Épreuve de section européenne

Hamming distance

Here are the six ways to take three letters a, b, c and arrange them in different ordered strings : *abc* / *acb* / *bac* / *bca* / *cab* / *cba*.

The strings are listed the way they would be in a dictionary, where words that start with a come before those that start with b, and these in turn come before those that start with c. Which word, *able* or *alert*, comes first in the dictionary ? Since both words start with a we have a tie, so we look at the letter in the second position. Since b comes before l, we list *able* before *alert*. This approach to ordering strings is known as lexicographical ordering. It begins with a system for an ordering of alphabetical strings of length 1 and extends to order strings of any length. Note that to carry such a process one has to know the "ordering" for the letters in the alphabet.

Hamming distance is the number of "columns" (positions) the two strings differ in, if the strings appear one on top of the other. Thus, *bleak* and *sneak* differ in two positions. The two strings differ in columns one and two, so they have Hamming distance 2.

If we compute the Hamming distance between the consecutive rows of our initial list and, when we get to the bottom of the list, think of the list as "cyclic," that is, returning to the start, or first row, and compute that Hamming distance as well, we get the total sum of distances between rows of 15.

Is this the smallest possible total that can be obtained ? No !

Since the strings have distinct symbols, the smallest number of changes between two consecutive strings must be 2. In the sequence list below, the Hamming distance between two consecutive strings, including the first and last string, is two. Hence, this must be an ordering where the sum of the consecutive distances (including the top and bottom strings) is a minimum. The minimum value is 12.

abc / *bac* / *bca* / *cba* / *cab* / *acb*

From *Gray Codes*, an AMS Feature Column, by Joseph Malkevitch

Questions

1. Explain in your own words the definition of the Hamming distance and compute it for the strings *abcde* and *ebadc*.
2. Check that the sum of Hamming distances in the first sequence is 15, and that it is 12 in the last sequence.
3. **a.** Find out the 24 possible ways to take four letters a, b, c, d and arrange them in different ordered strings. List them with the lexicographical ordering.
   b. Compute the sum of Hamming distances in the list you found in the question a.
   c. Try to find an ordering of the strings that will give a lower sum of Hamming distances.