CUCKOO : THEO $\forall$ pair of positions i, j $\forall c > 1$

if $m \geq 2 \cdot c n$ $\quad \mathbb{P}(i \overset{L}{—} j) = \dfrac{1}{mc^2}$

Base case: $L = 1$

$$\mathbb{P}(i — j) = \mathbb{P}\left(\exists k : \begin{array}{c} \underbrace{h_1(k) = j}_{=\alpha} \wedge \underbrace{h_2(k) = i}_{=\beta} \\ \vee \\ \underbrace{h_1(k) = i}_{=\gamma} \wedge \underbrace{h_2(k) = j}_{=\delta} \end{array}\right) =$$

edges

$$\leq n \cdot \mathbb{P}(k \text{ satisfy} \longrightarrow) =$$

$$= n \cdot \left[\begin{array}{c} \underbrace{\mathbb{P}(\alpha)}_{\leq \frac{1}{m}} \wedge \underbrace{\mathbb{P}(\beta)}_{\to \frac{1}{m}} \\ \vee \\ \underbrace{\mathbb{P}(\gamma)}_{\frac{1}{m}} \wedge \underbrace{\mathbb{P}(\delta)}_{\frac{1}{m}} \end{array} \right] \leq n \cdot \frac{2}{m^2}$$

$$\frac{2n}{m^2} \leq \frac{\cancel{m}}{c} \cdot \frac{1}{m^{\cancel{2}}} = \frac{1}{mc}$$

Inductive step:

$$\mathbb{P}(i \overset{L}{—} j) = \mathbb{P}(i \overset{L-1}{—} j \overset{L=1}{—} z) =$$

$$\leq \cancel{m} \cdot \frac{1}{mc^{L-1}} \cdot \frac{1}{\cancel{m} c} = \frac{1}{mc^L} \quad ✓$$

Prob I'm failing building a graph $\quad \dfrac{1}{c-1}$

$$\mathbb{P}(i \circlearrowleft \; j \circlearrowleft) \leq \mathbb{P}(i \circlearrowleft) =$$

$$= \mathbb{P}(\exists \, i \text{ node and edge } L : i \circlearrowleft^2) =$$

$$\leq \sum_{i} \sum_{l \geq 1} \mathbb{P}\left(C_i \geq c^{l}\right) = \cancel{m} \cdot \sum_{l \geq 1} \frac{1}{\cancel{m} c^{l}} = \sum_{l \geq 1} \left(\frac{1}{c}\right)^{l} = \frac{1}{c-1}$$

nodes are m

well known serie

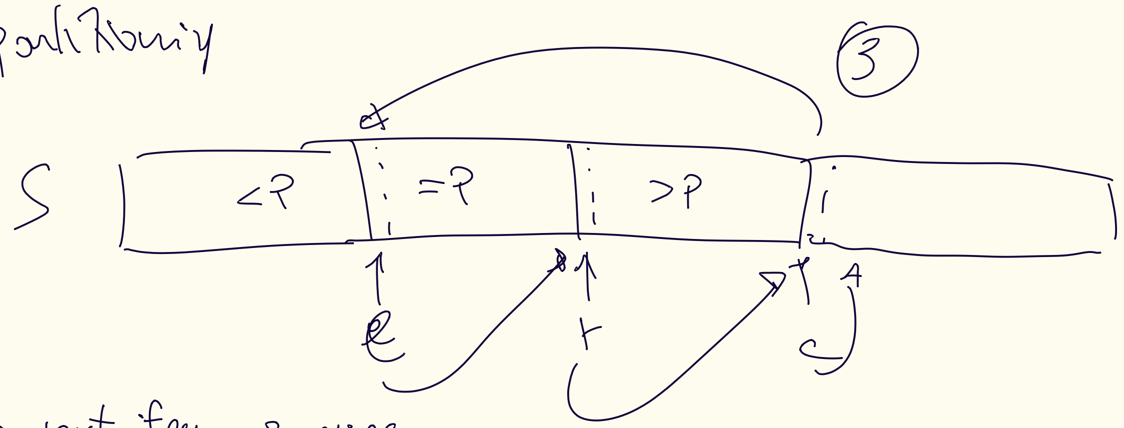
Snow plow

- Require U array of unsorted items
- Build a min-heap H over U items
- set $U = \emptyset$
- while $(H \neq \emptyset)$

min = extract min (root) from H
write min to output
next = read next item
if (next < min)
    put next in U
else
    put next in H
end if
end while

$$O\left(\frac{n}{B} \log_2 \frac{n}{2M}\right)$$

- 3-WAY partitioning



- c points to next item, 3 cases

1) $S[c] > p \Rightarrow c{+}{+}$
2) $S[c] == p \Rightarrow$ swap $S[r]$ with $S[c]$; r++; c++
3) $S[c] < p \Rightarrow$ swap $S[c]$ with $S[l]$
   swap $S[l]$ with $S[r]$
   swap $S[r]$ with $S[c]$

