Q1

(a) $E(X_N)_1 = 0p_0 + 1p_1 + 2p_2 + \cdots Np_N = \frac{c}{1} + 2\cdot\frac{c}{2} + 3\cdot\frac{c}{3} + \cdots N\cdot\frac{c}{N} = NC$

$\because$ this is distribution,

$\therefore \sum_N p_k = \sum_N \frac{c}{k} = 1 \Rightarrow C(1+\frac{1}{2}+\frac{1}{3}+\cdots\frac{1}{N}) = 1 \Rightarrow C = \frac{1}{\sum_{k=1}^{N}\frac{1}{k}}$

$\therefore E(X_N)_1 = NC = \frac{N}{\sum_{k=1}^{N}\frac{1}{k}}$

(b) The average $E(X_N)$ of uniform distribution

$$E(X_N)_2 = 0 * p_0 + 1 * p_1 + 2 * p_2 + \cdots N * p_N = \frac{1}{N} + 2*\frac{1}{N} + 3*\frac{1}{N} + \cdots + N*\frac{1}{N}$$

$= \frac{1+N}{2}$ when $N=10$, $E(X_N)_1 = \frac{10}{\sum_{k=1}^{10}\frac{1}{k}} = 3.414$

$E(X_N)_2 = \frac{1+10}{2} = 5.5$

(c) when $E(X_{N^*})_1 = E(X_{N^*})_2$, $\frac{N^*}{\sum_{k=1}^{N^*}\frac{1}{k}} = \frac{1+N^*}{2} \Rightarrow N^* = 1$

Since $E(X_{N^*})_1$ is divergent and the difference $E(X_{N^*})_2$ got larger after intersection.

$N^* = 1$ is a unique solution for equality in a range.

(d)

<i> if it's sequential search:

For each element, we'll keep searching until we find a element greater than the one we're searching for $\Rightarrow$ average number of comparisons: N/2

<ii> if it's binary search:For each element, the average number of comparisons is the total layer number of binary tree, which is log(N)

(e)

Record is present: On average, $\log_2 N$ comparisons, as heaps are tree-based structures (binary heap).

Record is not present: Up to N comparisons, as each node must be inspected to ensure the record isn't in the heap.

Q2

we have n=23,
with average storage utility 0.69, we have 23*0.69=15.87 ≈ 16
(i) Level1:
average number of nodes: 16
average number of key entries: 16*(16-1) = 16*15 = 240 average number of children pointers: 16*16 = 256
(ii) Level3:
average number of nodes: 163 = 4096
average number of key entries: 163 ∗ 15 = 61440 average number of children pointers: 164 = 65536
(iii) Level4:
average number of nodes: 164 = 65536
average number of key entries: 164 ∗ 15 = 983040 average number of children pointers: 165 = 1048576
(iv) height of tree is 2:
average number of entries = level0_key + level1_key + level2_key = 15 + 16*15 + 16*16*15 = 4095
(v) height of tree is 3:
average number of entries = level0_key + level1_key + level2_key + level3_key= 15 + 16*15 + 16*16*15 + 16*16*16*15 = 65535
(vi) height of tree is 4:
average number of entries = level0_key + level1_key + level2_key + level3_key + level4_key = 15 + 16*15 + 16*16*15 + 16*16*16*15 + 16*16*16*16*15 = 1048575 ⇒ average total number of entries hat such a tree:
we first have each node have a maximum of n children → 23 here
with average storage utility of B tree 0.69, we have average children we have is 0.69n → 0.69*23 ≈ 16 here
for i^{th} layer, we have $(0.69n)^i * (0.69n - 1) = 16^i * 15$ key entries
Therefore, for the tree with height h, we have:

$$\sum_{i=0}^{h}\left((0.69n)^i \cdot (0.69n-1)\right) = \sum_{i=0}^{h}(16^i \cdot 15) \Rightarrow \sum_{i=0}^{h}\left((0.69n)^i \cdot (0.69n-1)\right)$$

$$= (0.69n-1)\sum_{i=0}^{h}(0.69n)^i$$

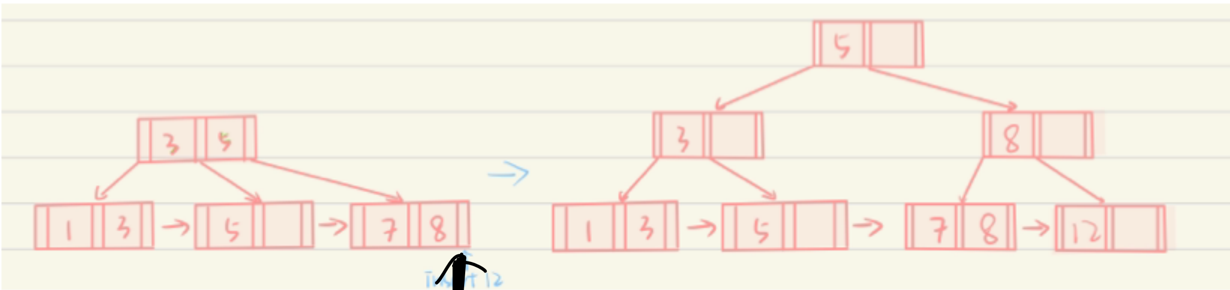
$$= (0.69n)^{h+1} - 1$$

Q3

Insert key 3:

