Solutions for Exercise Sheet 14

Richter, Yannick
MTK 03741982
ge78tup@mytum.de

Rodrigues, Diogo
MTK 03770446
diogo.rodrigues@tum.de

TUM – Query Optimization 2022/23
10th Febuary 2023

Our solutions for Exercise Sheet 14.

Exercise 1

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

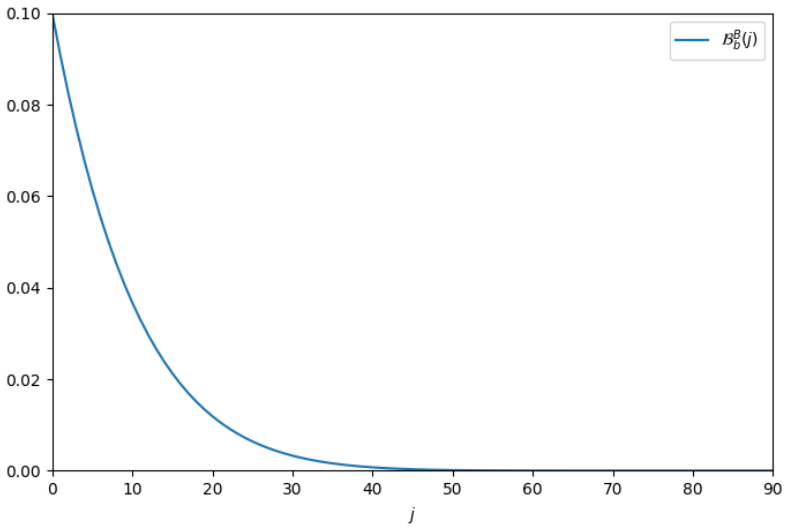
- Plot the probability of j pages being skipped for $0 \leq j \leq 90$
- Estimate the expected distance between two of the pages to be read
- Estimate the expected distance between the beginning of the Relation and the last page to be read
- Estimate the expected distance between the first and the last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

---

Item 1

$$\mathcal{B}_b^B(j) = \frac{\binom{B-j-1}{b-1}}{\binom{B}{b}}$$

$B = 100$, $b = 10$



Item 2

$$\overline{\mathcal{B}}_b^B = \sum_{j=0}^{B-b} j \cdot \mathcal{B}_b^B(j) = \frac{B-b}{b+1}$$

8.181818181818182

Item 3

$$\overline{\mathcal{B}}_{tot}(B,b) = \frac{Bb+b}{b+1}$$

$100 - 8.181818181818182 = 91.8181818182$

Item 4

$$\overline{\mathcal{B}}_{1-span}(B,b) = \frac{Bb-B+2b}{b+1}$$

$100 - 2 \cdot 8.181818181818182 = 83.6363636364$

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 $R.a \geq 55$ holds true.

$$\begin{aligned}\sigma &= \frac{\sum_{b \in B: c \in b} \frac{\max(b)-c}{\max(b)-min(b)} H_A(b) + \sum_{b \in B: \min(b) > c} H_A(b)}{\sum_{b \in B} H_A(b)} \\ &= \frac{\frac{60-55}{60-40} \cdot 4 + (2+0)}{1+3+4+2+0} \\ &= \frac{0.25 \cdot 4 + 2}{10} \\ &= \frac{3}{10}\end{aligned}$$

Therefore, the number of elements for which $R.a \geq 55$ is estimated at $N \cdot \sigma = 10 \cdot (3/10) = 3$.

Exercise 3

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

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

b'	[0, 10)	[10, 20)	[20, 40)	[40, 50)	[50, 60)	[60, 80)	[80, 100)
count	$1 * 0.5 * 2$	$1 * 0.5 * 4$	$3 * 1$	$4 * 0.5 * 6$	$4 * 0.5 * 4 * 0.2$	$2 * 4 * 0.4$	$0 * 4 * 0.4$
count	1	2	3	12	1.6	3.2	0

$A_1 = A_2$

$$\sigma = \frac{\sum_{b_1 \in B_1, b_2 \in B_2, b' = b_1 \cap b_2 : b' \neq \emptyset} \frac{\max(b')-\min(b')}{\max(b_1)-\min(b_1)} H_{A_1}(b_1) \frac{\max(b')-\min(b')}{\max(b_2)-\min(b_2)} H_{A_2}(b_2)}{\sum_{b_1 \in B_1} H_{A_1}(b_1) \sum_{b_2 \in B_2} H_{A_2}(b_2)} = \frac{22,8}{170} \approx 0,13$$