Solutions for Exercise Sheet 13

Richter, Yannick
MTK 03741982
ge78tup@mytum.de

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

TUM – Query Optimization 2022/23
3rd Febuary 2023

Our solutions for Exercise Sheet 13.

Exercise 1

Formally prove the following:

- $\binom{n}{k} = \binom{n}{n-k}$
- $\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$

The definition of binomial coefficients is:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Item 1

$$\binom{n}{n-k} = \frac{n!}{(n-k)!(n-(n-k))!} = \frac{n!}{(n-k)!k!} = \frac{n!}{k!(n-k)!} = \binom{n}{k}$$

Item 2

$$\begin{aligned}
\binom{n-1}{k} + \binom{n-1}{k-1} &= \frac{(n-1)!}{k!(n-1-k)!} + \frac{(n-1)!}{(k-1)!(n-1-(k-1))!} \\
&= \frac{(n-1)!}{k!(n-k-1)!} + \frac{(n-1)!}{(k-1)!(n-k)!} \\
&= \frac{(n-1)! \cdot (n-k)}{k!(n-k-1)! \cdot (n-k)} + \frac{(n-1)! \cdot k}{(k-1)!(n-k)! \cdot k} \\
&= \frac{(n-1)! \cdot (n-k)}{k!(n-k)!} + \frac{(n-1)! \cdot k}{k!(n-k)!} \\
&= \frac{(n-1)! \cdot (n-k) + (n-1)! \cdot k}{k!(n-k)!} \\
&= \frac{(n-1)! \cdot [(n-k)+k]}{k!(n-k)!} \\
&= \frac{(n-1)! \cdot n}{k!(n-k)!} = \frac{n!}{k!(n-k)!} = \binom{n}{k}
\end{aligned}$$

Exercise 2

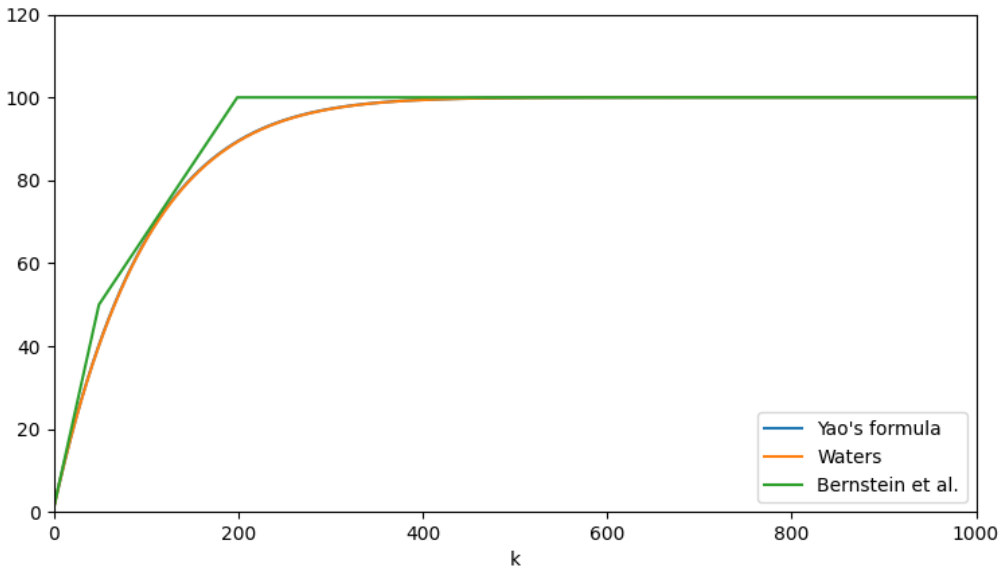
Given a relation with 3 pages and two tuples per page, compute the average number of accessed pages when reading 2 distinct tuples.
$N = 6$, $m = 3$, $B = 2$, $k = 2$
$\bar{\mathcal{Y}}_n^{N,m}(k) = m * \mathcal{Y}_n^M(k)$
$p = \frac{\binom{N-n}{k}}{\binom{N}{k}} = 0.4$
$\overline{\mathcal{Y}}_n^{N,m}(k) = m * \mathcal{Y}_n^M(k) = 3 * 0.6 = 1.8$

Exercise 3

Given a relation with 3 pages and two tuples per page, compute the average number of accessed pages when reading 4 not necessarily distinct tuples.
$N = 6$, $m = 3$, $B = 2$, $k = 4$
$\bar{Cheung}_n^{N,m}(k) = m * Cheung_n^M(k)$
$\tilde{p} = \frac{\binom{N-n+k-1}{k}}{\binom{N+k-1}{k}} = 0.278$
$\overline{Cheung}_n^{N,m}(k) = m * Cheung_n^M(k) = 3 * 0.722 = 2.167$

Exercise 4

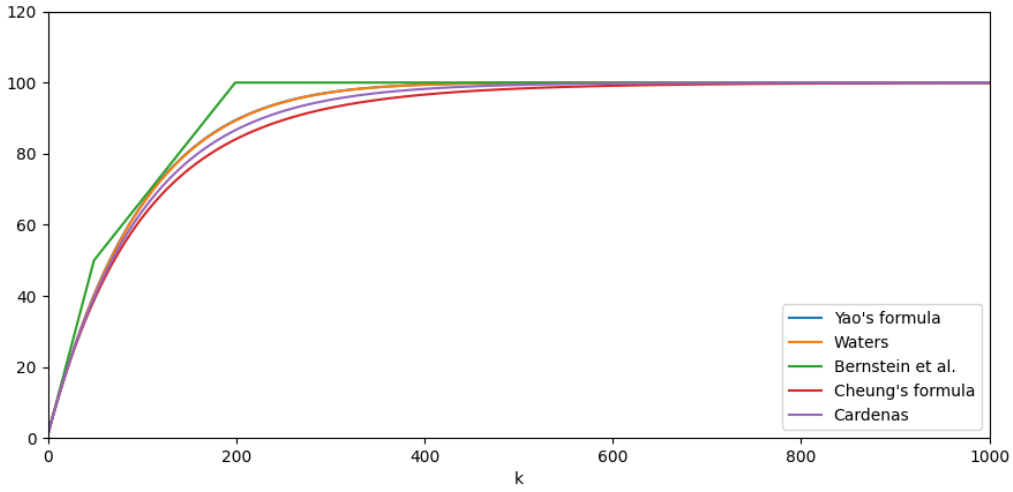
Given a relation with 100 pages with 10 tuples each, plot the expected number of accessed pages when reading 1, 2, ..., 1000 distinct tuples using Yao's formula. Also plot the approximations of Bernstein et al. and Waters. Submit all three graphs in a single chart.
$N = 1000$, $m = 100$, $B = 10$, $k = 4$



Note that Yao's formula and Waters overlap in this plot.

Exercise 5

Include the expected number of accessed pages when reading 1, 2, ..., 1000 not necessarily distinct tuples using Cheung's formula into your plot from the previous exercise. In addition, include the approximation of Cardenas into your plot. Submit a separate graphic including all five graphs.


120
100
80
60
40
20
0
0
200
400
600
800
1000
k
Yao's formula
Waters
Bernstein et al.
Cheung's formula
Cardenas