$l{+}{+};\ r{+}{+};\ c{+}{+}$

Prove QS is $n \log n$

$$X_{u,v} = \begin{cases} 1 & \text{iff } S[u] \text{ compared with } S[v] = \mathbb{P}_{u,v} \\ 0 & \text{otherwise} \end{cases}$$

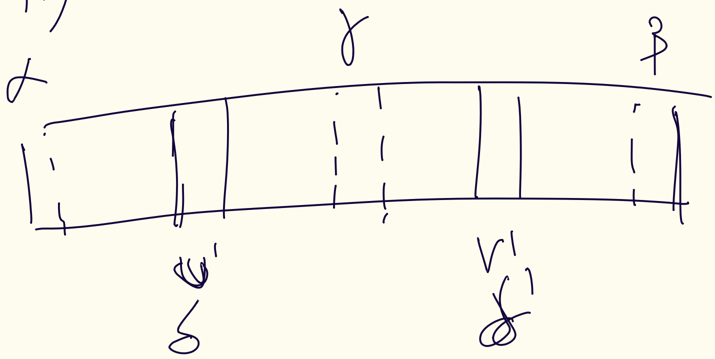
$$\mathbb{E}[\text{comparisons}] = \sum_{u=1}^{n} \sum_{v>u}^{n} 1 \cdot \mathbb{P}_{u,v} + 0 \cdot (1 - \mathbb{P}_{u,v}) =$$

$$= \sum_{u=1}^{n} \sum_{v>u}^{n} \mathbb{P}(u,v)$$

cases

consider

S sorted



(1) if pivot is α or β $\Rightarrow$ u' and v' not compared

$\Rightarrow$ not interesting

(2) if pivot is in between (γ) $\Rightarrow$ they go in $\neq$ partitions

(never compared)

(3) if pivot is u' or v'

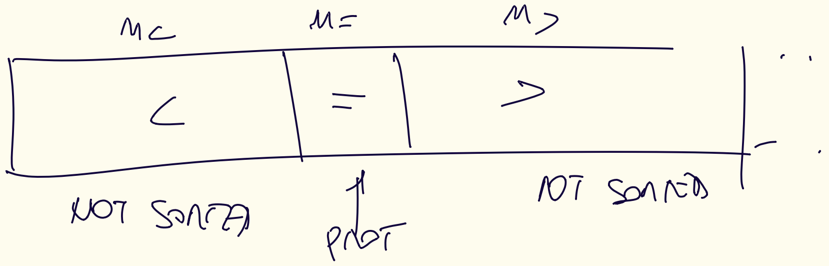
(δ and δ') $\Rightarrow$ they are compared

$$\Rightarrow \sum_{u=1}^{n} \sum_{u+1}^{n} \frac{2}{v'-u'+1} = \sum_{u' \neq v'}^{n} \frac{2}{v'-u'+1} \Rightarrow \text{harmonic } \cancel{\text{sum}} \text{ sum}$$

$$\simeq O(n \log_2 n)$$

CHOOSE K-RANKED item

- take arr
- take pivot
- 3-way partition



if PIVOT is $\boxed{=}$ => done

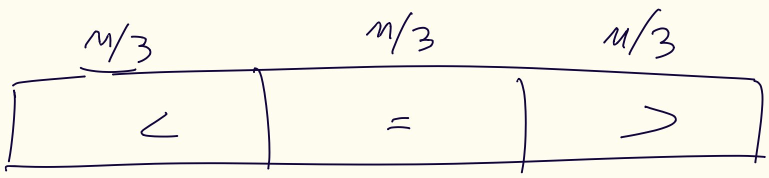
if PIVOT is in $\boxed{<}$ => RECURSE

if PIVOT is in $\boxed{>}$ => DISCARD some item

=> redefine $K = K - m_< - m_=$

=> COMPLEXITY

consider 3 SORTED



$$T(n) = \underbrace{O(n)}_{\text{partitioning}} + \underbrace{\frac{1}{3} T\left(\frac{2}{3} n\right)}_{\text{good case } (\boxed{=})} + \underbrace{\frac{2}{3} T(n-1)}_{\text{others bad cases}}$$

