Binary Search

Ryan Jian, Sreenath Are

October 18, 2013

1 What is Binary Search?

Binary search is one of the most fundamental algorithms in computer science. Recall that binary search allows us to quickly find a value in a sorted sequence by repeatedly halving the number of values we need to explicitly check, allowing us to find the location of a value in $O(\log N)$ time.

```
def binarysearch(A[0..N-1], target):
    sort(list) # if not already sorted
    low = 0, high = N - 1
    while low <= high:
        mid = (low + high) / 2
        if A[mid] > target:
            high = mid - 1
        else if A[mid] < target:
            low = mid + 1
        else: # we found it!
            return mid
    return -1 # not found
```

While most of you are used to using binary search on concrete sequences in the form of arrays, there is no need for this restriction. A sequence is really a function that maps indices to corresponding values, so binary search can be used to find the location x that any monotonic function $f(x)$ equals some target value.

This more abstract notion of a sequence can be very useful because for many problems, the function mapping an instance of the problem P to the corresponding answer is monotonic. Another way to say this is that if X is a possible answer to problem P and if the problem has the property where all $Y > X$ is a possible answer, then we can binary search on the answer. Since binary search has logarithmic time complexity, often times it can give us an efficient solution when other techniques fail.

2 Problems

1. I am thinking of a positive integer and for each guess, I will tell you whether the number is too high or too low. There is no upper bound on my numbers. Guess my numbers in $O(\log N)$ guesses, where N is my number.
   (Answer: Starting from 1, repeatedly double your guess until it is higher than N, giving you an upper bound. If k is the number of times you doubled, then binary search on between 2^{k-1} to 2^k)

2. (USACO NOV11 Silver) FJ's N cows are all standing at various positions along a line, each described by an integer position (i.e., its x coordinate) as well as an integer breed ID. FJ plans to take a photo of a contiguous range of cows along the line. The cost of the photo is the difference between the maximum and minimum x coordinates of the cows in the photo. Find the minimum cost of a photo in which there is at least one cow of each distinct breed appearing in FJ's herd.

3. (CodeJam 2013 Round 1A) A bullseye consists of a number of concentric rings. Maria is interested in manufacturing black-and-white bullseyes.

   Maria starts with t millilitres of black paint, which she will use to draw rings of thickness $1cm$ (one centimetre). A ring of thickness 1cm is the space between two concentric circles whose radii differ by $1cm$.

   Maria draws the first black ring around a white circle of radius r cm. Then she repeats the following process for as long as she has enough paint to do so:

   Maria imagines a white ring of thickness 1 cm around the last black ring. Then she draws a new black ring of thickness 1 cm around that white ring. Note that each "white ring" is simply the space between two black rings. The area of a disk with radius 1 cm is πcm^2. One millilitre of paint is required to cover area πcm^2. What is the maximum number of black rings that Maria can draw? Please note that: Maria only draws complete rings. If the remaining paint is not enough to draw a complete black ring, she stops painting immediately.

4. Given a list of N integers, find the consecutive subsequence of at least length K with the maximum average in $O(N \log N)$.

5. (Codeforces Round 200, Div 1) Mad scientist Mike does not use slow hard disks. His modification of a hard drive has not one, but n different heads that can read data in parallel.

   When viewed from the side, Mike's hard drive is an endless array of tracks. The tracks of the array are numbered from left to right with integers, starting with 1. In the initial state the i-th reading head is above the track number hi. For each of the reading heads, the hard drive's firmware can move the head exactly one track to the right or to the left, or leave it on the current track. During the operation each head's movement does not affect the movement of the other heads: the heads can change their relative order; there can be multiple reading heads above any of the tracks. A track is considered read if at least one head has visited this track. In particular, all of the tracks numbered h1, h2, ..., hn have been read at the beginning of the operation.

   Mike needs to read the data on m distinct tracks with numbers p1, p2, ..., pm. Determine the minimum time the hard drive firmware needs to move the heads and read all the given tracks. Note that an arbitrary number of other tracks can also be read.