

COMP 555 Bioalgorithms

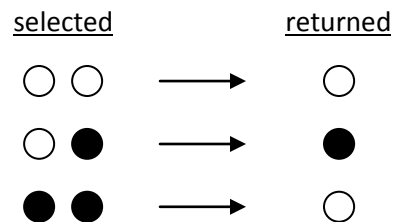
Fall 2012

Problem Set 1

Assigned: Tue Sep 11

Due: Tue Sep 25 (in class)

1. Here is another rock game, this time for a single player. A single pile contains rocks of two colors: black and white. The player repeatedly removes two rocks chosen at random from the pile (the pile is out of sight inside a box!) and then returns one rock according to the following rules:



In words, if the two selected rocks have the same color, a white rock is returned; if they are different in color, a black rock is returned (extra white rocks are available if needed). The pile will eventually be reduced to a single rock since the number of rocks is decreasing by one each time. Suppose the initial pile has twice as many black rocks as white rocks. What can you say about the chance that the last rock that remains is black? Explain your reasoning.

2. Study the idealized PCR reaction described in section 3.8 of the text. Starting from an initial DNA molecule, the Denaturation – Priming – Extension cycle creates two non-identical DNA molecules (see Fig. 3.3 and 3.4 in the text). The second cycle yields *four* non-identical DNA molecules. Identify the different DNA molecules present after n cycles and show how many copies are present of each.
3. Given a long text string T of length n and a short pattern string s of length m , find the first occurrence in T of a string s' such that $d_H(s, s') \leq 1$. What is the complexity of your algorithm?
4. The double digest problem discussed in class (lecture 4 slide 6) consists of sets A and B containing fragment lengths observed after treatment of DNA using two different restriction enzymes, respectively, and a set AB that contains fragment lengths produced after treatment of

DNA with both enzymes simultaneously. Assume, for simplicity, that all observed fragments in the digests differ in length (so A, B, and AB are truly sets rather than multisets), and that there are $O(n)$ fragments in A and B.

(a) We proposed an algorithm in class, shown here in pseudocode, to search for an ordering of fragments in A and B consistent with the lengths in AB:

```
foreach ab in perm(AB)
  foreach a in perm(A)
    foreach b in perm(B)
      if compatible1(a, ab) and compatible1(b, ab)
        return(a, b)
```

Describe an $O(n)$ complexity algorithm to compute `compatible1`.

(b) Consider the following rearrangement of the search:

```
foreach a in perm(A)
  foreach b in perm(B)
    if compatible2(a, b, AB)
      return(a, b)
```

Can you construct an $O(n)$ complexity algorithm for `compatible2`? You may preprocess AB in advance of the loops if you wish.

(c) Devise a branch and bound strategy for the algorithm in (b).

(d) Suppose you can select the restriction enzymes used for A and B. Discuss the advantages and disadvantages of selecting enzymes for which the expected fragment lengths are very different.