

CSE 582 Bioinformatics Final

I. QUESTION

What is the optimal global alignment for APPLE and HAPPE? Show all optimal alignments and the corresponding paths under the match premium +1, mismatch penalty -1, and indel penalty -1.

II. QUESTION

What is the optimal local alignment for APPLE and HAPPE? Show all optimal alignments and the corresponding paths under the match premium +1, mismatch penalty -1, and indel penalty -1.

III. QUESTION

Find the optimal pairwise global alignment of the sequences TACGAGTACGA and ACTGACGACTGAC with the condition that **G** nucleotides shown in bold font must be aligned together. The scoring parameters are defined as +2 for match, -1 for mismatch, and $d = -2$ for a linear gap penalty.

IV. QUESTION

Determine the number of different ways to partition a set of N elements into k clusters.

V. QUESTION

Suppose a dishonest dealer has two coins, one fair and one biased; the biased coin has heads probability $1/4$. Assume that the dealer never switches the coins. Which coin is more likely to have generated the sequence HTTTHHHTTTTTHHTT? ($\log_2(3) = 1.585$).

VI. QUESTION

Derive the expectation and the variance of the number of occurrences of patterns AA and AT in a random text of length N . Are the expectations and variances for AA and AT the same?

VII. QUESTION

A 4200 nt long DNA sequence is used as a training set for parameter estimation of the DNA statistical model. The observed counts of sixteen dinucleotides, N_{XY} , are as follows.

	T	C	A	G
T	510	380	210	190
C	240	170	360	230
A	370	200	220	210
G	190	170	220	220

Find the maximum likelihood estimates of (a) the transition probabilities P_{TT} , P_{AG} of the first order Markov model of the DNA sequence; (b) the transition probabilities P_{TT} , P_{AG} of the first order Markov model for the DNA sequence complementary to the given training sequence.

VIII. QUESTION

Construct unrooted phylogenetic tree from the following distance matrix:

	A	B	C	D	E	F
A	0	37	55	69	68	25
B	37	0	42	56	55	38
C	55	42	0	46	45	56
D	69	56	46	0	25	70
E	68	55	45	25	0	69
F	25	38	56	70	69	0

A, B, C, D, E, and F are the names of the species.