$$T(n-1) \simeq T(n)$$

$$T(n) - \frac{2}{3} T(n) = O(n) + \frac{1}{3} T\left(\frac{2}{3} n\right)$$

$$3 \cdot \frac{1}{3} T(n) = O(n)^{\cdot 3} + \frac{1}{3}^{\cdot 3} T\left(\frac{2}{3} n\right)$$

$$T(n) = O(n) + T\left(\frac{2}{3} n\right)$$

=> MASTER THEOREM

$$O(n)$$

BOUNDED QS => worst $\log_2 n$ recursive calls

(3 way partitioning

- if partition is short => not shoot with CANNOON of recursion

=> do insertion sort

BOUNDEDQS(S, i, j)

while $(j - i) \geq M_0$

 v = select a pivot

 swap S[i] with S[v]

 p = PARTITION(S, i, j)

 if $\left(p \leq \frac{i+j}{2}\right)$

  BOUNDEDQS(S, i, p-1)

  i = p+1

 else

  BOUNDEDQS(S, p+1, j)

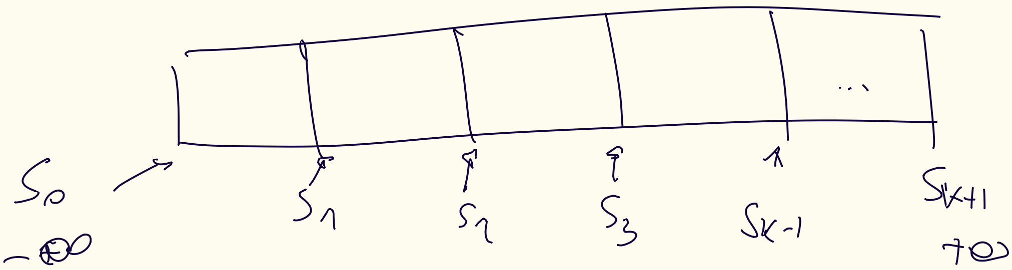
  j = p-1

 end if

end while

Insertion sort

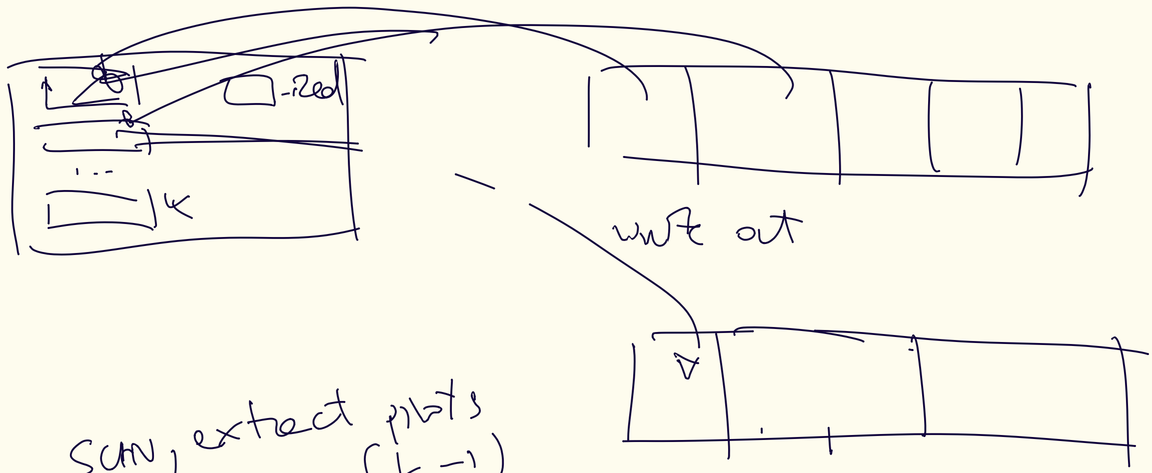
MULTIWAY QS



extract $(k-1)$ pivots $\Rightarrow$ k buckets $\qquad |B_i| = \frac{n}{k}$

sort them

Distribute items according to pivots



scan, extract pivots $(k-1)$

$$T(n) = O\left(\frac{n}{B}\right) + \sum_{i=1}^{k} T(|B_i|) \Rightarrow$$

$$T(n) = O\left(\frac{n}{B}\right) + \sum_{i=1}^{k} T\left(\frac{n}{k}\right) \Rightarrow$$

