Instructions

- The homework is due on **Friday 3/31 at 5pm ET**.
- No extension will be provided, unless for serious documented reasons.
- Start early!
- Study the material taught in class, and feel free to do so in small groups, but the solutions should be a product of your own work.
- This is not a multiple choice homework; reasoning, and mathematical proofs are required before giving your final answer.

1 Reservoir Sampling [15 points]

Design an algorithm that samples $k \geq 1$ elements uniformly at random from an insert-only stream, whose length is unknown. Present the pseudocode and prove the correctness of the proposed algorithm.

2 Median trick - a useful technique [15 points]

Prove the claim on slide 13. Be specific about the values of the constants C_1, C_2 you use in your proof, where $t = C_1 \log \frac{1}{\delta}$, $k = C_2 \frac{\mathrm{Var}[X]}{\epsilon^2 \mathbb{E}[X]^2}$.

3 Variance of Morris Counter [20 points]

Prove equation $\mathrm{Var}(Z) = \frac{m(m-1)}{2}$ on slide 47.

4 More on uniform RVs [10+10 points]

Let $X_1, \ldots, X_n$ be iid uniform random variables, $X_i \in U(0,1)$ for all i. (a) What is the pdf and (b) what is the expectation of the k-th smallest value among $X_1, \ldots, X_n$ for $k = 1, \ldots, n$?

5 Coding [40 points]

Check the Jupyter notebook on our Git repo.