What's selection?
Selection happens when favor mutation occurs. The population who have this favor gene will have higher chance to survive. The population of left explodes and the right group constricts. The goal of this project is that we are given a SNP matrix, then , we want to estimate how long does the selection onset and the strength of selection.

We use MSMS for simulation. Changing the coefficient after SAA and tau, we can get the selective SNP matrix in different period. Totally 3000 sample data and 200 test data were made. The figure on left is the allele spectrum of simulation result. Each line represents an average over 100 samples. The line on bottom represents the selection just onsets. The time step between 2 lines is 0.1N where N is total population. We observe the spectrum shifts from a U shape to a heading down line. Based on this figure, we can estimate the coefficient tau by using KNN algorithm.

However, the spectrum doesn't work as we expect. The spectrum for one SNP matrix looks like these. Because gene mutates randomly, it's hard for KNN to separate them. The correct rate for 3000 training samples and 200 test cases is 0.2%.

To figure out this problem, we first increase the time step in the sampling. We find that if we sample SNP matrix in the timeline further, correct rate increases to 20%. Then, we go one step further.