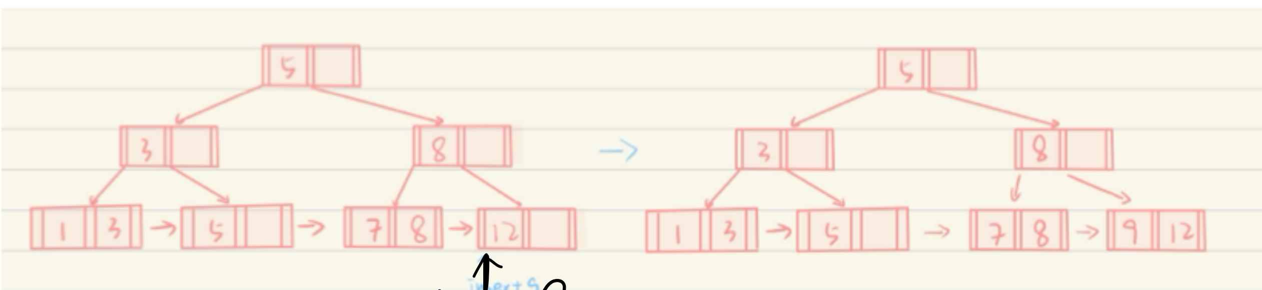
insert 3 (overflow, split)

Insert key 12:



insert 12 (overflow split)

Insert key 9:



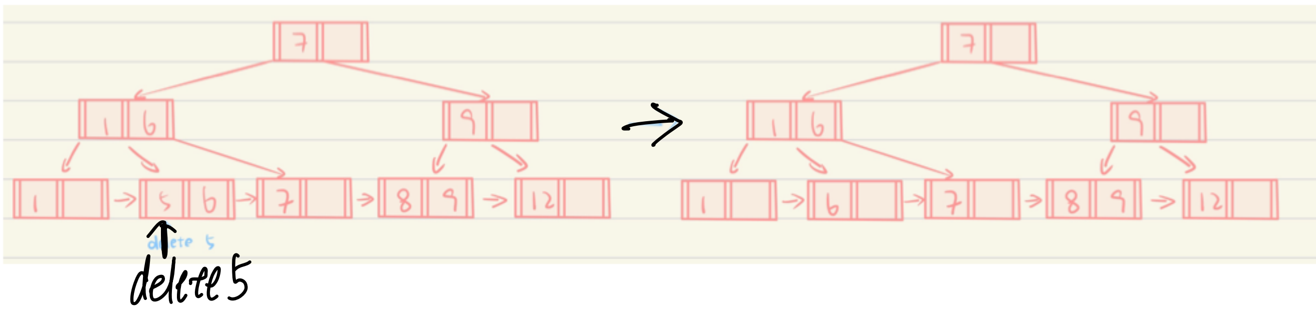
insert 9

Insert key 6:

