

Homework 1
Due September 25
to: teresa.Przytycka@verizon.net

1. (4p) Assume that it was observed that in a multiple sequence alignment for some protein family one column is well preserved. In fact, only three amino-acids are observed in this column: M is observed 40% of the time, L 25% of the time, V 10% of the time, and I is observed 25% of the time.
How much information (in bits) carries this position (column)? As a reference distribution, R , assume uniform distribution over all 20 amino-acids.

2. (6p) Given is a multiple alignment family of related RNA sequences. Two positions (columns) i and j are tested for possible interaction. It is observed that simultaneous occurrence of A at position i and U at position j happens 40% of the time, 20% of the time there is U at position i and A at position j , 15% of the time there is C at position i and G at position j and 10% of the time there is C at position i and U at position j , finally U at position i and C at position j is 15% of the time. No other combinations are observed.
 - a. What are the marginal probability distributions at positions i and j
 - b. What is the mutual information at positions i and j .