Problem Set 2

1) Please indicate the parameters of an HKY-SSR₅ model.

2) Use the character-by-taxon matrix below to demonstrate why one should never map characters onto a strict consensus tree.

   | A   | B   | C   | D   | E   |
---|-----|-----|-----|-----|-----|
A  | 0 0 | 1 0 | 0 1 | 0 1 | 1 1 |
B  | 0 0 | 1 0 | 1 1 | 0 1 | 1 1 |
C  | 0 0 | 1 0 | 1 1 | 1 1 | 0 1 |
D  | 0 0 | 1 0 | 1 1 | 1 1 | 0 1 |
E  | 1 1 | 1 1 | 1 1 | 1 1 | 0 0 |

3) Please draw a likelihood surface for two trees where one has a higher joint likelihood with respect to $p_{inv}$ and the other has a higher marginal likelihood.

4) Please write the formula for the AIC, and provide a simple justification and its theoretical justification.

5) What is the difference between consistency and efficiency?

6) What’s the difference between a KH test that uses the normal approximation and a KH test with a RELL bootstrap?

7) Why should the GTR+CAT model used in RAxML only be used for very large data sets?

8) Please write Bayes’ Theorem for the posterior distribution of trees.