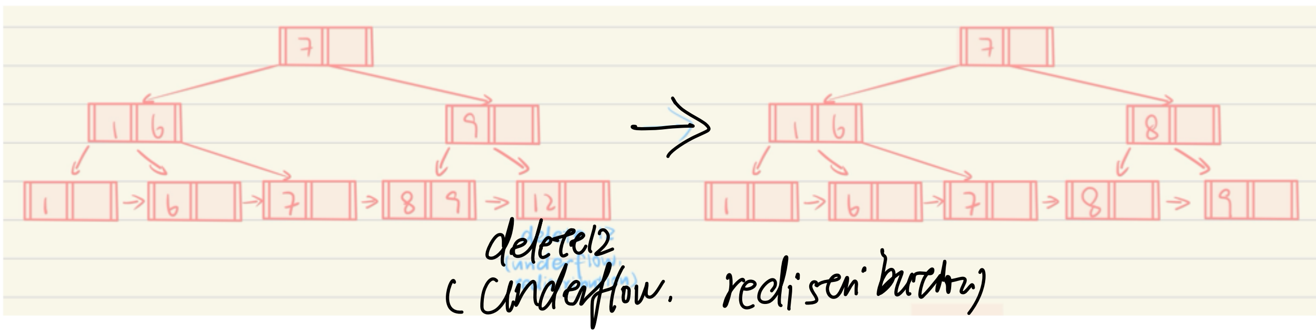
insert 6

Q4

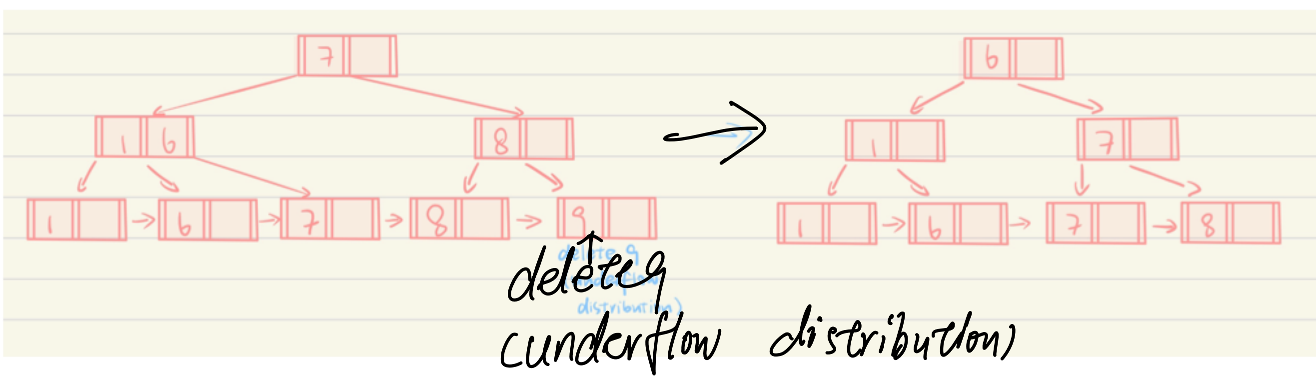
Delete Key5:



Delete Key12:



Delete Key9:



Q5

With K: total number of items in tree, n: maximum capacity of a node; N: random number of nodes in the tree
ρ: random storage utilization of the tree; f: minimum fullness factor, and f= 34 here
we have: ρ=K /Nn. minimum number of nodes is K/n, maximum number of nodes is 4K/3n $\Rightarrow N \sim U\left(\frac{K}{n}, \frac{4K}{3n}\right)$

(1) $E(\rho) = E\left(\frac{K}{Nn}\right) = \frac{K}{n} E\left(\frac{1}{N}\right) = \left(\frac{K}{n}\right) \times \left(\frac{3n}{K}\right) \int_{\frac{K}{n}}^{\frac{4K}{3n}} \left(\frac{1}{t}\right) dt$

$= 3 \cdot \left(\ln\left(\frac{4K}{3n}\right) - \ln\left(\frac{K}{n}\right)\right) = 3 \cdot \ln\left(\frac{4}{3}\right) = 86.3\%$

(2) $\sigma_f^2 = f - \left(\frac{f}{1-f}\right)^2 \left[\ln\left(\frac{1}{f}\right)\right]^2 = \frac{3}{4} - \left(\frac{\frac{3}{4}}{\frac{1}{4}}\right)^2 \left[\ln\left(\frac{4}{3}\right)\right]^2 = 0.02515$

$\Rightarrow sd(\rho) = \sqrt{0.02515} = 0.72$

Q6

(1)

Record#	2305	1168	2580	4871
Hash Index	1	0	4	7
	1620	2428	3943	4750
	4	4	7	6
Record#	5659	1821	1074	7115
Hash Index	3	5	2	3
	6975	4981	9280	
	7	5	0	

⇒

Hash Index	0	1	2	3
Records#	1168	2305	1074	5659
	9280			7115
Hash Index	4	5	6	7
Records#	2580	1821	4750	4871
	1620	4981		3943
	2428			6975

We have: bucket4 and bucket7 are overflow.

∴ average # block access for a random record retrieval on Part# is:

$$2 \times \frac{2}{15} + 1 \times \frac{13}{15} = 1.133$$

(2)

<step i> insert 2305, 1168. (2305) mod 128 = 1 = (00001)2. (1168) mod 128 = 16 = (10000)2
we have d=0, d'=0 with 2 elements in one local buckets
<step ii> insert 2580, 4871. (2580) mod 128 = 20 = (10100)2. (4871) mod 128 = 7 = (00111)2
we have d=1, d'=1 with 4 elements in two local buckets
<step iii> insert 5659, 1821. (5659) mod 128 = 27 = (11011)2. (1821) mod 128 = 29 = (11101)2 we have d=2,

Record#	2305	1168	2580	4871	5659	1821
Hash Index	1	16	20	7	27	29
Hash Value	00001	10000	10100	00111	11011	11101

⇒ global depth: d=2

global index:	Records:		local depth:
00 / 01	Record1: 2305	Record4: 4871	d' = 1
10	Record2: 1168	Record3: 2580	d' = 2
11	Record5: 5659	Record6: 1821	d' = 2