# ELEN 6889 Homework#2

I implement the first option, Operator Reordering.

My workload is a text file containing 10000 numbers in range (1,100), which are generated randomly.

I choose “filter out all even number of inputs” as Operator A, whose selectivity is fixed at 0.5. And I choose “filter out all number larger than a” as Operator B, where a is the variable to change selectivity of B in range (0.01,1).

Then I calculate the throughput for not-reordered case by 1/costA+secA*costB. For reordered case, just exchange the index A and index B. Next I normalize the not-reordered throughput to 1, and alter the reordered throughput in same scale so that the output graph will look almost the same with the paper’s.

To run the code, unzip the file into spark/bin and run *./spark-submit hw2.py*

Following is the output graph.