

Assignment #1

1. (a) Recover the marginals that were used to create the BLOSUM62 matrix using *only* that matrix.
 (b) Would this method work for *any* log-odds matrix?
2. (a) Recover the joint probability matrix $P = (p_{ij})$ of “real alignments” that was used to create BLOSUM62 (the H_1 hypothesis).
 (b) What is the expected number of mutations at each position under H_1 ?
 (c) What is it under H_0 ?
3. (a) Use figure 80 on p.346 in Dayhoff et al. to build the jump matrix Q .
 (b) Use the BLOSUM62 marginals computed above and the method showed in class to compute a new transition matrix corresponding to a distance of 1 PAM.
4. (a) Using this new “PAM1” matrix what is the expected number of mutations at a distance of 159 PAMs? Compare with Table 23 on p.351 of Dayhoff et al.
 (b) Construct a new PAM160 or PAM250 matrices and compare to the original ones.
5. Exercises 2.5 & 2.6 from Durbin et al. (modified slightly):
 - (a) Show that the number of ways of intercalating two sequences of lengths n and m to give a single sequence of length $n + m$, while preserving the order of the symbols in each, is $\binom{m+n}{m}$.
 - (b) By taking alternating symbols from the upper and lower sequences in an alignment, then discarding the gap characters, show that there is a one-to-one correspondence between gapped alignments of the two sequences and intercalated sequences of the type described above. Note: for this procedure to work you should rule out alignments with $\begin{matrix} \text{A-} \\ -\text{B} \end{matrix}$ (but $\begin{matrix} -\text{A} \\ \text{B-} \end{matrix}$ is OK).
 - (c) How many different such (i.e. subject to the rule above) global alignments are there?