Question 4

Consider the following way of representing a rooted tree:

11	1	5.0	12	4.0
12	2	1.0	3	1.0
13	4	9.0	5	9.0
14	6	10.0	15	4.0
15	16	2.0	17	4.0
16	7	4.0	8	4.0
17	9	2.0	10	2.0
18	11	14.0	19	3.0
19	13	7.0	14	6.0

Each line represents an internal node in the tree: the first column is the node label, and subsequent pairs of columns represent daughter nodes by their label and branch length. So for example, node 11 is the parent of nodes 1 and 12, connected to them by branches of length 5.0 and 4.0 respectively.

(a) Write a program which reads a data structure in this format and computes, for each internal node, the numbers of internal nodes and leaves below it. Print these out as two extra numbers at the end of each row in the table.

(b) Modify your program to calculate the branch length from each leaf to the root. Hence evaluate whether the tree is consistent with an evolutionary tree in which branch lengths represent times and all the leaves are at the present time.

(c) Assuming a Jukes-Cantor model and leaf values (x1, x2, x3, x4, x5, x6, x7, x8, x9, x10) = (A, C, A, G, G, A, T, C, A, T), make your program calculate the likelihood of this tree given mutation rate μ using Felsenstein's algorithm. Plot the log likelihood for μ in the range (0,1] and hence infer the maximum likelihood value of μ to two significant figures.

(d) Randomly assign new leaf values. Calculate the new maximum likelihood value of μ. Report your leaf assignments and the maximum likelihood estimate (MLE) to two significant figures. Calculate the distribution of μ MLEs for 1000 random assignments and briefly discuss (100 words).