Reservoir Sampling

Instructor: Kunpeng Zhang (kzhang6@uic.edu)
TA: Minghong Xu (mxu29@uic.edu)

The Deadline: May 4, 2015 (Saturday)

Steps to implement sequential reservoir sampling algorithm. The inputs of your algorithm are: a text file containing many integers, and a number indicating how many numbers you are sampling.

1) Create an array reservoir[0..k-1] and copy first k items of stream[0..n-1] to it.
2) Now one by one consider all items from $(k+1)^{th}$ item to n^{th} item: Generate a random number from 0 to *i* where *i* is the index of current item in stream. Let the generated random number be *j*.
3) If *j* is in range 0 to k-1, replace reservoir[*j*] with stream[*i*]

Notes:
You need to randomly generate at least 10,000 integers. They can be duplicated.

Requirements:

- Your codes must be readable and clean.
- When you submit your codes through blackboard, you need to put all source codes (.java files, NOT jar files) and some other optional files (e.g., a readme file) into one folder and name that folder as <YOUR UID>_ASSIGN9.
- Assignments not following this rule will not be graded. In addition, no resubmission after TA grades it. Late submission rule: 10% deduction for one day late. Late submission over a week is NOT acceptable.

DO NOT copy any codes from others. Otherwise, both will be penalized.