CS 809
Fall 2018
Homework #1

Due in class Friday September 21, 2018

Rules for Homework.

i.) Everyone must do his or her own work. Use of any sources other than class notes and recommended texts should be accompanied by a citation. In any case, there should be significant "value added" by the student's work.

ii.) Starred problems are optional. These are typically open-ended, and of unknown difficulty (at least to me). Think of these as questions with research potential.

iii.) Computer-aided math (e.g. Maple, Matlab) is acceptable, provided you play by these rules. It is OK to use the computer to execute routine tasks (e.g. solving linear equations) that you know how to do and could have done yourself. However, if the point of a problem is to gain experience with a particular procedure or algorithm, you should not use black-box software as a substitute for that experience.

1. In class, we studied the number of record values appearing in a sequence $X_1, \ldots, X_n$, where the X_i are chosen independently from some distribution with a density. For this exercise you are to consider discrete random variables.

   a) Prove the identity

$$\sum_{m=1}^{n} m a_m = n \sum_i a_i - \sum_{i=1}^{n-1} (a_1 + \ldots + a_i).$$

   Hint: Think of the left hand side as the sum of the entries in the following array:

$$\begin{array}{cccccc} a_1 & & & & & \\ a_2 & a_2 & & & & \\ a_3 & a_3 & a_3 & & & \\ & & \vdots & & & \\ a_n & a_n & a_n & a_n & \ldots & a_n \end{array}$$

b) Let the $X_1, \ldots, X_N$ be i.i.d. from the uniform distribution on $\{1, \ldots, n\}$. Records must now be strict, that is, greater than all predecessors. Find an exact expression for

$$E[\ \#\text{ of records }]$$

and prove that it is $\leq H_n$. Your expression should be capable of evaluation using $O(n \log N)$ operations.

Hint: Find the probability that X_i is not a record, by conditioning on $\max\{X_1, \ldots, X_{i-1}\}$ and then using a) to rewrite the sum. Complement this and sum over i. You will get a double summation that can be simplified using the sum of a finite geometric series.

c) More generally, let $X_1, \ldots, X_N$ be i.i.d. from any probability distribution on $\mathbf{Z}_{\geq 0}$. Show that

$$E[\ \#\text{ of records }] \leq H_N.$$

No computation is necessary; imagine dividing the interval [0,1] into bins.

d) (*) Consider other discrete distributions.