$$T(n) = O\left(\frac{n}{B}\right) + k\,T\left(\frac{n}{k}\right)$$

$$\Rightarrow T(n) = O\left(\frac{n}{B}\right) + \log_k\left(\frac{n}{B}\right)$$

setting $k = \Theta\left(\frac{M}{B}\right)$

$$\Rightarrow O\left(\frac{n}{B}\right) + \log_{\frac{M}{B}} \frac{n}{B}$$

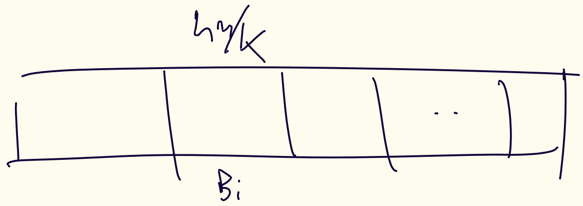
OVERSAMPLING : To get balanced partitions

=> extract more pivots then needed $(a+1)\cdot k - 1$

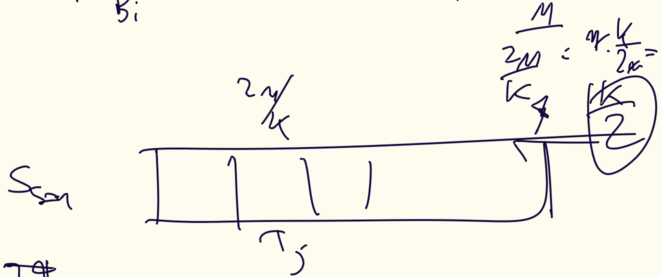
$= a = \Theta(\log_2 k)$ actually => $a = 12 \ln k$

$S_i = S[(a+1)\cdot i]$

Proof by contr.



$$\mathbb{P}[\exists B_i : |B_i| \geq \frac{4n}{k}] \leq \frac{1}{2}$$



$$\mathbb{P}[\exists B_i : |B_i| \geq \frac{4n}{k}] \leq \mathbb{P}(\exists \text{ bucket } T_j \text{ totally covered by } B_i)$$

$$\leq \mathbb{P}(\exists T_j : \text{contains} < (a+1) \text{ samples}) \leq$$ UNION BOUND

$$\leq \frac{k}{2}\cdot \mathbb{P}(\text{a } T_j \text{ contains} < (a+1) \text{ samples}) \quad (*)$$

$$\mathbb{P}(\text{sample occurs}) = 2n/k : n = \frac{2n}{k}\cdot\frac{1}{n} = \frac{2}{k}$$

$$\mathbb{E}[\text{samples}] = \mathbb{E}[X] = \mathbb{E}[X_i] \quad \text{where} \quad X_i = \begin{cases} 1 & \text{if a sample occurs in } T_j \\ 0 & \text{otherwise} \end{cases}$$

$$= \mathbb{E}[X_i) = \mathbb{E}\left[\sum_{i=1}^{x=(a+1)\cdot k-1} X_i\right] = \sum_{i=1}^{x} \mathbb{E}[X_i] = \frac{2}{k}\left[(a+1)k-1\right]$$

$$= 2(a+1) - \frac{2}{k} \geq \left(2(a+1)-1\right) \geq \frac{3}{2}(a+1)$$

$$E[X) \geq \frac{3}{2}(a+1) \qquad \frac{2}{3}\mathbb{E}[X] \geq (a+1)$$

$$\Rightarrow \mathbb{P}(T_j \text{ contains} < a+1 \text{ samples}) =$$

$$\mathbb{P}(\underbrace{T_j \text{ contains}}_{X} < \tfrac{2}{3}\mathbb{E}[X]) =$$

$$= \mathbb{P}(X < (1 - \underbrace{\tfrac{1}{3}}_{=\delta})\mathbb{E}[X]) \Leftarrow \text{CHERNOFF BOUND}$$

$$\leq e^{-\left(\frac{1}{3}\right)^2/2 \cdot \mathbb{E}[X]} = e^{-\left(\frac{1}{9}\right)/2 \cdot \mathbb{E}[X]} =$$

$$= e^{-\frac{1}{18}\cdot\mathbb{E}[X]} \leq e^{-\frac{1}{18}\cdot\frac{3}{2}(a+1)} =$$

$$= e^{-\frac{1}{12}(\overset{a+1}{\cancel{a+1}})} = e^{-\frac{1}{\cancel{12}}\cdot\cancel{12}\ln k} =$$

