Assignments I - Markovian modeling and Bayesian learning

1. Consider a state space $\mathcal{X} = \{A, C, G, T\}$. For each of the following conditions, define a discrete time Markov chain (DTMC) in $\mathcal{X}$ with corresponding properties and draw its network graph:

a) is reducible

b) has two communicating classes

c) is irreducible, aperiodic and recurrent, then show what the stationary distribution is.

2. Simulate 500 states from the DTMC you defined in 1c and demonstrate how the empirical distribution over the states of the process converges to the distribution derived in 1c. Report also commented code for simulation.

3. It is sometimes useful to try to relate frequentist probabilities with those derived using Bayesian reasoning. Assume we have a population of size 1,000 individuals from K-Pax out of which 10 have a certain characteristic, say 'A'. It is quite laborious to detect whether an individual has 'A', and therefore, we have constructed a machine which is fed with visual images of an individual and it judges whether the individual is 'A' or 'non-A'. We are quite clever scientists and the machine is able with probability 0.99 to correctly yield an 'A', when the individual in question has 'A'. That is, the machine yields an answer 'non-A' with probability 0.01 in those cases. Furthermore, the machine is equally accurate for detecting 'non-A', which also happens with probability 0.99. Assume now the machine spits out an 'A' for a randomly chosen individual from the mentioned K-Pax population. Using Bayes' theorem calculate the posterior probability that this individual in fact has the 'A' characteristic. Then, relate the answer to the empirical distribution of

$$\frac{\#\text{real 'A'}}{\#\text{real 'A'} + \#\text{fake 'A'}},$$

obtained through repeating the population experiment a large number of times. The above fraction is the fraction of real 'A' cases among all 'A' cases indicated by the machine when everyone among the 1,000 K-Pax individuals is tested. Notice that the 10/1000 fraction in the underlying population should be the same in every replicate and that the statistical variation stems from the properties of our machine.

4. In Section 3.1 of 'Bayesian statistics without tears' (see course literature) fundamental theorem of simulation is connected with Baysian learning that needs no input from calculus methods. This is useful for understanding the mechanics of Bayes' theorem when the random quantity in question is continuous. Assume θ is the unknown parameter in a Binomial(n, θ) experiment, where n is a fixed constant (fixed by experimental design). Assume further we observe $x = 7$ 'successful' outcomes out of $n = 10$ in the experiment. Specify a uniform prior Unif(0,1) for θ. Define the posterior of θ and simulate it by the rejection method using the prior. Notice that you can first sample m points uniformly in the unit rectangle $(0,1) \times (0,1)$, where the axes are θ and u in 3.1. Then, each of these points is either rejected or accepted, such that the accepted ones represent a sample from the posterior. Report your code and the plots of the samples and compare the empirical CDF of the posterior with the analytically derived form of the posterior which is a Beta distribution (see, e.g. the notes by L. Gu, the book by Koski & Noble, or lecture slides for 'Bayesian theory with applications').

5. Consider the two sequences of 15 binary observations with categories U/D: $\mathbf{x} = (U, D, U, U, D, D, D, U, D, U, U, U, D, D, U)$,
$\mathbf{y} = (U, U, U, U, U, U, U, U, D, D, D, D, D, D, D)$. Discuss consequences of assuming such sequences to be generated by an *i.i.d.* Bernoulli process, how reasonable is this? What are the particular consequences for a) prediction of future states of the process such as x_{16}, b) model learning, c) understanding uncertainty in the phenomenon from which the observations are interpreted to be generated and judging the qualitative structure of the model.