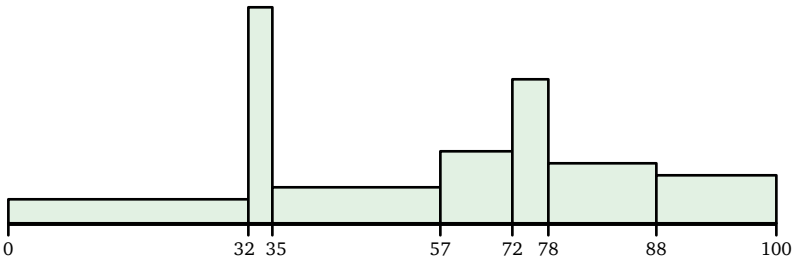
New CS 473: Algorithms, Spring 2015
Homework 4

Due Tuesday, February 24, 2015 at 5pm

Starting with this assignment, all homework must be submitted electronically via Moodle as separate PDF files, one for each numbered problem. Please see the course web site for more information.

---

1. Suppose we want to summarize a large set S of values—for example, course averages for students in CS 105—using a variable-width histogram. To construct a histogram, we choose a sorted sequence of ***breakpoints*** $b_0 < b_1 < \cdots < b_k$, such that every element of S lies between b_0 and b_k. Then for each interval $[b_{i-1}, b_i]$, the histogram includes a rectangle whose height is the number of elements of S that lie inside that interval.



A variable-width histogram with seven bins.

Unlike a standard histogram, which requires the intervals to have equal width, we are free to choose the breakpoints arbitrarily. For statistical purposes, it is useful for the *areas* of the rectangles to be as close to equal as possible. To that end, define the ***cost*** of a histogram to be the sum of the *squares* of the rectangle areas; we want to compute the histogram with minimum cost.

More formally, suppose we fix a sequence of breakpoints $b_0 < b_1 < \cdots < b_k$. For each index i, let n_i denote the number of input values in the ith interval:

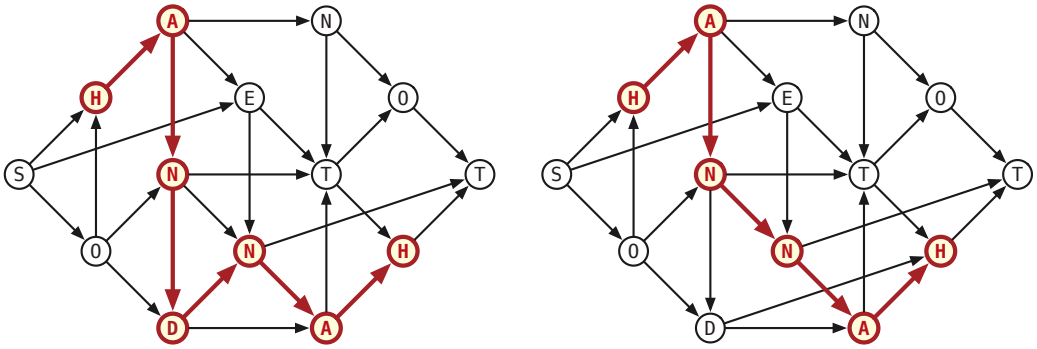
$$n_i := \#\left\{x \in S \;\middle|\; b_{i-1} \le x < b_i\right\}.$$

Then the ***cost*** of the resulting histogram is $\sum_{i=1}^{k} (n_i(b_i - b_{i-1}))^2$.

Describe and analyze an algorithm to compute a variable-width histogram with minimum cost for a given set of data values. Your input is an unsorted array $S[1..n]$ of ***distinct*** real numbers, all strictly between 0 and 1, and an integer k. Your algorithm should return a sorted array $B[0..k]$ of breakpoints that minimizes the cost of the resulting histogram, where $B[0] = 0$ and $B[k] = 1$***, and every other breakpoint $B[i]$ is equal to some input value $S[j]$***.

2. Suppose we are given a directed acyclic graph G with labeled vertices. Every path in G has a label, which is a string obtained by concatenating the labels of its vertices in order. Recall that a *palindrome* is a string that is equal to its reversal.

   Describe and analyze an algorithm to find the length of the longest palindrome that is the label of a path in G. For example, given the graph on the left below, your algorithm should return the integer 7, which is the length of the palindrome `HANDNAH`; given the graph on the right, your algorithm should return the integer 6, which is the length of the palindrome `HANNAH`.



New CS 473 Spring 2015 — Homework 4 Problem 1

Name:	NetID:
Name:	NetID:
Name:	NetID:

Describe and analyze an algorithm to compute a variable-width histogram with minimum cost for a given set of data values.

New CS 473 Spring 2015 — Homework 4 Problem 2

Name:	NetID:
Name:	NetID:
Name:	NetID:

Describe and analyze an algorithm to find the length of the longest palindrome path in a directed acyclic graph with labeled vertices.