ or $|S_{>u}| > \frac{n}{2}$ then output FAIL.
6. If $|C| \leq 4n^{3/4}$ then sort C else output FAIL.
7. Output $\left(\frac{n}{2} - |S_{<\ell}|\right)^{th}$ smallest of C.

$$\Pr\left(\overset{\text{OUTPUT}}{\text{FAIL}}\right) \leq n^{-1/4}$$

Let $\mathcal{E}_1 = \left\{ \left|\{r \in R : r \leq m\}\right| < \frac{1}{2} n^{3/4} - \sqrt{n} \right\}$

$\mathcal{E}_2 = \left\{ \left|\{r \in R : r \geq m\}\right| < \frac{1}{2} n^{3/4} - \sqrt{n} \right\}$

$\mathcal{E}_3 = |C| > 4n^{3/4}$

If FAIL in step 6 then $\mathcal{E}_3$ occurs.

If FAIL in step 5 then $|S_{<l}| > \frac{n}{2}$ or $|S_{>u}| > \frac{n}{2}$.

Say $|S_{<l}| > \frac{n}{2}$ then $l > m$

which is the same as $\mathcal{E}_1$

& similarly $|S_{>u}| > \frac{n}{2}$ is the same as $\mathcal{E}_2$.

Hence, $\Pr(\text{FAIL}) \leq \Pr(\mathcal{E}_1) + \Pr(\mathcal{E}_2) + \Pr(\mathcal{E}_3)$

We'll show $\leq \frac{1}{4} n^{-1/4}$ $\leq \frac{1}{4} n^{-1/4}$ $\leq \frac{1}{2} n^{-1/4}$

Claim: $Pr(\mathcal{E}_1) \leq \frac{1}{4} n^{-1/4}$

Let $r_1, \ldots, r_{n^{3/4}}$ be elements of R.

Let $X_i = \begin{cases} 1 & \text{if } r_i \leq m \\ 0 & \text{o/w} \end{cases}$

Let $Y = \sum_{i=1}^{n^{3/4}} X_i$

$$\mathcal{E}_1 \Leftrightarrow Y < \frac{1}{2} n^{3/4} - \sqrt{n}$$

$$Pr(X_i = 1) = \frac{\frac{n-1}{2} + 1}{n} = \frac{1}{2} + \frac{1}{2n}$$

What's $Var(Y)$?

$$Var(Y_i) = E[X_i^2] - (E[X_i])^2$$

$$= \left(\frac{1}{2} + \frac{1}{2n}\right) - \left(\frac{1}{2} + \frac{1}{2n}\right)^2 = \frac{1}{4} - \frac{1}{4n^2}$$

$$Var(Y) = n^{\frac{3}{4}}\left(\frac{1}{2} + \frac{1}{2n}\right)\left(\frac{1}{2} - \frac{1}{2n}\right) = \frac{n^{3/4}}{4} - \frac{1}{4n^{5/4}}$$

$$< \frac{1}{4} n^{3/4}$$

More generally, $X_i \sim \text{Bernoulli}(p)$, $Y = \sum_{i=1}^{n} X_i$

$$E[X_i] = p, \quad E[Y] = np$$

$$Var(X_i) = E[(X_i - E[X_i])^2] = E[X_i^2] - (E[X_i])^2$$
$$= p - p^2 = p(1-p)$$

What's $Var(Y)$?

For mutually independent random variables $X_1, \ldots, X_n$,

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i).$$

Hence, $Var(Y) = np(1-p)$

Chebyshev's inequality: for any $a > 0$,

$$\Pr(|X - E[X]| \geq a) \leq \frac{Var(X)}{a^2}$$

Hence,

$$\begin{aligned}\Pr(\mathcal{E}_1) &= \Pr\left(Y < \tfrac{1}{2}n^{3/4} - \sqrt{n}\right) \\ &\leq \Pr\left(|Y - E[Y]| > \sqrt{n}\right) \\ &\leq \frac{Var(Y)}{n} \\ &< \frac{\frac{1}{4}n^{3/4}}{n} = \frac{1}{4}n^{-1/4}\end{aligned}$$

Similarly for $\mathcal{E}_2$.

Claim: $\Pr(\mathcal{E}_3) \leq \frac{1}{2} n^{-1/4}$

Let $\tilde{\mathcal{F}}_1 = \geq 2n^{3/4}$ of C are $\geq m$

$\tilde{\mathcal{F}}_2 = \geq 2n^{3/4}$ of C are $\leq m$

if $\mathcal{E}_3$ occurs then $\tilde{\mathcal{F}}_1$ &/or $\tilde{\mathcal{F}}_2$ occurs.

Let's look at $\tilde{\mathcal{F}}_1$.

If $\geq 2n^{3/4}$ of C are $\geq m$

then v is $\geq \left(\frac{n}{2} + 2n^{3/4}\right)^{th}$ smallest

then R has $\geq \frac{1}{2}n^{3/4} - \sqrt{n}$

from $\underbrace{\frac{1}{2}n - 2n^{3/4} \text{ largest of } S}$

Let $X_i = \begin{cases} 1 & \text{if } r_i \text{ is in} \\ 0 & \text{o/w} \end{cases}$

$\cancel{E[X_i] = \frac{\frac{1}{2}n^{3/4} - \sqrt{n}}{n} = \frac{1}{2n^{1/4}} - \frac{1}{\sqrt{n}}}$

$\cancel{E[X] = n^{3/4}\left(\frac{1}{2n^{1/4}} - \frac{1}{\sqrt{n}}\right) = \frac{n^{1/4}}{2}}$

$\cancel{= \frac{\sqrt{n}}{2} - n^{1/4}}$

$$E[X_i] = \frac{\frac{1}{2}n - 2n^{3/4}}{n} = \underbrace{\frac{1}{2} - \frac{2}{n^{1/4}}}$$

$$Y = \sum_{i=1}^{n^{3/4}} X_i, \quad E[Y] = n^{3/4}\left(\quad\right) = \frac{n^{3/4}}{2} - 2\sqrt{n}$$

$$\mathrm{Var}(Y) = n^{3/4}\left(\frac{1}{2} - \frac{2}{n^{1/4}}\right)\left(\frac{1}{2} + \frac{2}{n^{1/4}}\right)$$

$$= \frac{1}{4}n^{3/4} - 4n^{1/4} < \frac{1}{4}n^{3/4}$$

By Chebyshev's,

$$\Pr\left(\tilde{\mathcal{J}}_1\right) = \Pr\left(Y \geq \frac{n^{3/4}}{2} - \sqrt{n}\right)$$

$$\leq \Pr\left(\left|Y - E[Y]\right| \geq \sqrt{n}\right)$$

$$\leq \frac{\mathrm{Var}(Y)}{n} < \frac{\frac{1}{4}n^{3/4}}{n} = \frac{1}{4}n^{-1/4}$$

Similarly for $\tilde{\mathcal{J}}_2$.