## Course on Models & Algorithms 2022-23 – Final Project

Choose one of the following problems and write a detailed notebook explaining your results.

### (a) Parameter inference from partial observations of a Lotka-Volterra process.

Consider a population of two interacting species, predators and prey. The state of the system at each time step (we consider here a discrete time description of the dynamic) is described by the number of individuals belonging to the two categories, $N_{pred}(t)$ and $N_{prey}(t)$ for predators and prey respectively. During the dynamic three events can happen:

1. Each one of the prey can reproduce (breed) with a probability $\lambda_{\text{breed}}$;
2. A predator can kill one of the prey and reproduce with a probability $\lambda_{\text{interaction}}$;
3. A predator dies with probability $\lambda_{\text{death}}$.

Questions:

- Simulate a dynamical system of this type using a population-based implementation. Study the behavior of the dynamical system when changing the values of parameters $(\lambda_{\text{breed}}, \lambda_{\text{interaction}}, \lambda_{\text{death}})$ and identify an interesting regime, i.e. the two species do not face extinction after a few steps of the simulation.
- In this regime, simulate a stationary dynamics and collect observations of the number of new prey and new predators over a time interval $\Delta T$ with a probability of observation $p_{obs} = 0.3$.
- Design an Approximate Bayesian Computation Monte Carlo (ABCMC) method to learn two of the three probabilities, while setting the third one equal to the true one.
- (Optional) Perform the inference of the three parameters in a unique ABCMC scheme.

### (b) Bayesian inference of a contact network of Potts variables.

Consider a system of $N = 5$ Potts variables and a set of $M$ given observations $\left\{\boldsymbol{x}^{(m)}\right\}_{m=1}^{M}$ where each $\boldsymbol{x}^{(m)} = \left(x_1^{(m)}, x_2^{(m)}, \ldots, x_N^{(m)}\right)$ represent a configuration of the $N$ variables. Each $x_i^{(m)} \in \{1, \ldots, q\}$ is a categorical variable representing one of the possible $q$ colors (for instance "1" $\to$ "red", "2" $\to$ "blue",etc.).

We model the interactions among these variables through the so-called Potts model

$$
\begin{aligned}
P(\boldsymbol{x}|\mathbf{J}) &= \frac{1}{Z} e^{\sum_{i,j} J_{i,j}(x_i,x_j)} \\
&= \frac{1}{Z} e^{\sum_{i,j}\sum_{a,b} J_{i,j}(a,b)\delta_{x_i,a}\delta_{x_j,b}}
\end{aligned}
$$

where each $J_{ij}$ is a coupling matrix of dimension $q \times q$. Given the data introduced above we consider the likelihood of the data as

$$
P\left(\left\{\boldsymbol{x}^{(m)}\right\}_{m=1}^{M} |\mathbf{J}\right) = \frac{1}{Z} e^{\sum_{i,j} J_{i,j}\left(x_i^{(m)},x_j^{(m)}\right)}
$$

Questions:

- Consider the file `data.dat` and design a Boltzmann machine learning scheme (or maximum likelihood estimator) aimed at inferring the coupling matrices. You can use a Metropolis-Hastings Monte Carlo or a Gibbs sampling to estimate the statistics of the model.
- From the inferred matrices $J^{\star}$, compute the so-called Frobenius norms

  $$
  \mathcal{F}_{ij} = \sqrt{\sum_{a,b} J_{ij}^{\star}(a,b)^2}
  $$

  and compare them to the true interaction graph in the file `groundtruth.dat` (in this file, the row $i^{th}$ contains the list of neighbors of individual $i$).
- (Optional) Consider an additional prior

  $$
  P(\mathbf{J}) = e^{-\lambda \sum_{i,j,a,b} |J_{i,j}(a,b)|}
  $$

for a small value of $\lambda$, and modify the Boltzmann machine according to the posterior distribution

$$
P\left(\mathbf{J}|\left\{\boldsymbol{x}^{(m)}\right\}_{m=1}^{M}\right) \propto P\left(\left\{\boldsymbol{x}^{(m)}\right\}_{m=1}^{M}|\mathbf{J}\right) P(\mathbf{J})
$$

- Repeat the inference process several times by changing the number of configurations $M$ used to estimate the statistics of the data. Evaluate how your final estimate (the Frobenius norms) is affected by $M$.
- At the convergence of your Boltzmann machine, run a Monte Carlo sampling and evaluate the auto-correlation function associated with the samples while varying $(M, T_{wait})$ where $T_{wait}$ is the sampling time.