

Given a sample of sequences from a population, it is important to calculate the probability that they evolved from a known ancestor in order to make estimates of mutation rates, and to produce accurate models of population structure, selection and recombination. The infinite sites model is commonly used to model mutations in sequences. There is no known algorithm that could sum the likelihood over the evolutionary histories quickly, and importance sampling methods also fail on large data sets. The aim of the project is to infer the theoretical mixing of Markov chain Monte Carlo methods under the infinite site model.