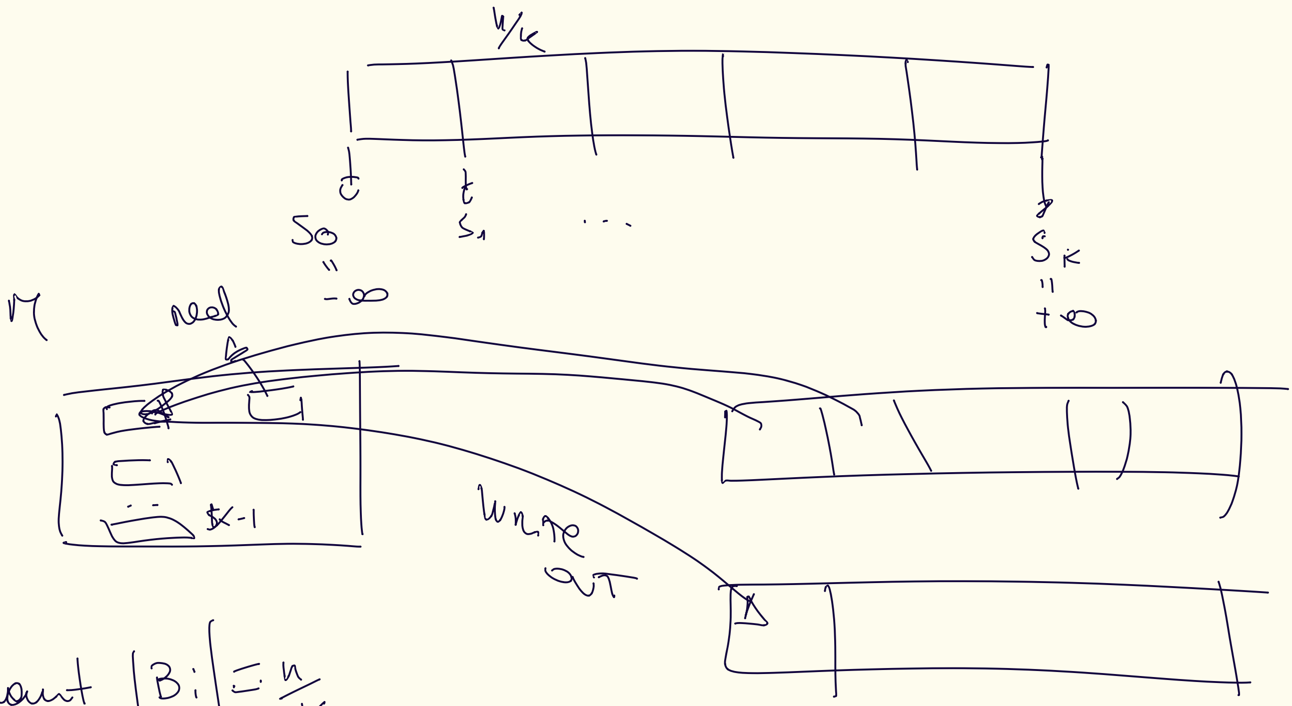
$$= e^{-\ln k} = \frac{1}{e^{\ln k}} = \frac{1}{k}$$

$\Rightarrow$ the probability I was searching was $\frac{k}{2}$.

$$\Rightarrow \frac{\cancel{k}}{2}\cdot\frac{1}{\cancel{k}} = \frac{1}{2}$$

MULTI-WAY QS

- extract more pivots than needed $(k-1)$
- sort them
- Distribute items according to them
- $S_{i-1} < \text{Bucket} \leq S_i$



want $|B_i| = \frac{n}{k}$

SCAN, SORT PIVOTS

$$T(n) = O\left(\frac{n}{B}\right) + \sum_{i=1}^{k} T(|B_i|)$$

$$T(n) = O\left(\frac{n}{B}\right) + k \cdot T\left(\frac{n}{k}\right) =$$

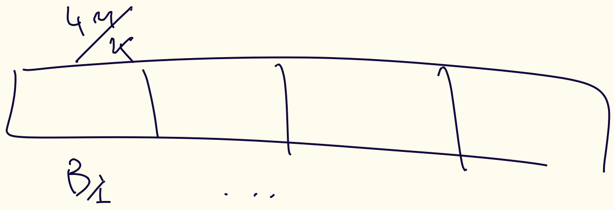
$$= O\left(\frac{n}{B}\right) + \log_k\left(\frac{n}{B}\right) = \quad \text{setting } k = \frac{M}{B}$$

