ex12.pdf

TU München, Fakultät für Informatik
Lehrstuhl III: Datenbanksysteme
Prof. Dr. Thomas Neumann

Query Optimization
12. Exercise
Due February 15, 2021, 9 AM

#number = 100 = B
1 = 10 = b
0 = 90 = B-b

Exercise 1
Given a Relation being consecutively stored on 100 pages on Disk. Assuming 10 random pages to be read:

- Plot the probability of j pages being skipped for 0 <= j <= 90
- Estimate the expected distance between two of the pages to be read
- Estimate the expected distance between the beginning of the Relation and the last page to be read
- Estimate the expected distance between the first and the last page to be read

Plot function

$f(j) = B_b^B(j) = \frac{\binom{100-j-1}{10-1}}{\binom{100}{10}} = \frac{\binom{99-j}{9}}{\binom{100}{10}}$

$j \in [0, 90]$

$\bar{B}_b^B = \frac{B-b}{b+1} = \frac{90}{11}$

$\bar{B}_{tot}(B,b) = \frac{Bb+b}{b+1} = \frac{100\times10+10}{10+1} = \frac{1010}{11}$

$\bar{B}_{1\text{-}Span}(B,b) = \frac{Bb-B+2b}{b+1} = \frac{100\times10-100+2\cdot10}{11} = \frac{920}{11}$

Exercise 2
Given the following histogram of an integer attribute *R.a*:

bucket	[0, 20)	[20, 40)	[40, 60)	[60, 80)	[80, 100)
count	1	3	4	2	0

Estimate the number of elements for which R.a >= 55 holds true.

$\frac{80-55}{80-40} \times 4 + 2 + 0$

$= \frac{5}{20} \times 4 + 2 + 0$

$= 1 + 2 + 0$

$= 3$

Exercise 3
Given the following histograms of an integer attributes R.a and S.b:

R.a bucket	[0, 20)	[20, 40)	[40, 60)	[60, 80)	[80, 100)
count	1	3	4	2	0

S.b bucket	[0, 10)	[10, 20)	[20, 40)	[40, 50)	[50, 100)
count	2	4	1	6	4

Estimate the number of elements for which R.a = S.b holds true.

$1\times(2+4) + 3\times1 + 4\times\frac{1}{2}\times6 + \left(4\times\frac{1}{2}+2+0\right)\times4$

$= 1\times6 + 3\times1 + 2\times6 + 4\times4$

$= 6 + 3 + 12 + 16$

$= 37$