Math 5233 PSSMs

Group members (1 to 3): _____
Due May 2nd in class.

(1) Calculate the score of matching the sequence AAGCCT with a PSSM derived from the aligned sequences

$$\begin{array}{l} \texttt{AACCGT} \\ \texttt{CACCCT} \\ \texttt{GCGCGC} \\ \texttt{TCCGCC} \end{array}$$

Recall that the score for position i and letter j is $S(i,j) = log_2(4f(i,j))$ where $f(i,j)$ is the frequency of letter j in column i, if we are comparing to the uniform background distribution of four DNA letters.

(2) Sketch the sequence logo for the four sequences above. The height of all the letters in each position is the relative entropy $\sum f(i,j) \log_2(4f(i,j))$, and within that the height of each letter is proportional to its frequency $f(i,j)$.