$$= O\left(\frac{n}{B} + \log_{\frac{M}{B}} \frac{n}{B}\right)$$

TO get BALANCED PARTITIONS

$\Rightarrow$ extract $(a+1)\cdot k-1$ pivots

S



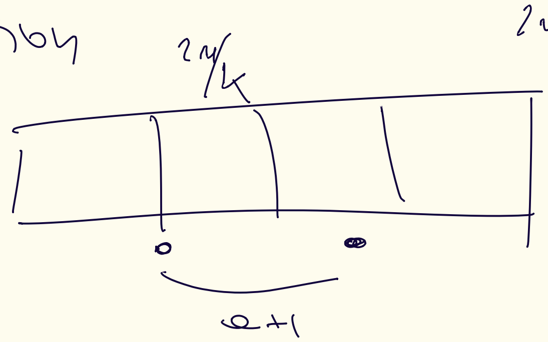
$$|B_i| = \frac{4m}{k}$$

To prove Balanced partition

$\Rightarrow$ proof by contradiction

$$\frac{m}{2m/k} = m \cdot \frac{k}{2m} = \frac{k}{2}$$

considered sorted S



$$|B_j| = \frac{2m}{k}$$

$$\mathbb{P}\left(\exists\, B_i : |B_i| \geq \frac{4m}{k}\right) \leq \frac{1}{2}$$

$\mathbb{P}$[sample occurs $= \frac{k}{2}$

$$\mathbb{P}\left(\exists\, B_i : |B_i| \geq \frac{4m}{k}\right) \leq \mathbb{P}\left(\exists\, T_j : |T_j| \text{ is covered by } B_i\right) \leq$$

$$\leq \left(\exists\, T_j : B_j \text{ contains} < (a+1) \text{ samples}\right) \leq \text{UNION BOUND}$$

$$\leq \frac{\cancel{k}}{\cancel{2}} \,\#T_j \cdot \mathbb{P}\left(a\, T_j \text{ contains} < (a+1) \text{ samples}\right)$$

$$X_i = \begin{cases} 1 & \text{if a sample occurs in } T_j \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbb{E}[\text{samples}] = \mathbb{E}[X_i] = \mathbb{E}\left[\sum_{i=1}^{x=(a+1)\cdot k-1} (X_i)\right] =$$

$$= \sum_{x=1}^{X} \mathbb{E}(X_i) = \frac{k}{2} \cdot \left[(a+1)\cdot k - 1\right] =$$

$$= 2(a+1) - \frac{k}{2} \geq 2(a+1) - 1 \geq \frac{3}{2}(a+1)$$

$$\mathbb{E}[X] \geq \frac{3}{2}(a+1)$$

$$\frac{2}{3}\mathbb{E}[X] \geq (a+1)$$

→ FORGET FOR THE MOMENT

$$\Rightarrow \frac{k}{2}\,\mathbb{P}(\underbrace{a\ T_j \text{ contain} < (a+1)}_{X} \text{ samples}) =$$

$$= \mathbb{P}(X < (a+1)) = \mathbb{P}(X \leq \frac{2}{3}\mathbb{E}[X]) =$$

$$= \mathbb{P}[X \leq (1 - \underbrace{\tfrac{1}{3}}_{\delta})\mathbb{E}[X]] \leq \text{Chernoff Bound}$$

$$e^{-\left(\frac{1}{3}\right)^2/2 \cdot \mathbb{E}[X]} = e^{-\frac{1}{18}\mathbb{E}[X]} \leq e^{-\frac{1}{\cancel{18}\,6}\cdot\frac{\cancel{3}}{2}(a+1)} =$$

$$= e^{-1/12\cdot(a+1)} \quad a+1 = 12\ln k \quad = e^{-\frac{\cancel{12}}{\cancel{12}}\ln k} =$$

$$= e^{-\ln k} = \frac{1}{e^{\ln k}} = \frac{1}{k}$$

thus $\frac{\cancel{k}}{2}\cdot\mathbb{P}(X < \ldots) =$

$$= \frac{\cancel{k}}{2}\cdot\frac{1}{\cancel{k}} = \frac{1}{2} \quad \checkmark$$

