CS 6104
Homework Assignment 4

Given: June 15, 2000 **Due:** June 20, 2000

The point value of each problem is shown in []. The assignment is due by 9:30AM on June 20, 2000. Prepare your solutions in LaTeX, preferably using this file as a starting point. You may submit your solutions in printed form or by email to `cs6104@courses.cs.vt.edu`. Explain your solution to each problem, including references to the appropriate material in the textbook.

Help is available by email as well as during my office hours. It is especially helpful to request clarification or hints by email to `cs6104@courses.cs.vt.edu`, so I can send the response to everyone.

[1] **1.** Consider the following alignment of 6 DNA sequences:

C	T	T	A	G	C	T
C	C	-	-	G	C	A
T	C	-	-	-	C	T
G	C	T	A	-	G	A
C	T	T	-	-	G	A
C	T	T	T	G	C	T

From the alignment, we see that not every base is equally likely. Use the data in the alignment to fill in this table of probabilities:

	A	C	G	T
PROBABILITY				

---

[3] **2.** Find a Markov chain for the alignment in question 1. Derive a consensus (highest scoring) sequence from your Markov chain.

---

[3] **3.** Find a hidden Markov model (HMM) for the alignment in question 1. Derive a consensus (highest scoring) sequence from your HMM.

---

[3] **4.** Find a Bayesian network for the last three columns of the alignment in question 1. Derive a consensus (highest scoring) 3-base sequence from your Bayesian network. N.B. A Bayesian network must be acyclic.