LOWER BOUND SORTING STRINGS

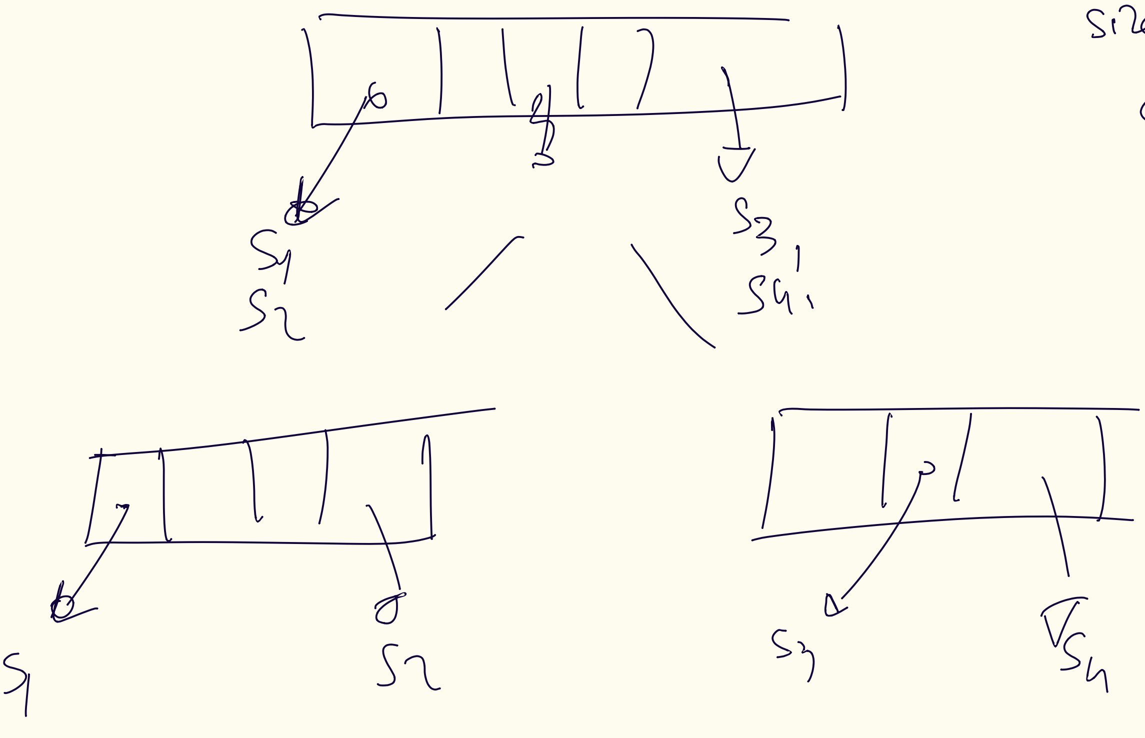
1) Sort string to 1^{st} digit $= \Omega(n \log_2 n)$
(is sorting

2) I've to consider all the distinguish prefixes of the string to compare

$$\Rightarrow \sum_{S} d_S = d$$

$$\Rightarrow \Omega(d + n \log_2 n)$$

MSD: $\longrightarrow$; string move left



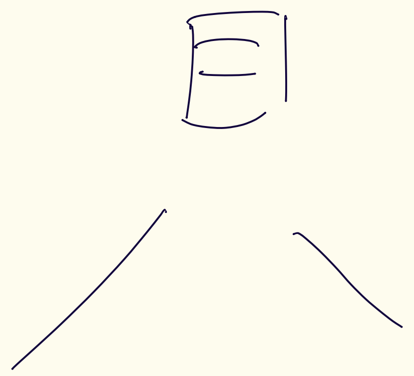
size of array
$\sigma = |\Sigma|$

stop when 1 string per bucket

I've to consider the array (which have a dimension $= \sigma$)
and I'm making ~~path~~ the confront for distinguish prefixes

$$\Rightarrow \Theta(d \cdot \sigma) \Rightarrow \text{best lower bound}$$

Hashing $\Rightarrow$ Trie



~~#~~ 1 hash table per node

edges

$\Rightarrow$ the thing is I make $n \lg n$ comparisons but n in this case is #edges

$$O(\underbrace{\#\text{edges}}_{e_v} \lg \underbrace{\#\text{edges}}_{e_v})$$

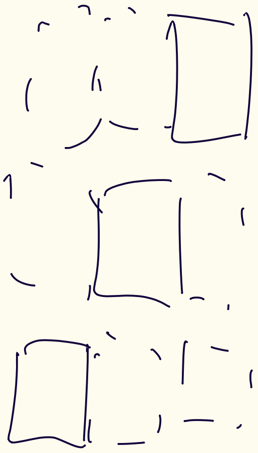
$$O(e_v \lg e_v) \leq O(\cancel{e_v \lg \sigma})$$

$$= O\left(\sum_s e_v \lg e_v\right) \leq \left(\sum_s e_v \lg \sigma\right) =$$

no more than sigma

$$= O(d \cdot \lg \sigma)$$

LSD



$\Rightarrow$ counting sort (n + max up

$\Rightarrow$ in this case

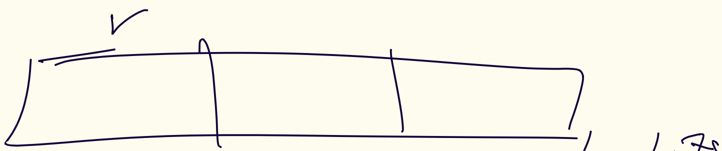
I make

$O(L(n+B))$

confronting 2 chars at time

$O\left(\frac{L}{r}(n+B^r)\right)$

$\Rightarrow$ consider bites instead of digits



($B = 2$ in bits)

$O\left(\frac{b}{r}(n+2^r)\right)$ want r that minimize the function

$2^r > n \Rightarrow \frac{b\,2^r}{r}$

$r > \lg n$

$\Rightarrow \frac{b\,2^{\lg n}}{\lg n} = \frac{b \cdot n}{\lg n} = \frac{L}{\lg n}$ ✓

but this is PROPER THE cost

multikey qs

$MKQ(S, i)$

if $|R| \leq 1$

return R

P = choose a pivot

else

$R_< : \{ s \in S : S[i] < P(i) \}$

$R_= : \{ s \in S : S[i] = P(i) \}$

$R_> : \{ s \in S : S[i] > P(i) \}$

$A = MKQ(R_<, i)$

$B = MKQ(R_=, i+1)$

$C = MKQ(R_<, i)$

$\Rightarrow$ return $A \circ B \circ C$

complexity : ordinary sorted partitions

Then that of string go in $R_<$ or $R_>$

i doesn't increase

$\Rightarrow$ occurs $n \log n$ times

1 string in $R_=$:

i++;

$\Rightarrow i \leq \sum_S d_S \Leftrightarrow i \leq L$

$\Rightarrow (n \log n + d)$

$$\Omega\;\frac{\frac{\cancel{N}}{DB}\cdot \log_{\frac{M}{DB}}\frac{N}{M}}{\frac{\cancel{N}}{DB}\cdot \log_{\frac{M}{B}}\frac{N}{M}} = \frac{\cancel{\log_2\frac{N}{M}}}{\log_2\frac{M}{DB}}\cdot\frac{\log_2\frac{M}{B}}{\cancel{\log_2\frac{N}{M}}} = \frac{\log_2\frac{M}{B}}{\log\frac{M}{B}-\log D} =$$

$$= \frac{\cancel{\log_2\frac{M}{B}}}{\cancel{\log\frac{M}{B}}\left(1-\frac{\log_2 D}{\log_2\frac{M}{B}}\right)} = \frac{1}{1-\log_{\frac{M}{B}} D} \geq 1$$

More disk i put
Denom grows
=> primal
with 1 disk

Lower bound for sort

new page $\binom{M}{B}\cdot B!$ $\quad\#\ \frac{N}{B}$

old $\binom{M}{B}$ $\quad\#\ t-\frac{N}{B}$

$$\left[\cancel{\binom{M}{B}}\,B!\right]^{\frac{N}{B}}\cdot\binom{M}{B}^{t-\cancel{\frac{N}{B}}} \geq N!$$

$$B!^{\frac{N}{B}}\cdot\binom{M}{B}^{t} \geq N!$$

$$\log_2 (B!)^{\frac{N}{B}} + \log_2\binom{M}{B}^{t} \geq \log_2 N!$$

$$\cancel{B}\cdot\frac{N}{\cancel{B}}\log_2 B + t\cdot B\log_2\frac{M}{B} \geq N\log N$$

$$n \lg_2 B + t B \lg \frac{M}{B} \geq n \lg n$$

$$t B \lg \frac{M}{B} \geq n \lg n - n \lg_2 B$$

$$t \frac{\left(\cancel{B \lg \frac{M}{B}}\right)}{\cancel{B \lg \frac{M}{B}}} \geq \frac{n \left(\lg_2 \frac{n}{B} \right)}{B \lg \frac{M}{B}} \qquad t \geq \frac{n}{B} \lg_{\frac{M}{B}} \frac{n}{B}$$

```
Building SA (char *T, char *n, char **SA)
    for (i=0 ; i<n ; i++) {
        SA[i] = T+i }
    qsort (SA, n, sizeof (char *), suffix_cmp)

suffix_cmp (char **p, char ** e)
    return strg_cmp (*p, *e)
```